Bruce K. Driver

# Probability Tools with Examples

September 28, 2018 *File:prob.tex*

# Contents

## Part Homework Problems



## Part I Background Material



## Part II Formal Development

# Part

# Homework Problems

# Math 280A Homework Problems Fall 2018

Problems are from Resnick, S. A Probability Path, Birkhauser, or from the lecture notes. The problems from the lecture notes are restated here.

## -3.1 Homework 1. Due Friday, October 5, 2018

- Read over Lecture notes Chapter 1.
- Lecture note Exercises: 1.1, 1.2, and 1.3.

# Part I

# Background Material

# 1

# Limsups, Liminfs and Extended Limits

**Notation 1.1** *The* ***extended real numbers*** *is the set* $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, *i.e. it is* $\mathbb{R}$ *with two new points called* $\infty$ *and* $-\infty$. *We use the following conventions,* $\pm\infty \cdot 0 = 0$, $\pm\infty \cdot a = \pm\infty$ *if* $a \in \mathbb{R}$ *with* $a > 0$, $\pm\infty \cdot a = \mp\infty$ *if* $a \in \mathbb{R}$ *with* $a < 0$, $\pm\infty + a = \pm\infty$ *for any* $a \in \mathbb{R}$, $\infty + \infty = \infty$ *and* $-\infty - \infty = -\infty$ *while* $\infty - \infty$ *is not defined. A sequence* $a_n \in \bar{\mathbb{R}}$ *is said to converge to* $\infty$ $(-\infty)$ *if for all* $M \in \mathbb{R}$ *there exists* $m \in \mathbb{N}$ *such that* $a_n \geq M$ $(a_n \leq M)$ *for all* $n \geq m$.

**Lemma 1.2.** *Suppose* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *are convergent sequences in* $\bar{\mathbb{R}}$, *then:*

1. *If* $a_n \leq b_n$ *for*[1] *a.a.* $n$, *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
2. *If* $c \in \mathbb{R}$, *then* $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$.
3. $\{a_n + b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \tag{1.1}$$
*provided the right side is not of the form* $\infty - \infty$.
4. $\{a_n b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{1.2}$$
*provided the right hand side is not of the for* $\pm\infty \cdot 0$ *of* $0 \cdot (\pm\infty)$.

Before going to the proof consider the simple example where $a_n = n$ and $b_n = -\alpha n$ with $\alpha > 0$. Then

$$\lim (a_n + b_n) = \begin{cases} \infty & \text{if } \alpha < 1 \\ 0 & \text{if } \alpha = 1 \\ -\infty & \text{if } \alpha > 1 \end{cases}$$

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{“}=\text{”} \infty - \infty.$$

This shows that the requirement that the right side of Eq. (1.1) is not of form $\infty - \infty$ is necessary in Lemma 1.2. Similarly by considering the examples $a_n = n$ and $b_n = n^{-\alpha}$ with $\alpha > 0$ shows the necessity for assuming right hand side of Eq. (1.2) is not of the form $\infty \cdot 0$.

**Proof.** The proofs of items 1. and 2. are left to the reader.

**Proof of Eq.** (1.1). Let $a := \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Case 1., suppose $b = \infty$ in which case we must assume $a > -\infty$. In this case, for every $M > 0$, there exists $N$ such that $b_n \geq M$ and $a_n \geq a - 1$ for all $n \geq N$ and this implies

$$a_n + b_n \geq M + a - 1 \text{ for all } n \geq N.$$

Since $M$ is arbitrary it follows that $a_n + b_n \to \infty$ as $n \to \infty$. The cases where $b = -\infty$ or $a = \pm\infty$ are handled similarly. Case 2. If $a, b \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|a - a_n| \leq \varepsilon \text{ and } |b - b_n| \leq \varepsilon \text{ for all } n \geq N.$$

Therefore,

$$|a + b - (a_n + b_n)| = |a - a_n + b - b_n| \leq |a - a_n| + |b - b_n| \leq 2\varepsilon$$

for all $n \geq N$. Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

**Proof of Eq.** (1.2). It will be left to the reader to prove the case where $\lim a_n$ and $\lim b_n$ exist in $\mathbb{R}$. I will only consider the case where $a = \lim_{n\to\infty} a_n \neq 0$ and $\lim_{n\to\infty} b_n = \infty$ here. Let us also suppose that $a > 0$ (the case $a < 0$ is handled similarly) and let $\alpha := \min\left(\frac{a}{2}, 1\right)$. Given any $M < \infty$, there exists $N \in \mathbb{N}$ such that $a_n \geq \alpha$ and $b_n \geq M$ for all $n \geq N$ and for this choice of $N$, $a_n b_n \geq M\alpha$ for all $n \geq N$. Since $\alpha > 0$ is fixed and $M$ is arbitrary it follows that $\lim_{n\to\infty} (a_n b_n) = \infty$ as desired. ■

For any subset $\Lambda \subset \bar{\mathbb{R}}$, let $\sup \Lambda$ and $\inf \Lambda$ denote the least upper bound and greatest lower bound of $\Lambda$ respectively. The convention being that $\sup \Lambda = \infty$ if $\infty \in \Lambda$ or $\Lambda$ is not bounded from above and $\inf \Lambda = -\infty$ if $-\infty \in \Lambda$ or $\Lambda$ is not bounded from below. We will also use the **conventions** that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

**Notation 1.3** *Suppose that* $\{x_n\}_{n=1}^\infty \subset \bar{\mathbb{R}}$ *is a sequence of numbers. Then*

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k \geq n\} \text{ and} \tag{1.3}$$
$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{1.4}$$

[1] Here we use “a.a. $n$” as an abbreviation for almost all $n$. So $a_n \leq b_n$ a.a. $n$ iff there exists $N < \infty$ such that $a_n \leq b_n$ for all $n \geq N$.

*We will also write* $\underline{\lim}$ *for* $\liminf_{n\to\infty}$ *and* $\overline{\lim}$ *for* $\limsup\limits_{n\to\infty}$.

*Remark 1.4.* Notice that if $a_k := \inf\{x_k : k \ge n\}$ and $b_k := \sup\{x_k : k \ge n\}$, then $\{a_k\}$ is an increasing sequence while $\{b_k\}$ is a decreasing sequence. Therefore the limits in Eq. (1.3) and Eq. (1.4) always exist in $\bar{\mathbb{R}}$ and

$$\liminf_{n\to\infty} x_n = \sup_n \inf\{x_k : k \ge n\} \text{ and}$$
$$\limsup_{n\to\infty} x_n = \inf_n \sup\{x_k : k \ge n\}.$$

The following proposition contains some basic properties of liminfs and limsups.

**Proposition 1.5.** *Let* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *be two sequences of real numbers. Then*

1. $\liminf_{n\to\infty} a_n \le \limsup\limits_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} a_n$ *exists in* $\bar{\mathbb{R}}$ *iff*
$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n \in \bar{\mathbb{R}}.$$
2. *There is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \limsup\limits_{n\to\infty} a_n$. *Similarly, there is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \liminf_{n\to\infty} a_n$.
3. $$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{1.5}$$
*whenever the right side of this equation is not of the form* $\infty - \infty$.
4. *If* $a_n \ge 0$ *and* $b_n \ge 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty}(a_n b_n) \le \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{1.6}$$
*provided the right hand side of (1.6) is not of the form* $0 \cdot \infty$ *or* $\infty \cdot 0$.

**Proof.** 1. Since

$$\inf\{a_k : k \ge n\} \le \sup\{a_k : k \ge n\}\ \forall\, n,$$

$$\liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n.$$

Now suppose that $\liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n = a \in \mathbb{R}$. Then for all $\varepsilon > 0$, there is an integer $N$ such that

$$a - \varepsilon \le \inf\{a_k : k \ge N\} \le \sup\{a_k : k \ge N\} \le a + \varepsilon,$$

i.e.

$$a - \varepsilon \le a_k \le a + \varepsilon \text{ for all } k \ge N.$$

Hence by the definition of the limit, $\lim_{k\to\infty} a_k = a$. If $\liminf_{n\to\infty} a_n = \infty$, then we know for all $M \in (0,\infty)$ there is an integer $N$ such that

$$M \le \inf\{a_k : k \ge N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup\limits_{n\to\infty} a_n = -\infty$ is handled similarly.

Conversely, suppose that $\lim_{n\to\infty} a_n = A \in \bar{\mathbb{R}}$ exists. If $A \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that $|A - a_n| \le \varepsilon$ for all $n \ge N(\varepsilon)$, i.e.

$$A - \varepsilon \le a_n \le A + \varepsilon \text{ for all } n \ge N(\varepsilon).$$

From this we learn that

$$A - \varepsilon \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le A + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$A \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le A,$$

i.e. that $A = \liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$. If $A = \infty$, then for all $M > 0$ there exists $N = N(M)$ such that $a_n \ge M$ for all $n \ge N$. This show that $\liminf_{n\to\infty} a_n \ge M$ and since $M$ is arbitrary it follows that

$$\infty \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n.$$

The proof for the case $A = -\infty$ is analogous to the $A = \infty$ case.

2. – 4. The remaining items are left as an exercise to the reader. It may be useful to keep the following simple example in mind. Let $a_n = (-1)^n$ and $b_n = -a_n = (-1)^{n+1}$. Then $a_n + b_n = 0$ so that

$$0 = \lim_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty}(a_n + b_n) = \limsup_{n\to\infty}(a_n + b_n)$$

while

$$\liminf_{n\to\infty} a_n = \liminf_{n\to\infty} b_n = -1 \text{ and}$$
$$\limsup_{n\to\infty} a_n = \limsup_{n\to\infty} b_n = 1.$$

Thus in this case we have

$$\limsup_{n\to\infty}(a_n + b_n) < \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \text{ and}$$
$$\liminf_{n\to\infty}(a_n + b_n) > \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n.$$

■

## 1.1 Infinite sums

**Definition 1.6.** *For* $a_n \in [0,\infty]$, *let*

$$\sum_{n=1}^{\infty} a_n := \lim_{N\to\infty} \sum_{n=1}^{N} a_n = \sup_N \sum_{n=1}^{N} a_n.$$

*Remark 1.7.* If $a_n, b_n \in [0,\infty]$ and $\lambda \geq 0$, then

$$\sum_{n=1}^{\infty} (a_n + \lambda b_n) = \sum_{n=1}^{\infty} a_n + \lambda \cdot \sum_{n=1}^{\infty} b_n.$$

Indeed,

$$\begin{aligned}\sum_{n=1}^{\infty} (a_n + \lambda b_n) &= \lim_{N\to\infty} \sum_{n=1}^{N} (a_n + \lambda b_n) = \lim_{N\to\infty} \left[\sum_{n=1}^{N} a_n + \lambda \sum_{n=1}^{N} b_n\right] \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} a_n + \lambda \lim_{N\to\infty} \sum_{n=1}^{N} b_n = \sum_{n=1}^{\infty} a_n + \lambda \cdot \sum_{n=1}^{\infty} b_n.\end{aligned}$$

We will refer to the following basic proposition as the monotone convergence theorem for sums (MCT for short).

**Proposition 1.8 (MCT for sums).** *Suppose that for each* $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^{\infty}$ *is a sequence in* $[0,\infty]$ *such that* $\uparrow \lim_{n\to\infty} f_n(i) = f(i)$ *by which we mean* $f_n(i) \uparrow f(i)$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} f(i), \text{ i.e.}$$

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} \lim_{n\to\infty} f_n(i).$$

*We allow for the possibility that these expression may equal to* $+\infty$.

**Proof.** Let $M :=\uparrow \lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i)$. As $f_n(i) \leq f(i)$ for all $n$ it follows that $\sum_{i=1}^{\infty} f_n(i) \leq \sum_{i=1}^{\infty} f(i)$ for all $n$ and therefore passing to the limit shows $M \leq \sum_{i=1}^{\infty} f(i)$. If $N \in \mathbb{N}$ we have,

$$\sum_{i=1}^{N} f(i) = \sum_{i=1}^{N} \lim_{n\to\infty} f_n(i) = \lim_{n\to\infty} \sum_{i=1}^{N} f_n(i) \leq \lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = M.$$

Letting $N \uparrow \infty$ in this equation then shows $\sum_{i=1}^{\infty} f(i) \leq M$ which completes the proof. ■

**Proposition 1.9 (Tonelli's theorem for sums).** *If* $\{a_{kn}\}_{k,n=1}^{\infty} \subset [0,\infty]$, *then*

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn}.$$

*Here we allow for one and hence both sides to be infinite.*

**Proof. First Proof.** Let $S_N(k) := \sum_{n=1}^{N} a_{kn}$, then by the MCT (Proposition 1.8),

$$\lim_{N\to\infty} \sum_{k=1}^{\infty} S_N(k) = \sum_{k=1}^{\infty} \lim_{N\to\infty} S_N(k) = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn}.$$

On the other hand,

$$\sum_{k=1}^{\infty} S_N(k) = \sum_{k=1}^{\infty} \sum_{n=1}^{N} a_{kn} = \sum_{n=1}^{N} \sum_{k=1}^{\infty} a_{kn}$$

so that

$$\lim_{N\to\infty} \sum_{k=1}^{\infty} S_N(k) = \lim_{N\to\infty} \sum_{n=1}^{N} \sum_{k=1}^{\infty} a_{kn} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn}.$$

**Second Proof.** Let

$$M := \sup\left\{\sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn} : K, N \in \mathbb{N}\right\} = \sup\left\{\sum_{n=1}^{N} \sum_{k=1}^{K} a_{kn} : K, N \in \mathbb{N}\right\}$$

and

$$L := \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn}.$$

Since

$$L = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = \lim_{K\to\infty} \sum_{k=1}^{K} \sum_{n=1}^{\infty} a_{kn} = \lim_{K\to\infty} \lim_{N\to\infty} \sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn}$$

and $\sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn} \leq M$ for all $K$ and $N$, it follows that $L \leq M$. Conversely,

$$\sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn} \leq \sum_{k=1}^{K} \sum_{n=1}^{\infty} a_{kn} \leq \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = L$$

and therefore taking the supremum of the left side of this inequality over $K$ and $N$ shows that $M \leq L$. Thus we have shown

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = M.$$

By symmetry (or by a similar argument), we also have that $\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} a_{kn} = M$ and hence the proof is complete. ■

**Definition 1.10.** *A sequence* $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ *is* ***summable*** *(absolutely convergent) if* $\sum_{n=1}^{\infty} |a_n| < \infty$. *When* $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ *is summable we let* $a_n^{\pm} = \max(\mp a_n, 0)$ *and define,*

$$\sum_{n=1}^{\infty} a_n := \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-. \tag{1.7}$$

*Remark 1.11.* From Eq. (1.7) it follows that

$$\left|\sum_{n=1}^{\infty} a_n\right| \le \sum_{n=1}^{\infty} a_n^+ + \sum_{n=1}^{\infty} a_n^- = \sum_{n=1}^{\infty} \left(a_n^+ + a_n^-\right) = \sum_{n=1}^{\infty} |a_n|.$$

**Proposition 1.12 (Linearity).** *If* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *are summable* $\lambda \in \mathbb{R}$, *then* $\{a_n + \lambda b_n\}_{n=1}^{\infty}$ *is summable and*

$$\sum_{n=1}^{\infty} (a_n + \lambda b_n) = \sum_{n=1}^{\infty} a_n + \lambda \sum_{n=1}^{\infty} b_n.$$

**Proof.** Let $c_n := a_n + \lambda b_n$ so that $|c_n| \le |a_n| + |\lambda| |b_n|$ and hence

$$\sum_{n=1}^{\infty} |c_n| = \sum_{n=1}^{\infty} |a_n + \lambda b_n| \le \sum_{n=1}^{\infty} |a_n| + |\lambda| \sum_{n=1}^{\infty} |b_n| < \infty.$$

This shows $\{c_n\}_{n=1}^{\infty}$ is summable. Let us now suppose that $\lambda \ge 0$ for the moment. in which case we have

$$c_n = c_n^+ - c_n^- = a_n + \lambda b_n = a_n^+ - a_n^- + \lambda b_n^+ - \lambda b_n^-$$

and therefore,

$$c_n^+ + a_n^- + \lambda b_n^- = c_n^- + a_n^+ + \lambda b_n^+.$$

Summing this equation on $n$ while making use of Remark 1.7, then shows,

$$\begin{aligned}\sum_{n=1}^{\infty} c_n^+ + \sum_{n=1}^{\infty} a_n^- + \lambda \sum_{n=1}^{\infty} b_n^- \\ &= \sum_{n=1}^{\infty} \left(c_n^+ + a_n^- + \lambda b_n^-\right) = \sum_{n=1}^{\infty} \left(c_n^- + a_n^+ + \lambda b_n^+\right) \\ &= \sum_{n=1}^{\infty} c_n^- + \sum_{n=1}^{\infty} a_n^+ + \lambda \sum_{n=1}^{\infty} b_n^+.\end{aligned}$$

Rearranging these terms gives,

$$\begin{aligned}\sum_{n=1}^{\infty} c_n &= \sum_{n=1}^{\infty} c_n^+ - \sum_{n=1}^{\infty} c_n^- = \sum_{n=1}^{\infty} a_n^+ + \lambda \sum_{n=1}^{\infty} b_n^+ - \sum_{n=1}^{\infty} a_n^- - \lambda \sum_{n=1}^{\infty} b_n^- \\ &= \sum_{n=1}^{\infty} a_n + \lambda \sum_{n=1}^{\infty} b_n.\end{aligned}$$

To finish the proof we need only observe that $(-1 \cdot a_n)^{\pm} = a_n^{\mp}$ and hence

$$\sum_{n=1}^{\infty} -1 \cdot a_n = \sum_{n=1}^{\infty} a_n^- - \sum_{n=1}^{\infty} a_n^+ = -\left(\sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-\right) = -\sum_{n=1}^{\infty} a_n.$$

■

You are asked to prove the next three results in the exercises.

**Proposition 1.13 (Fubini for sums).** *Suppose* $\{a_{kn}\}_{k,n=1}^{\infty} \subset \mathbb{R}$ *such that*

$$\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} |a_{kn}| = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} |a_{kn}| < \infty.$$

*Then*

$$\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} a_{kn} = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} a_{kn}.$$

*Example 1.14 (Counter example).* Let $\{S_{mn}\}_{m,n=1}^{\infty}$ be any sequence of complex numbers such that $\lim_{m\to\infty} S_{mn} = 1$ for all $n$ and $\lim_{n\to\infty} S_{mn} = 0$ for all $n$. For example, take $S_{mn} = 1_{m\ge n} + \frac{1}{n} 1_{m<n}$. Then define $\{a_{ij}\}_{i,j=1}^{\infty}$ so that

$$S_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}.$$

Then

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} = \lim_{m\to\infty}\lim_{n\to\infty} S_{mn} = 0 \ne 1 = \lim_{n\to\infty}\lim_{m\to\infty} S_{mn} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{ij}.$$

To find $a_{ij}$, set $S_{mn} = 0$ if $m = 0$ or $n = 0$, then

$$S_{mn} - S_{m-1,n} = \sum_{j=1}^{n} a_{mj}$$

and

$$\begin{aligned}a_{mn} &= S_{mn} - S_{m-1,n} - (S_{m,n-1} - S_{m-1,n-1}) \\ &= S_{mn} - S_{m-1,n} - S_{m,n-1} + S_{m-1,n-1}.\end{aligned}$$

**Proposition 1.15 (Fatou's Lemma for sums).** *Suppose that for each $n \in \mathbb{N}$, $\{h_n(i)\}_{i=1}^\infty$ is any sequence in $[0,\infty]$, then*

$$\sum_{i=1}^\infty \liminf_{n\to\infty} h_n(i) \le \liminf_{n\to\infty} \sum_{i=1}^\infty h_n(i).$$

The next proposition is referred to as the dominated convergence theorem (DCT for short) for sums.

**Proposition 1.16 (DCT for sums).** *Suppose that for each $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^\infty \subset \mathbb{R}$ is a sequence and $\{g_n(i)\}_{i=1}^\infty$ is a sequence in $[0,\infty)$ such that;*

1. *$\sum_{i=1}^\infty g_n(i) < \infty$ for all $n$,*
2. *$f(i) = \lim_{n\to\infty} f_n(i)$ and $g(i) := \lim_{n\to\infty} g_n(i)$ exists for each $i$,*
3. *$|f_n(i)| \le g_n(i)$ for all $i$ and $n$,*
4. *$\lim_{n\to\infty} \sum_{i=1}^\infty g_n(i) = \sum_{i=1}^\infty g(i) < \infty$.*

*Then*

$$\lim_{n\to\infty} \sum_{i=1}^\infty f_n(i) = \sum_{i=1}^\infty \lim_{n\to\infty} f_n(i) = \sum_{i=1}^\infty f(i).$$

*(Often this proposition is used in the special case where $g_n = g$ for all $n$.)*

**Exercise 1.1 (Prove Proposition 1.13).** Suppose $\{a_{kn}\}_{k,n=1}^\infty \subset \mathbb{R}$ such that

$$\sum_{k=1}^\infty \sum_{n=1}^\infty |a_{kn}| = \sum_{n=1}^\infty \sum_{k=1}^\infty |a_{kn}| < \infty.$$

Then

$$\sum_{k=1}^\infty \sum_{n=1}^\infty a_{kn} = \sum_{n=1}^\infty \sum_{k=1}^\infty a_{kn}.$$

**Hint:** Let $a_{kn}^+ := \max(a_{kn}, 0)$ and $a_{kn}^- = \max(-a_{kn}, 0)$ and observe that; $a_{kn} = a_{kn}^+ - a_{kn}^-$ and $|a_{kn}^+| + |a_{kn}^-| = |a_{kn}|$. Now apply Proposition 1.9 with $a_{kn}$ replaced by $a_{kn}^+$ and $a_{kn}^-$. You should be careful to verify that $S_k := \sum_{n=1}^\infty a_{kn}$ exists and $\{S_k\}_{k=1}^\infty$ is summable so that $\sum_{k=1}^\infty \sum_{n=1}^\infty a_{kn} = \sum_{k=1}^\infty S_k$ exits, etc. etc.

**Exercise 1.2 (Prove Proposition 1.15).** Suppose that for each $n \in \mathbb{N}$, $\{h_n(i)\}_{i=1}^\infty$ is any sequence in $[0,\infty]$, then

$$\sum_{i=1}^\infty \liminf_{n\to\infty} h_n(i) \le \liminf_{n\to\infty} \sum_{i=1}^\infty h_n(i).$$

**Hint:** apply the MCT by applying the monotone convergence theorem with $f_n(i) := \inf_{m\ge n} h_m(i)$.

**Exercise 1.3 (Prove Proposition 1.16).** Suppose that for each $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^\infty \subset \mathbb{R}$ is a sequence and $\{g_n(i)\}_{i=1}^\infty$ is a sequence in $[0,\infty)$ such that;

1. $\sum_{i=1}^\infty g_n(i) < \infty$ for all $n$,
2. $f(i) = \lim_{n\to\infty} f_n(i)$ and $g(i) := \lim_{n\to\infty} g_n(i)$ exists for each $i$,
3. $|f_n(i)| \le g_n(i)$ for all $i$ and $n$,
4. $\lim_{n\to\infty} \sum_{i=1}^\infty g_n(i) = \sum_{i=1}^\infty g(i) < \infty$.

Then

$$\lim_{n\to\infty} \sum_{i=1}^\infty f_n(i) = \sum_{i=1}^\infty \lim_{n\to\infty} f_n(i) = \sum_{i=1}^\infty f(i).$$

**Hint:** Apply Fatou's lemma twice. Once with $h_n(i) = g_n(i) + f_n(i)$ and once with $h_n(i) = g_n(i) - f_n(i)$.

# Basic Metric and Topological Space Notions

The reader may refer to this chapter when the need arises later.

**Definition 2.1 (Pseudo-Metrics).** *Let $X$ be a non-empty set. A function $d : X \times X \to [0, \infty)$ is called a* ***pseudo-metric*** *on $X$ if $d$ is symmetric and satisfies the triangle inequality, i.e.*

1. *(**Symmetry**) $d(x, y) = d(y, x)$ for all $x, y \in X$, and*
2. *(**Triangle inequality**) $d(x, z) \le d(x, y) + d(y, z)$ for all $x, y, z \in X$.*

*If we further assume that $d$ is* ***non-degenerate*** *in the sense that $d(x, y) = 0$ if and only if $x = y \in X$, then we say $d$ is a* ***metric*** *on $X$.*

Notice that any subset, $Y$, of a (pseudo) metric space $(X, d)$ is a metric space by simply restricting $d$ to $Y \times Y$.

*Example 2.2.* Let us mention just a very few examples of (pseudo-metric) spaces.

1. Let $X = \mathbb{R}$. Then $d(x, y) = |y - x|$ is the usual metric on $\mathbb{R}$. Another useful metric is $d(x, y) = \left|\tan^{-1}(y) - \tan^{-1}(x)\right|$.
2. If $X = \mathbb{R}^d$, then $d(x, y) = \sqrt{\sum_{j=1}^d (y_j - x_j)^2}$ is the usual Euclidean distance metric on $\mathbb{R}^d$. Subsets like the unit sphere in $\mathbb{R}^d$ are metric spaces as well.
3. Let $X = C([0,1], \mathbb{C})$ be the continuous function the $d(f, g) := \max_{x\in[0,1]} |f(x) - g(x)|$ is a metric while $d(f, g) := \max_{x\in[0,1/2]} |f(x) - g(x)|$ is a Pseudo - metric on $X$.
4. Any normed space $(X, \|\cdot\|)$ (see Definition **??**) is a metric space with $d(x, y) := \|x - y\|$. Thus the space $\ell^p(\mu)$ (as in Theorem **??**) is a metric space for all $p \in [1, \infty]$.
5. Let $X$ denote the $C^1$ – periodic fucntions on $\mathbb{R}$. Then $d(f, g) := \max_{x\in\mathbb{R}} |f'(x) - g'(x)|$ is a pseudo-metric on $X$.

Througout this chapter, let $(X, d)$ be a pseudo-metric space and we will often just say $(X, d)$ is a metric space even though we may allow $d$ to be degenerate unless explicitly noted.

**Definition 2.3.** *Let $(X, d)$ be a metric space. The* ***open ball*** *$B(x, \delta) \subset X$ centered at $x \in X$ with radius $\delta > 0$ is the set*

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

*We will often also write $B(x, \delta)$ as $B_x(\delta)$. We also define the* ***closed ball*** *centered at $x \in X$ with radius $\delta > 0$ as the set $C_x(\delta) := \{y \in X : d(x, y) \le \delta\}$.*

**Definition 2.4.** *A sequence $\{x_n\}_{n=1}^\infty \subset X$ is said to* ***converge*** *to a point $x \in X$ if $\lim_{n\to\infty} d(x, x_n) = 0$ and abbreviate this by writing $x_n \to x$ or $x_n \overset{d}{\to} x$ as $n \to \infty$.*

If $x_n \to x$ and $x_n \to y$, then

$$d(x, y) \le d(x, x_n) + d(x_n, y) \to 0 \text{ as } n \to \infty$$

and so $d(x, y) = 0$. If $d$ is non-degenerate, then $x = y$ and limits are unique otherwise they are not.

**Definition 2.5.** *A set $E \subset X$ is* ***bounded*** *if $E \subset B(x, R)$ for some $x \in X$ and $R < \infty$. A set $F \subset X$ is* ***closed*** *iff every convergent sequence $\{x_n\}_{n=1}^\infty$ which is contained in $F$ has its limits back in $F$.*[1] *A set $V \subset X$ is* ***open*** *iff $V^c$ is closed. We will write $F \sqsubset X$ to indicate $F$ is a closed subset of $X$ and $V \subset_o X$ to indicate the $V$ is an open subset of $X$. We also let $\tau_d$ denote the collection of open subsets of $X$ relative to the metric $d$.*

**Exercise 2.1.** Let $\mathcal{F}$ be a collection of closed subsets of $X$, show $\cap\mathcal{F} := \cap_{F\in\mathcal{F}} F$ is closed. Also show that finite unions of closed sets are closed, i.e. if $\{F_k\}_{k=1}^K$ are closed sets then $\cup_{k=1}^K F_k$ is closed. (By taking complements, this shows that the collection of open sets, $\tau_d$, is closed under finite intersections and arbitrary unions.) Show by example that a countable union of closed sets need not be closed.

**Exercise 2.2.** Show that $V \subset X$ is open iff for every $x \in V$ there is a $\delta > 0$ such that $B_x(\delta) \subset V$. In particular show $B_x(\delta)$ is open for all $x \in X$ and $\delta > 0$. **Hint:** by definition $V$ is not open iff $V^c$ is not closed.

**Definition 2.6.** *A subset $A \subset X$ is a* ***neighborhood*** *of $x$ if there exists an open set $V \subset_o X$ such that $x \in V \subset A$. We will say that $A \subset X$ is an* ***open neighborhood*** *of $x$ if $A$ is open and $x \in A$.*

[1] When $d$ is non-degenerate we require all the possible limits of $\{x_n\}$ to be in $F$. This then implies that if $x \in F$ and $y \in X$ with $d(x, y) = 0$, then $y \in F$ as well.

The following "continuity" facts of the metric $d$ will be used frequently in the remainder of this book.

**Lemma 2.7.** *For any non empty subset $A \subset X$, let $d_A(x) := \inf\{d(x,a)|a \in A\}$, then*

$$|d_A(x) - d_A(y)| \le d(x,y) \;\forall\, x,y \in X \tag{2.1}$$

*and in particular if $x_n \to x$ in $X$ then $d_A(x_n) \to d_A(x)$ as $n \to \infty$. Moreover the set $F_\varepsilon := \{x \in X : d_A(x) \ge \varepsilon\}$ is closed in $X$.*

**Proof.** Let $a \in A$ and $x, y \in X$, then

$$d_A(x) \le d(x,a) \le d(x,y) + d(y,a).$$

Take the infimum over $a$ in the above equation shows that

$$d_A(x) \le d(x,y) + d_A(y) \quad \forall x,y \in X.$$

Therefore, $d_A(x) - d_A(y) \le d(x,y)$ and by interchanging $x$ and $y$ we also have that $d_A(y) - d_A(x) \le d(x,y)$ which implies Eq. (2.1). If $x_n \to x \in X$, then by Eq. (2.1),

$$|d_A(x) - d_A(x_n)| \le d(x,x_n) \to 0 \text{ as } n \to \infty$$

so that $\lim_{n\to\infty} d_A(x_n) = d_A(x)$. Now suppose that $\{x_n\}_{n=1}^\infty \subset F_\varepsilon$ and $x_n \to x$ in $X$, then

$$d_A(x) = \lim_{n\to\infty} d_A(x_n) \ge \varepsilon$$

since $d_A(x_n) \ge \varepsilon$ for all $n$. This shows that $x \in F_\varepsilon$ and hence $F_\varepsilon$ is closed. ■

**Corollary 2.8.** *The function $d$ satisfies,*

$$|d(x,y) - d(x',y')| \le d(y,y') + d(x,x').$$

*In particular $d : X \times X \to [0,\infty)$ is "continuous" in the sense that $d(x,y)$ is close to $d(x',y')$ if $x$ is close to $x'$ and $y$ is close to $y'$. (The notion of continuity will be developed shortly.)*

**Proof.** By Lemma 2.7 for single point sets and the triangle inequality for the absolute value of real numbers,

$$\begin{aligned}|d(x,y) - d(x',y')| &\le |d(x,y) - d(x,y')| + |d(x,y') - d(x',y')| \\ &\le d(y,y') + d(x,x').\end{aligned}$$

■

*Example 2.9.* Let $x \in X$ and $\delta > 0$, then $C_x(\delta)$ and $B_x(\delta)^c$ are closed subsets of $X$. For example if $\{y_n\}_{n=1}^\infty \subset C_x(\delta)$ and $y_n \to y \in X$, then $d(y_n, x) \le \delta$ for all $n$ and using Corollary 2.8 it follows $d(y,x) \le \delta$, i.e. $y \in C_x(\delta)$. A similar proof shows $B_x(\delta)^c$ is closed, see Exercise 2.2.

**Lemma 2.10 (Approximating open sets from the inside by closed sets).** *Let $A$ be a closed subset of $X$ and $F_\varepsilon := \{x \in X | d_A(x) \ge \varepsilon\} \sqsubset X$ be as in Lemma 2.7. Then $F_\varepsilon \uparrow A^c$ as $\varepsilon \downarrow 0$.*

**Proof.** It is clear that $d_A(x) = 0$ for $x \in A$ so that $F_\varepsilon \subset A^c$ for each $\varepsilon > 0$ and hence $\cup_{\varepsilon>0} F_\varepsilon \subset A^c$. Now suppose that $x \in A^c \subset_o X$. By Exercise 2.2 there exists an $\varepsilon > 0$ such that $B_x(\varepsilon) \subset A^c$, i.e. $d(x,y) \ge \varepsilon$ for all $y \in A$. Hence $x \in F_\varepsilon$ and we have shown that $A^c \subset \cup_{\varepsilon>0} F_\varepsilon$. Finally it is clear that $F_\varepsilon \subset F_{\varepsilon'}$ whenever $\varepsilon' \le \varepsilon$. ■

**Definition 2.11.** *Given a set $A$ contained in a metric space $X$, let $\bar{A} \subset X$ be the **closure of** $A$ defined by*

$$\bar{A} := \{x \in X : \exists\, \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\}.$$

*That is to say $\bar{A}$ contains all **limit points** of $A$. We say $A$ is **dense in** $X$ if $\bar{A} = X$, i.e. every element $x \in X$ is a limit of a sequence of elements from $A$. A metric space is said to be **separable** if it contains a countable dense subset, $D$.*

**Exercise 2.3.** Given $A \subset X$, show $\bar{A}$ is a closed set and in fact

$$\bar{A} = \cap\{F : A \subset F \subset X \text{ with } F \text{ closed}\}. \tag{2.2}$$

That is to say $\bar{A}$ is the smallest closed set containing $A$.

**Exercise 2.4.** If $D$ is a dense subset of a metric space $(X,d)$ and $E \subset X$ is a subset such that to every point $x \in D$ there exists $\{x_n\}_{n=1}^\infty \subset E$ with $x = \lim_{n\to\infty} x_n$, then $E$ is also a dense subset of $X$. If points in $E$ well approximate every point in $D$ and the points in $D$ well approximate the points in $X$, then the points in $E$ also well approximate all points in $X$.

**Exercise 2.5.** Suppose $(X,d)$ is a metric space which contains an uncountable subset $\Lambda \subset X$ with the property that there exists $\varepsilon > 0$ such that $d(a,b) \ge \varepsilon$ for all $a, b \in \Lambda$ with $a \ne b$. Show that $(X,d)$ is **not** separable.

## 2.1 Metric spaces as topological spaces

Let $(X,d)$ be a metric space and let $\tau = \tau_d$ denote the collection of open subsets of $X$. (Recall $V \subset X$ is open iff $V^c$ is closed iff for all $x \in V$ there exists an $\varepsilon = \varepsilon_x > 0$ such that $B(x, \varepsilon_x) \subset V$ iff $V$ can be written as a (possibly uncountable) union of open balls.) Although we will stick with metric spaces in this chapter, it will be useful to introduce the definitions needed here in the more general context of a general "topological space," i.e. a space equipped with a collection of "open sets."

**Definition 2.12 (Topological Space).** *Let $X$ be a set. A **topology** on $X$ is a collection of subsets ($\tau$) of $X$ with the following properties;*

1. *$\tau$ contains both the empty set ($\emptyset$) and $X$.*
2. *$\tau$ is closed under arbitrary unions.*
3. *$\tau$ is closed under finite intersections.*

*The elements $V \in \tau$ are called **open** subsets of $X$. A subset $F \subset X$ is said to be **closed** if $F^c$ is open. I will write $V \subset_o X$ to indicate that $V \subset X$ and $V \in \tau$ and similarly $F \sqsubset X$ will denote $F \subset X$ and $F$ is closed. Given $x \in X$ we say that $V \subset X$ is an **open neighborhood** of $x$ if $V \in \tau$ and $x \in V$. Let $\tau_x = \{V \in \tau : x \in V\}$ denote the collection of open neighborhoods of $x$.*

Of course every metric space $(X, d)$ is also a topological space where we take $\tau = \tau_d$.

**Definition 2.13.** *Let $(X, \tau)$ be a topological space and $A$ be a subset of $X$.*

1. *The **closure** of $A$ is the smallest closed set $\bar{A}$ containing $A$, i.e.*
$$\bar{A} := \cap \{F : A \subset F \sqsubset X\}.$$
*(Because of Exercise 2.3 this is consistent with Definition 2.11 for the closure of a set in a metric space.)*
2. *The **interior** of $A$ is the largest open set $A^o$ contained in $A$, i.e.*
$$A^o = \cup \{V \in \tau : V \subset A\}.$$
3. *$A \subset X$ is a **neighborhood of a point** $x \in X$ if $x \in A^o$.*
4. *The **accumulation points** of $A$ is the set*
$$\mathrm{acc}(A) = \{x \in X : V \cap [A \setminus \{x\}] \neq \emptyset \text{ for all } V \in \tau_x\}.$$
5. *The **boundary** of $A$ is the set $\mathrm{bd}(A) := \bar{A} \setminus A^o$.*
6. *$A$ is **dense** in $X$ if $\bar{A} = X$ and $X$ is said to be **separable** if there exists a countable dense subset of $X$.*

*Remark 2.14.* The relationships between the interior and the closure of a set are:

$$(A^o)^c = \bigcap \{V^c : V \in \tau \text{ and } V \subset A\} = \bigcap \{C : C \text{ is closed } C \supset A^c\} = \overline{A^c}$$

and similarly, $(\bar{A})^c = (A^c)^o$. Hence the boundary of $A$ may be written as

$$\mathrm{bd}(A) := \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = \bar{A} \cap \overline{A^c}, \tag{2.3}$$

which is to say $\mathrm{bd}(A)$ consists of the points in both the closures of $A$ and $A^c$.

### 2.1.1 Continuity

Suppose now that $(X, \rho)$ and $(Y, d)$ are two metric spaces and $f : X \to Y$ is a function.

**Definition 2.15.** *A function $f : X \to Y$ is **continuous at** $x \in X$ if for all $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$d(f(x), f(x')) < \varepsilon \text{ provided that } \rho(x, x') < \delta. \tag{2.4}$$

*The function $f$ is said to be **continuous** if $f$ is continuous at all points $x \in X$.*

The following lemma gives two other characterizations of continuity of a function at a point.

**Lemma 2.16 (Local Continuity Lemma).** *Suppose that $(X, \rho)$ and $(Y, d)$ are two metric spaces and $f : X \to Y$ is a function defined in a neighborhood of a point $x \in X$. Then the following are equivalent:*

1. *$f$ is continuous at $x \in X$.*
2. *For all neighborhoods $A \subset Y$ of $f(x)$, $f^{-1}(A)$ is a neighborhood of $x \in X$.*
3. *For all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x = \lim_{n\to\infty} x_n$, $\{f(x_n)\}$ is convergent in $Y$ and*
$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

**Proof.** 1 $\implies$ 2. If $A \subset Y$ is a neighborhood of $f(x)$, there exists $\varepsilon > 0$ such that $B_{f(x)}(\varepsilon) \subset A$ and because $f$ is continuous there exists a $\delta > 0$ such that Eq. (2.4) holds. Therefore

$$B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right) \subset f^{-1}(A)$$

showing $f^{-1}(A)$ is a neighborhood of $x$.

2 $\implies$ 3. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ and $x = \lim_{n\to\infty} x_n$. Then for any $\varepsilon > 0$, $B_{f(x)}(\varepsilon)$ is a neighborhood of $f(x)$ and so $f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ is a neighborhood of $x$ which must contain $B_x(\delta)$ for some $\delta > 0$. Because $x_n \to x$, it follows that $x_n \in B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ for a.a. $n$ and this implies $f(x_n) \in B_{f(x)}(\varepsilon)$ for a.a. $n$, i.e. $d(f(x), f(x_n)) < \varepsilon$ for a.a. $n$. Since $\varepsilon > 0$ is arbitrary it follows that $\lim_{n\to\infty} f(x_n) = f(x)$.

3. $\implies$ 1. We will show not 1. $\implies$ not 3. If $f$ is not continuous at $x$, there exists an $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a point $x_n \in X$ with $\rho(x_n, x) < \frac{1}{n}$ yet $d(f(x_n), f(x)) \geq \varepsilon$. Hence $x_n \to x$ as $n \to \infty$ yet $f(x_n)$ does not converge to $f(x)$. ■

Here is a global version of the previous lemma.

**Lemma 2.17 (Global Continuity Lemma).** *Suppose that $(X,\rho)$ and $(Y,d)$ are two metric spaces and $f: X \to Y$ is a function defined on all of $X$. Then the following are equivalent:*

1. *$f$ is continuous.*
2. *$f^{-1}(V) \in \tau_\rho$ for all $V \in \tau_d$, i.e. $f^{-1}(V)$ is open in $X$ if $V$ is open in $Y$.*
3. *$f^{-1}(C)$ is closed in $X$ if $C$ is closed in $Y$.*
4. *For all convergent sequences $\{x_n\} \subset X$, $\{f(x_n)\}$ is convergent in $Y$ and*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

**Proof.** Since $f^{-1}(A^c) = \left[f^{-1}(A)\right]^c$, it is easily seen that 2. and 3. are equivalent. So because of Lemma 2.16 it only remains to show 1. and 2. are equivalent. If $f$ is continuous and $V \subset Y$ is open, then for every $x \in f^{-1}(V)$, $V$ is a neighborhood of $f(x)$ and so $f^{-1}(V)$ is a neighborhood of $x$. Hence $f^{-1}(V)$ is a neighborhood of all of its points and from this and Exercise 2.2 it follows that $f^{-1}(V)$ is open. Conversely, if $x \in X$ and $A \subset Y$ is a neighborhood of $f(x)$ then there exists $V \subset_o X$ such that $f(x) \in V \subset A$. Hence $x \in f^{-1}(V) \subset f^{-1}(A)$ and by assumption $f^{-1}(V)$ is open showing $f^{-1}(A)$ is a neighborhood of $x$. Therefore $f$ is continuous at $x$ and since $x \in X$ was arbitrary, $f$ is continuous. ■

**Definition 2.18 (Continuity at a point in topological terms).** *Let $(X,\tau_X)$ and $(Y,\tau_Y)$ be topological spaces. A function $f: X \to Y$ is **continuous at a point** $x \in X$ if for every open neighborhood $V$ of $f(x)$ there is an open neighborhood $U$ of $x$ such that $U \subset f^{-1}(V)$. See Figure 2.1.*

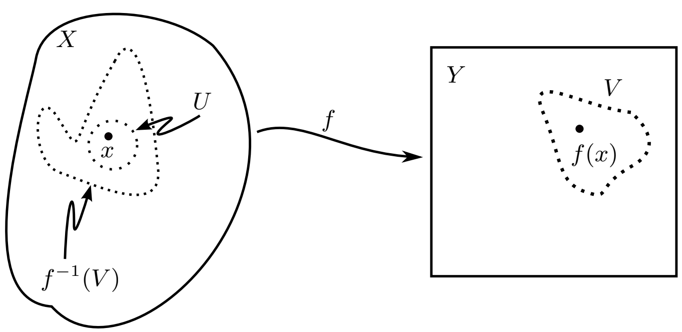



**Fig. 2.1.** Checking that a function is continuous at $x \in X$.

**Definition 2.19 (Global continuity in topological terms).** *Let $(X,\tau_X)$ and $(Y,\tau_Y)$ be topological spaces. A function $f: X \to Y$ is **continuous** if*

$$f^{-1}(\tau_Y) := \left\{f^{-1}(V) : V \in \tau_Y\right\} \subset \tau_X.$$

*We will also say that $f$ is $\tau_X/\tau_Y$ –continuous or $(\tau_X,\tau_Y)$ – continuous. Let $C(X,Y)$ denote the set of continuous functions from $X$ to $Y$.*

**Exercise 2.6.** Show $f: X \to Y$ is continuous (Definition 2.19) iff $f$ is continuous at all points $x \in X$.

**Exercise 2.7.** Show $f: X \to Y$ is continuous iff $f^{-1}(C)$ is closed in $X$ for all closed subsets $C$ of $Y$.

**Definition 2.20.** *A map $f: X \to Y$ between topological spaces is called a **homeomorphism** provided that $f$ is bijective, $f$ is continuous and $f^{-1}: Y \to X$ is continuous. If there exists $f: X \to Y$ which is a homeomorphism, we say that $X$ and $Y$ are homeomorphic. (As topological spaces $X$ and $Y$ are essentially the same.)*

*Example 2.21.* The function $d_A$ defined in Lemma 2.7 is continuous for each $A \subset X$. In particular, if $A = \{x\}$, it follows that $y \in X \to d(y,x)$ is continuous for each $x \in X$.

**Exercise 2.8.** Use Example 2.21 and Lemma 2.17 to recover the results of Example 2.9.

**Exercise 2.9 (A joint continuity criteria).** Let $X, Y, Z$ be three metric spaces and $F: X \times Y \to Z$ be a function such that;

1. For each $x \in X$ the map $Y \ni y \to F(x,y) \in Z$ is continuous and moreover are locally equi-continuous in $x$. In more detail, assume for all $(a,b) \in X \times Y$ there exists $\kappa := \kappa(a,b) > 0$ such that

$$\lim_{y\to b} \sup_{x \in B_X(a,\varepsilon)} d(F(x,y), F(x,b)) = 0.$$

2. There exists a dense subset $Y_0 \subset Y$ such that $X \ni x \to F(x,y) \in Z$ is continuous for any fixed $y \in Y_0$.

Show;

1. $X \ni x \to F(x,y) \in Z$ is continuous for any fixed $y \in Y$ and then show
2. $F: X \times Y \to Z$ is jointly continuous on $X \times Y$.

**Lemma 2.22 (Urysohn's Lemma for Metric Spaces).** *Let $(X,d)$ be a metric space and suppose that $A$ and $B$ are two disjoint closed subsets of $X$. Then*

$$f(x)=\frac{d_B(x)}{d_A(x)+d_B(x)} \text{ for } x\in X \tag{2.5}$$

*defines a continuous function, $f:X\to[0,1]$, such that $f(x)=1$ for $x\in A$ and $f(x)=0$ if $x\in B$.*

**Proof.** By Lemma 2.7, $d_A$ and $d_B$ are continuous functions on $X$. Since $A$ and $B$ are closed, $d_A(x)>0$ if $x\notin A$ and $d_B(x)>0$ if $x\notin B$. Since $A\cap B=\emptyset$, $d_A(x)+d_B(x)>0$ for all $x$ and $(d_A+d_B)^{-1}$ is continuous as well. The remaining assertions about $f$ are all easy to verify. ■

Sometimes Urysohn's lemma will be use in the following form. Suppose $F\subset V\subset X$ with $F$ being closed and $V$ being open, then there exists $f\in C(X,[0,1]))$ such that $f=1$ on $F$ while $f=0$ on $V^c$. This of course follows from Lemma 2.22 by taking $A=F$ and $B=V^c$.

**Corollary 2.23.** *If $A$ and $B$ are two disjoint closed subsets of $X$, then there exists disjoint open subsets $U$ and $V$ of $X$ such that $A\subset U$ and $B\subset V$.*

**Proof.** Let $f$ be as in Lemma 2.22 so that $f\in C(X\to[0,1])$ such that $f=1$ on $A$ and $f=0$ on $B$. Then set $U=\left\{f>\frac{1}{2}\right\}$ and $V=\{f<1/2\}$. ■

## 2.2 Completeness in Metric Spaces

**Definition 2.24 (Cauchy sequences).** *A sequence $\{x_n\}_{n=1}^\infty$ in a metric space $(X,d)$ is* ***Cauchy*** *provided that*

$$\lim_{m,n\to\infty} d(x_n,x_m)=0.$$

**Exercise 2.10.** Let $(X,d)$ be a pseudo metric space.

1. Show every convergent sequence, $\{x_n\}_{n=1}^\infty\subset X$, is Cauchy.
2. If $x_n\to x$ and $y_n\to y$ show $d(x_n,y_n)\to d(x,y)$.
3. If $\{x_n\}_{n=1}^\infty,\{y_n\}_{n=1}^\infty\subset X$ are Cauchy sequences, show $\lim_{n\to\infty} d(x_n,y_n)$ exists in $[0,\infty)$.

As you showed in Exercise 2.10, convergent sequences are always Cauchy sequences. The converse is not always true. For example, let $X=\mathbb{Q}$ be the set of rational numbers and $d(x,y)=|x-y|$. Choose a sequence $\{x_n\}_{n=1}^\infty\subset\mathbb{Q}$ which converges to $\sqrt{2}\in\mathbb{R}$, then $\{x_n\}_{n=1}^\infty$ is $(\mathbb{Q},d)$ – Cauchy but not $(\mathbb{Q},d)$ – convergent. The sequence does converge in $\mathbb{R}$ however.

**Definition 2.25.** *A metric space $(X,d)$ is* ***complete*** *if all Cauchy sequences are convergent sequences.*

**Exercise 2.11.** Let $(X,d)$ be a complete metric space. Let $A\subset X$ be a subset of $X$ viewed as a metric space using $d|_{A\times A}$. Show that $(A,d|_{A\times A})$ is complete iff $A$ is a closed subset of $X$.

*Example 2.26.* Examples 2. – 4. of complete metric spaces will be verified in Chapter **??** below.

1. $X=\mathbb{R}$ and $d(x,y)=|x-y|$, see Theorem **??** above.
2. $X=\mathbb{R}^n$ and $d(x,y)=\|x-y\|_2=\left(\sum_{i=1}^n(x_i-y_i)^2\right)^{1/2}$.
3. $X=\ell^p(\mu)$ for $p\in[1,\infty]$ and any weight function $\mu:X\to(0,\infty)$.
4. $X=C([0,1],\mathbb{R})$ – the space of continuous functions from $[0,1]$ to $\mathbb{R}$ and
$$d(f,g):=\max_{t\in[0,1]}|f(t)-g(t)|.$$
This is a special case of Lemma **??** below.
5. Let $X=C([0,1],\mathbb{R})$ and
$$d(f,g):=\int_0^1|f(t)-g(t)|\,dt.$$
You are asked in Exercise **??** to verify that $(X,d)$ is a metric space which is **not** complete.

**Exercise 2.12 (Completions of Metric Spaces).** Suppose that $(X,d)$ is a (not necessarily complete) metric space. Using the following outline show there exists a complete metric space $\left(\bar{X},\bar{d}\right)$ and an isometric map $i:X\to\bar{X}$ such that $i(X)$ is dense in $\bar{X}$, see Definition 2.11.

1. Let $\mathcal{C}$ denote the collection of Cauchy sequences $a=\{a_n\}_{n=1}^\infty\subset X$. Given two element $a,b\in\mathcal{C}$ show $d_{\mathcal{C}}(a,b):=\lim_{n\to\infty}d(a_n,b_n)$ exists, $d_{\mathcal{C}}(a,b)\geq 0$ for all $a,b\in\mathcal{C}$ and $d_{\mathcal{C}}$ satisfies the triangle inequality,
$$d_{\mathcal{C}}(a,c)\leq d_{\mathcal{C}}(a,b)+d_{\mathcal{C}}(b,c) \text{ for all } a,b,c\in\mathcal{C}.$$
Thus $(\mathcal{C},d_{\mathcal{C}})$ would be a metric space if it were true that $d_{\mathcal{C}}(a,b)=0$ iff $a=b$. This however is false, for example if $a_n=b_n$ for all $n\geq 100$, then $d_{\mathcal{C}}(a,b)=0$ while $a$ need not equal $b$.
2. Define two elements $a,b\in\mathcal{C}$ to be equivalent (write $a\sim b$) whenever $d_{\mathcal{C}}(a,b)=0$. Show " $\sim$ " is an equivalence relation on $\mathcal{C}$ and that $d_{\mathcal{C}}(a',b')=d_{\mathcal{C}}(a,b)$ if $a\sim a'$ and $b\sim b'$. (**Hint:** see Corollary 2.8.)

3. Given $a \in \mathcal{C}$ let $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ denote the equivalence class containing $a$ and let $\bar{X} := \{\bar{a} : a \in \mathcal{C}\}$ denote the collection of such equivalence classes. Show that $\bar{d}\left(\bar{a}, \bar{b}\right) := d_{\mathcal{C}}(a, b)$ is well defined on $\bar{X} \times \bar{X}$ and verify $\left(\bar{X}, \bar{d}\right)$ is a metric space.
4. For $x \in X$ let $i(x) = \bar{a}$ where $a$ is the constant sequence, $a_n = x$ for all $n$. Verify that $i : X \to \bar{X}$ is an isometric map and that $i(X)$ is dense in $\bar{X}$.
5. Verify $\left(\bar{X}, \bar{d}\right)$ is complete. **Hint:** if $\{\bar{a}(m)\}_{m=1}^{\infty}$ is a Cauchy sequence in $\bar{X}$ choose $b_m \in X$ such that $\bar{d}(i(b_m), \bar{a}(m)) \le 1/m$. Then show $\bar{a}(m) \to \bar{b}$ where $b = \{b_m\}_{m=1}^{\infty}$.

**Definition 2.27 (Lip-$K$ functions).** *Suppose that $(X, d)$ is a pseudo-metric space, $U$ be a non-empty subset of $X$, and $K > 0$. A function $f : U \to \mathbb{C}$ is **Lip**-$K$ on $U$ if*

$$|f(y) - f(x)| \le K d(x, y) \text{ for all } x, y \in U.$$

The basic fact about pseudo-metrics and Lip-$K$ - functions that we will use is contained in the next lemma.

**Lemma 2.28.** *Suppose that $(X, d)$ is a pseudo-metric space, $U$ is a non-empty subset of $X$, and $f : U \to \mathbb{C}$ is a Lip-$K$ function. Then there exists a unique Lip-$K$ function, $\hat{f} : \bar{U} \to \mathbb{C}$ such that $f = \hat{f}|_{\bar{U}}$. [See Exercise 2.18 for a generalization of this result.]*

**Exercise 2.13.** Prove Lemma 2.28. [It is useful to observe that every Lip-$K$ function on $U$ as above is continuous on $U$.]

## 2.3 Supplementary Remarks

### 2.3.1 Word of Caution

*Example 2.29.* Let $(X, d)$ be a metric space. It is always true that $\overline{B_x(\varepsilon)} \subset C_x(\varepsilon)$ since $C_x(\varepsilon)$ is a closed set containing $B_x(\varepsilon)$. However, it is not always true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$. For example let $X = \{1, 2\}$ and $d(1, 2) = 1$, then $B_1(1) = \{1\}$, $\overline{B_1(1)} = \{1\}$ while $C_1(1) = X$. For another counterexample, take

$$X = \left\{(x, y) \in \mathbb{R}^2 : x = 0 \text{ or } x = 1\right\}$$

with the usually Euclidean metric coming from the plane. Then

$$\begin{aligned} B_{(0,0)}(1) &= \left\{(0, y) \in \mathbb{R}^2 : |y| < 1\right\}, \\ \overline{B_{(0,0)}(1)} &= \left\{(0, y) \in \mathbb{R}^2 : |y| \le 1\right\}, \text{ while} \\ C_{(0,0)}(1) &= \overline{B_{(0,0)}(1)} \cup \{(1, 0)\}. \end{aligned}$$

In spite of the above examples, Lemmas 2.30 and 2.31 below shows that for certain metric spaces of interest it is true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$.

**Lemma 2.30.** *Suppose that $(X, |\cdot|)$ is a normed vector space and $d$ is the metric on $X$ defined by $d(x, y) = |x - y|$. Then*

$$\begin{aligned} \overline{B_x(\varepsilon)} &= C_x(\varepsilon) \text{ and} \\ \mathrm{bd}(B_x(\varepsilon)) &= \{y \in X : d(x, y) = \varepsilon\}. \end{aligned}$$

*where the boundary operation, $\mathrm{bd}(\cdot)$ is defined in Definition* **??** *(BRUCE: Forward Reference.) below.*

**Proof.** We must show that $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. For $y \in C$, let $v = y - x$, then

$$|v| = |y - x| = d(x, y) \le \varepsilon.$$

Let $\alpha_n = 1 - 1/n$ so that $\alpha_n \uparrow 1$ as $n \to \infty$. Let $y_n = x + \alpha_n v$, then $d(x, y_n) = \alpha_n d(x, y) < \varepsilon$, so that $y_n \in B_x(\varepsilon)$ and $d(y, y_n) = (1 - \alpha_n)|v| \to 0$ as $n \to \infty$. This shows that $y_n \to y$ as $n \to \infty$ and hence that $y \in \bar{B}$. ∎

### 2.3.2 Riemannian Metrics

This subsection is not completely self contained and may safely be skipped.

**Lemma 2.31.** *Suppose that $X$ is a Riemannian (or sub-Riemannian) manifold and $d$ is the metric on $X$ defined by*

$$d(x, y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}$$

*where $\ell(\sigma)$ is the length of the curve $\sigma$. We define $\ell(\sigma) = \infty$ if $\sigma$ is not piecewise smooth.*

*Then*

$$\begin{aligned} \overline{B_x(\varepsilon)} &= C_x(\varepsilon) \text{ and} \\ \mathrm{bd}(B_x(\varepsilon)) &= \{y \in X : d(x, y) = \varepsilon\} \end{aligned}$$

*where the boundary operation, $\mathrm{bd}(\cdot)$ is defined in Definition* **??** *below.*

**Proof.** Let $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. We will show that $C \subset \bar{B}$ by showing $\bar{B}^c \subset C^c$. Suppose that $y \in \bar{B}^c$ and choose $\delta > 0$ such that $B_y(\delta) \cap \bar{B} = \emptyset$. In particular this implies that

$$B_y(\delta) \cap B_x(\varepsilon) = \emptyset.$$

We will finish the proof by showing that $d(x, y) \ge \varepsilon + \delta > \varepsilon$ and hence that $y \in C^c$. This will be accomplished by showing: if $d(x, y) < \varepsilon + \delta$ then $B_y(\delta) \cap$

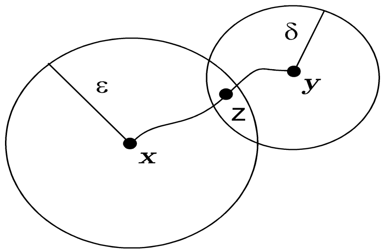



**Fig. 2.2.** An almost length minimizing curve joining $x$ to $y$.

$B_x(\varepsilon) \neq \emptyset$. If $d(x,y) < \max(\varepsilon, \delta)$ then either $x \in B_y(\delta)$ or $y \in B_x(\varepsilon)$. In either case $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. Hence we may assume that $\max(\varepsilon,\delta) \leq d(x,y) < \varepsilon + \delta$. Let $\alpha > 0$ be a number such that

$$\max(\varepsilon,\delta) \leq d(x,y) < \alpha < \varepsilon + \delta$$

and choose a curve $\sigma$ from $x$ to $y$ such that $\ell(\sigma) < \alpha$. Also choose $0 < \delta' < \delta$ such that $0 < \alpha - \delta' < \varepsilon$ which can be done since $\alpha - \delta < \varepsilon$. Let $k(t) = d(y, \sigma(t))$ a continuous function on $[0,1]$ and therefore $k([0,1]) \subset \mathbb{R}$ is a connected set which contains 0 and $d(x,y)$. Therefore there exists $t_0 \in [0,1]$ such that $d(y, \sigma(t_0)) = k(t_0) = \delta'$. Let $z = \sigma(t_0) \in B_y(\delta)$ then

$$d(x,z) \leq \ell(\sigma|_{[0,t_0]}) = \ell(\sigma) - \ell(\sigma|_{[t_0,1]}) < \alpha - d(z,y) = \alpha - \delta' < \varepsilon$$

and therefore $z \in B_x(\varepsilon) \cap B_x(\delta) \neq \emptyset$. ■

*Remark 2.32.* Suppose again that $X$ is a Riemannian (or sub-Riemannian) manifold and

$$d(x,y) = \inf \{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}.$$

Let $\sigma$ be a curve from $x$ to $y$ and let $\varepsilon = \ell(\sigma) - d(x,y)$. Then for all $0 \leq u < v \leq 1$,

$$\begin{aligned} d(x,y) + \varepsilon = \ell(\sigma) &= \ell(\sigma|_{[0,u]}) + \ell(\sigma|_{[u,v]}) + \ell(\sigma|_{[v,1]}) \\ &\geq d(x, \sigma(u)) + \ell(\sigma|_{[u,v]}) + d(\sigma(v), y) \end{aligned}$$

and therefore, using the triangle inequality,

$$\begin{aligned} \ell(\sigma|_{[u,v]}) &\leq d(x,y) + \varepsilon - d(x,\sigma(u)) - d(\sigma(v), y) \\ &\leq d(\sigma(u), \sigma(v)) + \varepsilon. \end{aligned}$$

This leads to the following conclusions. If $\sigma$ is within $\varepsilon$ of a length minimizing curve from $x$ to $y$ then $\sigma|_{[u,v]}$ is within $\varepsilon$ of a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. In particular if $\sigma$ is a length minimizing curve from $x$ to $y$ then $\sigma|_{[u,v]}$ is a length minimizing curve from $\sigma(u)$ to $\sigma(v)$.

## 2.4 Exercises

**Exercise 2.14.** Let $(X,d)$ be a metric space. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ is a sequence and set $\varepsilon_n := d(x_n, x_{n+1})$. Show that for $m > n$ that

$$d(x_n, x_m) \leq \sum_{k=n}^{m-1} \varepsilon_k \leq \sum_{k=n}^{\infty} \varepsilon_k.$$

Conclude from this that if

$$\sum_{k=1}^\infty \varepsilon_k = \sum_{n=1}^\infty d(x_n, x_{n+1}) < \infty$$

then $\{x_n\}_{n=1}^\infty$ is Cauchy. Moreover, show that if $\{x_n\}_{n=1}^\infty$ is a convergent sequence and $x = \lim_{n\to\infty} x_n$ then

$$d(x, x_n) \leq \sum_{k=n}^\infty \varepsilon_k.$$

**Exercise 2.15.** Show that $(X,d)$ is a complete metric space iff every sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\sum_{n=1}^\infty d(x_n, x_{n+1}) < \infty$ is a convergent sequence in $X$. You may find it useful to prove the following statements in the course of the proof.

1. If $\{x_n\}$ is Cauchy sequence, then there is a subsequence $y_j := x_{n_j}$ such that $\sum_{j=1}^\infty d(y_{j+1}, y_j) < \infty$.
2. If $\{x_n\}_{n=1}^\infty$ is Cauchy and there exists a subsequence $y_j := x_{n_j}$ of $\{x_n\}$ such that $x = \lim_{j\to\infty} y_j$ exists, then $\lim_{n\to\infty} x_n$ also exists and is equal to $x$.

**Exercise 2.16.** Suppose that $f : [0,\infty) \to [0,\infty)$ is a $C^2$ – function such that $f(0) = 0$, $f' > 0$ and $f'' \leq 0$ and $(X,\rho)$ is a metric space. Show that $d(x,y) = f(\rho(x,y))$ is a metric on $X$. In particular show that

$$d(x,y) := \frac{\rho(x,y)}{1 + \rho(x,y)}$$

is a metric on $X$. (Hint: use calculus to verify that $f(a+b) \leq f(a) + f(b)$ for all $a, b \in [0,\infty)$.)

**Exercise 2.17.** Let $\{(X_n, d_n)\}_{n=1}^\infty$ be a sequence of metric spaces, $X := \prod_{n=1}^\infty X_n$, and for $x = (x(n))_{n=1}^\infty$ and $y = (y(n))_{n=1}^\infty$ in $X$ let

$$d(x,y) = \sum_{n=1}^\infty 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}. \tag{2.6}$$

Show:

1. $(X, d)$ is a metric space,
2. a sequence $\{x_k\}_{k=1}^{\infty} \subset X$ converges to $x \in X$ iff $x_k(n) \to x(n) \in X_n$ as $k \to \infty$ for each $n \in \mathbb{N}$ and
3. $X$ is complete if $X_n$ is complete for all $n$.

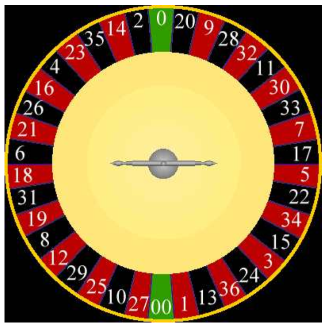
**Exercise 2.18.** Suppose $(X, \rho)$ and $(Y, d)$ are metric spaces and $A$ is a dense subset of $X$.

1. Show that if $F : X \to Y$ and $G : X \to Y$ are two continuous functions such that $F = G$ on $A$ then $F = G$ on $X$. **Hint:** consider the set $C := \{x \in X : F(x) = G(x)\}$.
2. Now suppose that $(Y, d)$ is complete. If $f : A \to Y$ is a function which is uniformly continuous (i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that
$$d(f(a), f(b)) < \varepsilon \text{ for all } a, b \in A \text{ with } \rho(a, b) < \delta),$$
show there is a unique continuous function $F : X \to Y$ such that $F = f$ on $A$. **Hint:** each point $x \in X$ is a limit of a sequence consisting of elements from $A$.
3. Let $X = \mathbb{R} = Y$ and $A = \mathbb{Q} \subset X$, find a function $f : \mathbb{Q} \to \mathbb{R}$ which is continuous on $\mathbb{Q}$ but does **not** extend to a continuous function on $\mathbb{R}$.

# 3

# Basic Probabilistic Notions

**Definition 3.1.** *A sample space $\Omega$ is a set which is to represents all possible outcomes of an "experiment."*




*Example 3.2.*
1. The sample space for flipping a coin one time could be taken to be, $\Omega = \{0,1\}$.
2. The sample space for flipping a coin $N$ -times could be taken to be, $\Omega = \{0,1\}^N$ and for flipping an infinite number of times,
$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0,1\}\} = \{0,1\}^{\mathbb{N}}.$$
3. If we have a roulette wheel with 38 entries, then we might take
$$\Omega = \{00, 0, 1, 2, \dots, 36\}$$
for one spin,
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^N$$
for $N$ spins, and
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^{\mathbb{N}}$$
for an infinite number of spins.
4. If we have a spinner (a board with an arrow attached to the board by a bearing at its base), we could take
$$\Omega = S^1 = \{z \in \mathbb{C} : |z| = 1\} = \left\{e^{i\theta} : -\pi \le \theta \le \pi\right\}.$$
5. If we throw darts at a board of radius $R$, we may take
$$\Omega = D_R := \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le R^2\right\}$$
for one throw,
$$\Omega = D_R^N$$
for $N$ throws, and
$$\Omega = D_R^{\mathbb{N}}$$
for an infinite number of throws.
6. Suppose we release a perfume particle at location $x \in \mathbb{R}^3$ and follow its motion for all time, $0 \le t < \infty$. In this case, we might take,
$$\Omega = \left\{\omega \in C\left([0,\infty), \mathbb{R}^3\right) : \omega(0) = x\right\}.$$

**Definition 3.3.** *An **event**, $A$, is a subset of $\Omega$. Given $A \subset \Omega$ we also define the **indicator function of** $A$ by*
$$1_A(\omega) := \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}.$$

*Example 3.4.* Suppose that $\Omega = \{0,1\}^{\mathbb{N}}$ is the sample space for flipping a coin an infinite number of times. Here $\omega_n = 1$ represents the fact that a head was thrown on the $n^{\text{th}}$ – toss, while $\omega_n = 0$ represents a tail on the $n^{\text{th}}$ – toss.

1. $A = \{\omega \in \Omega : \omega_3 = 1\}$ represents the event that the third toss was a head.
2. $A = \cup_{i=1}^\infty \{\omega \in \Omega : \omega_i = \omega_{i+1} = 1\}$ represents the event that (at least) two heads are tossed twice in a row at some time.
3. $A = \cap_{N=1}^\infty \cup_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where there are infinitely many heads tossed in the sequence.
4. $A = \cup_{N=1}^\infty \cap_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where heads occurs from some time onwards, i.e. $\omega \in A$ iff there exists, $N = N(\omega)$ such that $\omega_n = 1$ for all $n \ge N$.

Ideally we would like to assign a probability, $P(A)$, to all events $A \subset \Omega$. Given a physical experiment, we think of assigning this probability as follows. Run the experiment many times to get sample points, $\omega(n) \in \Omega$ for each $n \in \mathbb{N}$, then try to "define" $P(A) = \lim_{N \to \infty} P_N(A)$ where

$$P_N(A) := \frac{1}{N}\sum_{k=1}^{N} 1_A(\omega(k)) \tag{3.1}$$
$$:= \frac{1}{N}\#\{1 \le k \le N : \omega(k) \in A\}. \tag{3.2}$$

Here $P_N(A)$ is the (empirical) relative frequency that event $A$ happened during the first $N$ trials. The properties of this function are indicated in the next simple lemma.

**Lemma 3.5.** *The function $P_N(\cdot)$ satisfies;*

1. $P_N(A) \in [0,1]$ *for all* $A \subset \Omega$.
2. $P_N(\emptyset) = 0$ *and* $P_N(\Omega) = 1$.
3. ***Additivity.*** *If $A$ and $B$ are disjoint event, i.e. $A \cap B = AB = \emptyset$, then* $P_N(A \cup B) = P_N(A) + P_N(B)$.
4. ***Countable Additivity.*** *More generally, if $\{A_j\}_{j=1}^{\infty}$ are pairwise disjoint events (i.e. $A_j \cap A_k = \emptyset$ for all $j \neq k$), then*
$$P_N\left(\cup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P_N(A_j).$$

**Proof.** Items 1. and 2. are obvious. For the additivity of item 3. first observe that if $A$ and $B$ are disjoint events, i.e. $A \cap B = AB = \emptyset$, then $1_{A\cup B} = 1_A + 1_B$. Therefore we have

$$P_N(A \cup B) = \frac{1}{N}\sum_{k=1}^{N} 1_{A\cup B}(\omega(k)) = \frac{1}{N}\sum_{k=1}^{N}[1_A(\omega(k)) + 1_B(\omega(k))]$$
$$= P_N(A) + P_N(B).$$

Similarly for item 4., if $\{A_j\}_{j=1}^{\infty}$ are pairwise disjoint events (i.e. $A_j \cap A_k = \emptyset$ for all $j \neq k$), then again, $1_{\cup_{j=1}^{\infty} A_j} = \sum_{j=1}^{\infty} 1_{A_j}$ and therefore

$$P_N\left(\cup_{j=1}^{\infty} A_j\right) = \frac{1}{N}\sum_{k=1}^{N} 1_{\cup_{j=1}^{\infty} A_j}(\omega(k)) = \frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{\infty} 1_{A_j}(\omega(k))$$
$$= \sum_{j=1}^{\infty}\frac{1}{N}\sum_{k=1}^{N} 1_{A_j}(\omega(k)) = \sum_{j=1}^{\infty} P_N(A_j).$$

■

We expect that $P_N(A)$ is an approximation to the "true" probability of the event $A$ which gets closer to the truth as $N \to \infty$. Thus we wish to define $P(A)$ to be the relative long term relative frequency that the event $A$ occurred for the given sequence of experiments, $\{\omega(k)\}_{k=1}^{\infty}$, i.e.

$$\text{“}P(A) := \lim_{N\to\infty} P_N(A) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} 1_A(\omega(k)).\text{”} \tag{3.3}$$

There are of course a number of problems with defining $P$ as in Eq. (3.3) of which the most important is how do we know the limit even exits. Even if the limit exists we may wonder if the answer is independent of the sequence of experiments we used to compute $P(A)$. Nevertheless, we will take it for granted in this chapter that the limit does exist and is well defined, i.e. independent of the given sequence of experiments.

Under the postulate that limit in Eq. (3.3) exists for each even $A \subset \Omega$, we may formally pass to the limit in the expressions in items 1. – 4. of Lemma 3.5 in order to "show" $P$ satisfies;

1. $P(A) \in [0,1]$ for all $A \subset \Omega$.
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
3. **Additivity.** If $A$ and $B$ are disjoint event, i.e. $A \cap B = AB = \emptyset$, then $1_{A\cup B} = 1_A + 1_B$ so that
$$P(A \cup B) = \lim_{N\to\infty} P_N(A \cup B) = \lim_{N\to\infty}[P_N(A) + P_N(B)] = P(A) + P(B).$$
4. **Countable Additivity.** If $\{A_j\}_{j=1}^{\infty}$ are pairwise disjoint events (i.e. $A_j \cap A_k = \emptyset$ for all $j \neq k$), then again, $1_{\cup_{j=1}^{\infty} A_j} = \sum_{j=1}^{\infty} 1_{A_j}$ and therefore we might hope that (by another leap of faith) that
$$P\left(\cup_{j=1}^{\infty} A_j\right) = \lim_{N\to\infty} P_N\left(\cup_{j=1}^{\infty} A_j\right) = \lim_{N\to\infty}\sum_{j=1}^{\infty} P_N(A_j)$$
$$\stackrel{?}{=} \sum_{j=1}^{\infty}\lim_{N\to\infty} P_N(A_j) = \sum_{j=1}^{\infty} P(A_j).$$

**Definition 3.6 (Probability Measures (Provisional)).** *Loosely speaking a* ***probability measure*** *is a function, $P : 2^{\Omega} \to [0,1]$ for which the 4 conditions above are satisfied. (Probability theory is the study of such functions.)*

*Example 3.7.* Let us consider the tossing of a fair coin $N$ times. In this case we would expect that every $\omega \in \Omega$ is equally likely, i.e. $P(\{\omega\}) = \frac{1}{2^N}$. Assuming this we are then forced to define

$$P(A) = \frac{1}{2^N}\#(A).$$

Observe that this probability has the following property. Suppose that $\sigma \in \{0,1\}^k$ is a given sequence, then

$$P(\{\omega : (\omega_1, \ldots, \omega_k) = \sigma\}) = \frac{1}{2^N} \cdot 2^{N-k} = \frac{1}{2^k}.$$

That is if we ignore the flips after time $k$, the resulting probabilities are the same as if we only flipped the coin $k$ times.

*Example 3.8.* The previous example suggests that if we flip a fair coin an infinite number of times, so that now $\Omega = \{0,1\}^{\mathbb{N}}$, then we should define

$$P(\{\omega \in \Omega : (\omega_1, \ldots, \omega_k) = \sigma\}) = \frac{1}{2^k} \tag{3.4}$$

for any $k \geq 1$ and $\sigma \in \{0,1\}^k$. Assuming there exists a probability, $P : 2^\Omega \to [0,1]$ such that Eq. (3.4) holds, we would like to compute, for example, the probability of the event $B$ where an infinite number of heads are tossed. To try to compute this, let

$$A_n = \{\omega \in \Omega : \omega_n = 1\} = \{\text{heads at time } n\}$$
$$B_N := \cup_{n \geq N} A_n = \{\text{at least one heads at time } N \text{ or later}\}$$

and

$$B = \cap_{N=1}^\infty B_N = \{A_n \text{ i.o.}\} = \cap_{N=1}^\infty \cup_{n \geq N} A_n.$$

Since

$$B_N^c = \cap_{n \geq N} A_n^c \subset \cap_{M \geq n \geq N} A_n^c = \{\omega \in \Omega : \omega_N = \omega_{N+1} = \cdots = \omega_M = 0\},$$

we see that

$$P(B_N^c) \leq \frac{1}{2^{M-N}} \to 0 \text{ as } M \to \infty.$$

Therefore, $P(B_N) = 1$ for all $N$. If we assume that $P$ is continuous under taking decreasing limits[1] we may conclude, using $B_N \downarrow B$, that

$$P(B) = \lim_{N \to \infty} P(B_N) = 1.$$

Without this continuity assumption we would not be able to compute $P(B)$.

The unfortunate fact is that we can not always assign a desired probability function, $P(A)$, for all $A \subset \Omega$. For example we have the following negative theorem.

**Theorem 3.9 (No-Go Theorem for fair spinners).** *Let $\Omega = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle. Then there is no probability function, $P : 2^\Omega \to [0,1]$ such that $P$ is invariant under rotations.*

[1] We will see a little later this is a consequence of countable additivity.

**Proof.** We are going to use the fact proved below in Proposition 6.4, that the continuity condition on $P$ is equivalent to the $\sigma$ – additivity of $P$. For $z \in \Omega$ and $N \subset \Omega$ let

$$zN := \{zn \in \Omega : n \in N\}, \tag{3.5}$$

that is to say $e^{i\theta}N$ is the set $N$ rotated counter clockwise by angle $\theta$. By assumption, we are supposing that

$$P(zN) = P(N) \tag{3.6}$$

for all $z \in \Omega$ and $N \subset \Omega$.

Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of $\Omega$. As above $R$ acts on $\Omega$ by rotations and divides $\Omega$ up into equivalence classes, where $z, w \in \Omega$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point $n$ from each of these equivalence classes and let $N \subset \Omega$ be the set of these representative points. Then every point $z \in \Omega$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$\Omega = \sum_{r \in R} (rN) \tag{3.7}$$

where $\sum_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (3.6) and (3.7),

$$1 = P(\Omega) = \sum_{r \in R} P(rN) = \sum_{r \in R} P(N). \tag{3.8}$$

We have thus arrived at a contradiction, since the right side of Eq. (3.8) is either equal to 0 or to $\infty$ depending on whether $P(N) = 0$ or $P(N) > 0$. ■

Here are some other related results which should give one pause even when thinking about desirable finitely additive measures.

**Theorem 3.10 (Banach–Tarski paradox 1942).** *Given any two bounded subsets $A$ and $B$ of $\mathbb{R}^d$ with $d \geq 3$, both of which have a non-empty interior, there are partitions of $A$ and $B$ into a finite number of disjoint subsets, $A = A_1 \cup \cdots \cup A_k$ and $B = B_1 \cup \cdots \cup B_k$ such that $A_i$ and $B_i$ are congruent for each $i$.*

**Theorem 3.11 (Robinson's doubling of the sphere 1947).** *It is possible to double the ball in $\mathbb{R}^3$ by decomposing it into five pieces. To be more precise if $B$ is the unit ball in $\mathbb{R}^3$ there exists $\{A_i\}_{i=1}^5 \subset B$ such that $B = \sum_{i=1}^5 A_i$ while $\sum_{i=1}^5 A_i' = B \cup B'$ where $B'$ is a translate of $B$ such that $B' \cap B = \emptyset$ an each $A_i'$ is congruent to $A_i$ for $1 \leq i \leq 5$.*

To avoid the issues inherent in the last three theorems we are going to have to relinquish the idea that $P$ should necessarily be defined on all of $2^{\Omega}$. So we are going to only define $P$ on particular subsets, $\mathcal{B} \subset 2^{\Omega}$. We will developed this below under the title of $\sigma$ – algebras.

*Remark 3.12 (What is probability about.).* Given a sample space $\Omega$ our goals in a nutshell are as follows;

1. Start with a "natural" probability $P_0$ defined on some relatively small collection of events ($\mathcal{A}$) in $\Omega$, see Proposition 5.7.
2. Verify that $P_0$ has the continuity property of being countably additive – this is a substantial restriction on $P_0$, see Proposition 6.56.
3. Show that continuous $P_0$'s have a unique extension ($P$) to ($\mathcal{B}$) – the "closure" of $\mathcal{A}$, see Theorem 6.44 and 6.52.
4. Now try to compute as explicitly as possible $P(A)$ for $A \in \mathcal{B}$ and more generally "expectations of random variables" relative to $P$. [This is essentially the main content of the course and the rest of these notes.]

# Part II

# Formal Development

# Preliminaries

## 4.1 Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers, and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

**Notation 4.1** *Given two sets $X$ and $Y$, let $Y^X$ denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in $Y$ and often write $f_n$ for $f(n)$ and express $f$ as $\{f_n\}_{n=1}^{\infty}$. If $X = \{1, 2, \dots, N\}$, we will write $Y^N$ in place of $Y^{\{1,2,\dots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \dots, f_N)$ where $f_n = f(n)$.*

**Notation 4.2** *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space $X$, then we will write $\prod_{\alpha \in A} X_\alpha$ as $X^A$ rather than $X_A$.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

**Notation 4.3** *Given a set $X$, let $2^X$ denote the **power set** of $X$ – the collection of all subsets of $X$ including the empty set.*

The reason for writing the power set of $X$ as $2^X$ is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of $X$.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = B \cap A^c.$$

We also define the **symmetric difference** of $A$ and $B$ by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of $X$ we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \,\forall\, \alpha \in I\}.$$

**Notation 4.4** *We will also write $\sum_{\alpha \in I} A_\alpha$ for $\cup_{\alpha \in I} A_\alpha$ in the case that $\{A_\alpha\}_{\alpha \in I}$ are pairwise disjoint, i.e. $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$.*

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^{\infty}$ be a sequence of subsets from $X$ and define

$$\inf_{k \geq n} A_n := \cap_{k \geq n} A_k, \quad \sup_{k \geq n} A_n := \cup_{k \geq n} A_k,$$

$$\limsup_{n \to \infty} A_n := \inf_n \sup_{k \geq n} A_k = \{x \in X : \#\{n : x \in A_n\} = \infty\} =: \{A_n \text{ i.o.}\}$$

and

$$\liminf_{n \to \infty} A_n := \sup_n \inf_{k \geq n} A_k = \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\} =: \{A_n \text{ a.a.}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as $A_n$ infinitely often and $\{A_n \text{ a.a.}\}$ as $A_n$ almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\; \exists\, n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^{\infty} \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\, N \in \mathbb{N} \ni \forall\, n \geq N,\ \ x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^{\infty} \cap_{n \geq N} A_n.$$

**Definition 4.5.** *Given a set $A \subset X$, let*

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

*be the* ***indicator function*** *of $A$.*

*Example 4.6.* Here are some example identities involving indicator fuctions. Let $A$ and $B$ be subsets of $X$, then

$$\begin{aligned} 1_{A\cap B} &= 1_A \cdot 1_B = \min(1_A, 1_B), \\ 1_{A\cup B} &= \max(1_A, 1_B), \\ 1_{A^c} &= 1 - 1_A, \text{ and } 1_{A\triangle B} = |1_A - 1_B| \end{aligned}$$

**Lemma 4.7 (Properties of** inf **and** sup**).** *We have:*

1. $(\cup_n A_n)^c = \cap_n A_n^c$,
2. $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ a.a.}\}$,
3. $\limsup_{n\to\infty} A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n\to\infty} A_n = \left\{x \in X : \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\right\}$,
5. $\sup_{k\ge n} 1_{A_k}(x) = 1_{\cup_{k\ge n} A_k} = 1_{\sup_{k\ge n} A_k}$,
6. $\inf_{k\ge n} 1_{A_k}(x) = 1_{\cap_{k\ge n} A_k} = 1_{\inf_{k\ge n} A_k}$,
7. $1_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} 1_{A_n}$, and
8. $1_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} 1_{A_n}$.

**Proof.** These results follow fairly directly from the definitions and so the proof is left to the reader – some of the results are in the exercises below. (The reader should definitely provide a proof for herself.) ∎

**Definition 4.8.** *A set $X$ is said to be* ***countable*** *if is empty or there is an injective function $f : X \to \mathbb{N}$, otherwise $X$ is said to be* ***uncountable.***

**Lemma 4.9 (Basic Properties of Countable Sets).**

1. *If $A \subset X$ is a subset of a countable set $X$ then $A$ is countable.*
2. *Any infinite subset $\Lambda \subset \mathbb{N}$ is in one to one correspondence with $\mathbb{N}$.*
3. *A non-empty set $X$ is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If $X$ and $Y$ are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that $A_m$ is a countable subset of a set $X$, then $A = \cup_{m=1}^\infty A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If $X$ is an infinite set and $Y$ is a set with at least two elements, then $Y^X$ is uncountable. In particular $2^X$ is uncountable for any infinite set $X$.*

**Proof.** We take each item in turn.

1. If $f : X \to \mathbb{N}$ is an injective map then so is the restriction, $f|_A$, of $f$ to the subset $A$.
2. Let $f(1) = \min \Lambda$ and define $f$ inductively by
$$f(n+1) = \min(\Lambda \setminus \{f(1), \dots, f(n)\}).$$
Since $\Lambda$ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection.
3. If $g : \mathbb{N} \to X$ is a surjective map, let
$$f(x) = \min g^{-1}(\{x\}) = \min\{n \in \mathbb{N} : f(n) = x\}.$$
Then $f : X \to \mathbb{N}$ is injective which combined with item 2. (taking $\Lambda = f(X)$) shows $X$ is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise.
4. Let us first construct a bijection, $h$, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form
$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \dots \\ (2,1) & (2,2) & (2,3) & \dots \\ (3,1) & (3,2) & (3,3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$
and then "count" these elements by counting the sets $\{(i,j) : i + j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$ and so on. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$.
5. If $A = \emptyset$ then $A$ is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing $A_m$ by $A_1$ if necessary we may also assume $A_m \neq \emptyset$ for all $m$. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^\infty A_m$ by $f(m,n) := a_m(n)$. The function $f$ is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to \cup_{m=1}^\infty A_m$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above.
6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \dots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all $n$ and so $a \notin f(\mathbb{N})$. This contradicts the assumption that $f$ is surjective and shows $2^{\mathbb{N}}$ is uncountable.
For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if $Y_0^X$ is uncountable then so is $Y^X$. In this way we may assume $Y_0$ is a two point

set which may as well be $Y_0 = \{0, 1\}$ . Moreover, since $X$ is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(A) := \{x_n : n \in \mathbb{N}\} \subset X$ for all $A \subset \mathbb{N}$. If $2^X$ were countable we could find an injective map $f : 2^X \to \mathbb{N}$. We then have $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ is also injective which would imply $2^{\mathbb{N}}$ is countable which it is not.

■

## 4.2 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i\in I}$ be an indexed family of subsets of $Y$, verify the following assertions.

**Exercise 4.1.** $(\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c$.

**Exercise 4.2.** Suppose that $B \subset Y$, show that $B \setminus (\cup_{i\in I} A_i) = \cap_{i\in I}(B \setminus A_i)$.

**Exercise 4.3.** Let $\{B_i\}_{i\in I}$ be another collection of subsets of $Y$. Show $[\cup_{i\in I} A_i] \setminus [\cup_{i\in I} B_i] \subset \cup_{i\in I} (A_i \setminus B_i)$ and then use this inclusion twice to show $[\cup_{i\in I} A_i] \triangle [\cup_{i\in I} B_i] \subset \cup_{i\in I} (A_i \triangle B_i)$ .

**Exercise 4.4 (Triangle inclusion for sets).** If $A, B, C$ are subsets of $X$, show $A \setminus C \subset [A \setminus B] \cup [B \setminus C]$ and use this identity twice to show

$$A \triangle C \subset [A \triangle B] \cup [B \triangle C] . \tag{4.1}$$

**Exercise 4.5.** $f^{-1}(\cup_{i\in I} A_i) = \cup_{i\in I} f^{-1}(A_i)$.

**Exercise 4.6.** $f^{-1}(\cap_{i\in I} A_i) = \cap_{i\in I} f^{-1}(A_i)$.

**Exercise 4.7.** Find a function $f : X = \{a, b, c\} \to Y = \{1, 2\}$ and subsets $C$ and $D$ of $X$ such that

$$f(C \cap D) \neq f(C) \cap f(D) \text{ and } f(C^c) \neq [f(C)]^c .$$

## 4.3 Algebraic sub-structures of sets

**Definition 4.10.** *A collection of subsets $\mathcal{A}$ of a set $X$ is a $\pi$ – **system** or **multiplicative system** if $\mathcal{A}$ is closed under taking finite intersections.*

**Definition 4.11.** *A collection of subsets $\mathcal{A}$ of a set $X$ is an **algebra (Field)** if*

*1. $\emptyset, X \in \mathcal{A}$*
*2. $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$*
*3. $\mathcal{A}$ is closed under finite unions, i.e. if $A_1, \ldots, A_n \in \mathcal{A}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{A}$. In view of conditions 1. and 2., 3. is equivalent to*
*3′. $\mathcal{A}$ is closed under finite intersections.*

**Definition 4.12.** *A collection of subsets $\mathcal{B}$ of $X$ is a $\sigma$ – **algebra** (or sometimes called a $\sigma$ – **field**) if $\mathcal{B}$ is an algebra which also closed under countable unions, i.e. if $\{A_i\}_{i=1}^\infty \subset \mathcal{B}$, then $\cup_{i=1}^\infty A_i \in \mathcal{B}$. (Notice that since $\mathcal{B}$ is also closed under taking complements, $\mathcal{B}$ is also closed under taking countable intersections.)*

*Example 4.13.* Here are some examples of algebras.

1. $\mathcal{B} = 2^X$, then $\mathcal{B}$ is a $\sigma$ – algebra.
2. $\mathcal{B} = \{\emptyset, X\}$ is a $\sigma$ – algebra called the trivial $\sigma$ – field.
3. Let $X = \{1, 2, 3\}$, then $\mathcal{A} = \{\emptyset, X, \{1\}, \{2, 3\}\}$ is an algebra while, $\mathcal{S} := \{\emptyset, X, \{2, 3\}\}$ is a not an algebra but is a $\pi$ – system.

**Proposition 4.14.** *Let $\mathcal{E}$ be any collection of subsets of $X$. Then there exists a unique smallest algebra $\mathcal{A}(\mathcal{E})$ and $\sigma$ – algebra $\sigma(\mathcal{E})$ which contains $\mathcal{E}$.*

**Proof.** Simply take

$$\mathcal{A}(\mathcal{E}) := \bigcap \{\mathcal{A} : \mathcal{A} \text{ is an algebra such that } \mathcal{E} \subset \mathcal{A}\}$$

and

$$\sigma(\mathcal{E}) := \bigcap \{\mathcal{M} : \mathcal{M} \text{ is a } \sigma - \text{algebra such that } \mathcal{E} \subset \mathcal{M}\}.$$

■

*Example 4.15.* Suppose $X = \{1, 2, 3\}$ and $\mathcal{E} = \{\emptyset, X, \{1, 2\}, \{1, 3\}\}$, see Figure 4.1. Then

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = 2^X.$$

On the other hand if $\mathcal{E} = \{\{1, 2\}\}$ , then $\mathcal{A}(\mathcal{E}) = \{\emptyset, X, \{1, 2\}, \{3\}\}$.

**Exercise 4.8.** Suppose that $\mathcal{E}_i \subset 2^X$ for $i = 1, 2$. Show that $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \mathcal{A}(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \mathcal{A}(\mathcal{E}_1)$ . Similarly show, $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \sigma(\mathcal{E}_1)$ . Give a simple example where $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ while $\mathcal{E}_1 \neq \mathcal{E}_2$.

In this course we will often be interested in the Borel $\sigma$ – algebra on a topological space.

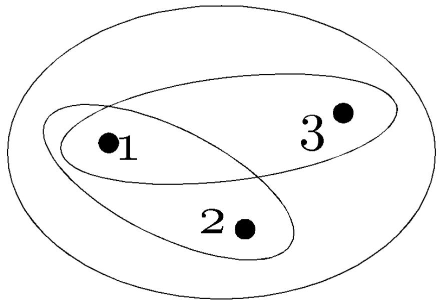



**Fig. 4.1.** A collection of subsets.

**Definition 4.16 (Borel $\sigma$ – field).** *The **Borel** $\sigma$ – **algebra**, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$, on $\mathbb{R}$ is the smallest $\sigma$ -field containing all of the open subsets of $\mathbb{R}$. More generally if $(X, \tau)$ is a topological space, the Borel $\sigma$ – algebra on $X$ is $\mathcal{B}_X := \sigma(\tau)$ – i.e. the smallest $\sigma$ – algebra containing all open (closed) subsets of $X$.*

**Exercise 4.9.** Verify the Borel $\sigma$ – algebra, $\mathcal{B}_{\mathbb{R}}$, is generated by any of the following collection of sets:

1. $\mathcal{E}_1 := \{(a,\infty) : a \in \mathbb{R}\}$, 2. $\mathcal{E}_2 := \{(a,\infty) : a \in \mathbb{Q}\}$ or 3. $\mathcal{E}_3 := \{[a,\infty) : a \in \mathbb{Q}\}$.

**Hint:** make use of the ideas in Exercise 4.8.

We will postpone a more in depth study of $\sigma$ – algebras until later. For now, let us concentrate on understanding the the simpler notion of an algebra.

**Definition 4.17.** *Let $X$ be a set. We say that a family of sets $\mathcal{F} \subset 2^X$ is a **partition** of $X$ if distinct members of $\mathcal{F}$ are disjoint and if $X$ is the union of the sets in $\mathcal{F}$.*

*Example 4.18.* Let $X$ be a set and $\mathcal{E} = \{A_1, \dots, A_n\}$ where $A_1, \dots, A_n$ is a partition of $X$. In this case

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset \{1,2,\dots,n\}\}$$

where $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. Notice that

$$\#\left(\mathcal{A}(\mathcal{E})\right) = \#(2^{\{1,2,\dots,n\}}) = 2^n.$$

*Example 4.19.* Suppose that $X$ is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#(\mathcal{A}) < \infty$. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{A} : x \in A\} \in \mathcal{A},$$

wherein we have used $\mathcal{A}$ is finite to insure $A_x \in \mathcal{A}$. Hence $A_x$ is the smallest set in $\mathcal{A}$ which contains $x$.

Now suppose that $y \in X$. If $x \in A_y$ then $A_x \subset A_y$ so that $A_x \cap A_y = A_x$. On the other hand, if $x \notin A_y$ then $x \in A_x \setminus A_y$ and therefore $A_x \subset A_x \setminus A_y$, i.e. $A_x \cap A_y = \emptyset$. Therefore we have shown, either $A_x \cap A_y = \emptyset$ or $A_x \cap A_y = A_x$. By reversing the roles of $x$ and $y$ it also follows that either $A_y \cap A_x = \emptyset$ or $A_y \cap A_x = A_y$. Therefore we may conclude, either $A_x = A_y$ or $A_x \cap A_y = \emptyset$ for all $x, y \in X$.

[Alternatively, let $x, y \in X$. If $x \notin A_y$, then $x \in A_x \setminus A_y \in \mathcal{A}$ and therefore $A_x \setminus A_y = A_x$, i.e. $A_y \cap A_x = \emptyset$. Similarly if $y \notin A_x$, then $A_y \cap A_x = \emptyset$. From these remarks we may now also conclude that if $x \in A_y$, then $y \in A_x$ (for otherwise $A_y \cap A_x = \emptyset$) and therefore $A_x \subset A_y$ and $A_y \subset A_x$, i.e. $A_x = A_y$.]

Let us now define $\{B_i\}_{i=1}^k$ to be an enumeration of $\{A_x\}_{x\in X}$. It is a straightforward to conclude that

$$\mathcal{A} = \{\cup_{i\in\Lambda} B_i : \Lambda \subset \{1,2,\dots,k\}\}.$$

For example observe that for any $A \in \mathcal{A}$, we have $A = \cup_{x\in A} A_x = \cup_{i\in\Lambda} B_i$ where $\Lambda := \{i : B_i \subset A\}$.

**Proposition 4.20.** *Suppose that $\mathcal{B} \subset 2^X$ is a $\sigma$ – algebra and $\mathcal{B}$ is at most a countable set. Then there exists a unique **finite** partition $\mathcal{F}$ of $X$ such that $\mathcal{F} \subset \mathcal{B}$ and every element $B \in \mathcal{B}$ is of the form*

$$B = \cup\{A \in \mathcal{F} : A \subset B\}. \tag{4.2}$$

*In particular $\mathcal{B}$ is actually a finite set and $\#(\mathcal{B}) = 2^n$ for some $n \in \mathbb{N}$.*

**Proof.** We proceed as in Example 4.19. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{B} : x \in A\} \in \mathcal{B},$$

wherein we have used $\mathcal{B}$ is a countable $\sigma$ – algebra to insure $A_x \in \mathcal{B}$. Just as above either $A_x \cap A_y = \emptyset$ or $A_x = A_y$ and therefore $\mathcal{F} = \{A_x : x \in X\} \subset \mathcal{B}$ is a (necessarily countable) partition of $X$ for which Eq. (4.2) holds for all $B \in \mathcal{B}$.

Enumerate the elements of $\mathcal{F}$ as $\mathcal{F} = \{P_n\}_{n=1}^N$ where $N \in \mathbb{N}$ or $N = \infty$. If $N = \infty$, then the correspondence

$$a \in \{0,1\}^{\mathbb{N}} \to A_a = \cup\{P_n : a_n = 1\} \in \mathcal{B}$$

is bijective and therefore, by Lemma 4.9, $\mathcal{B}$ is uncountable. Thus any countable $\sigma$ – algebra is necessarily finite. This finishes the proof modulo the uniqueness assertion which is left as an exercise to the reader. ■

*Example 4.21 (Countable/Co-countable $\sigma$ – Field).* Let $X = \mathbb{R}$ and $\mathcal{E} := \{\{x\} : x \in \mathbb{R}\}$. Then $\sigma(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that $A$ is countable or $A^c$ is countable. Similarly, $\mathcal{A}(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that $A$ is finite or $A^c$ is finite. More generally we have the following exercise.

**Exercise 4.10 (Look at but do not hand in.).** Let $X$ be a set, $I$ be an **infinite** index set, and $\mathcal{E} = \{A_i\}_{i\in I}$ be a partition of $X$. Prove the algebra, $\mathcal{A}(\mathcal{E})$, and that $\sigma -$ algebra, $\sigma(\mathcal{E})$, generated by $\mathcal{E}$ are given by

$$\mathcal{A}(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \#(\Lambda) < \infty \text{ or } \#(\Lambda^c) < \infty\} \tag{4.3}$$

and

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \Lambda \text{ countable or } \Lambda^c \text{ countable}\} \tag{4.4}$$

respectively. Here we are using the convention that $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. In particular if $I$ is countable, then

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I\}.$$

**Proposition 4.22.** *Let $X$ be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

$$\mathcal{A}(\mathcal{E}) := \{\text{finite unions of finite intersections of elements from } \mathcal{E}_c\}. \tag{4.5}$$

**Proof.** Let $\mathcal{A}$ denote the right member of Eq. (4.5). From the definition of an algebra, it is clear that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$. Hence to finish that proof it suffices to show $\mathcal{A}$ is an algebra. The proof of these assertions are routine except for possibly showing that $\mathcal{A}$ is closed under complementation. To check $\mathcal{A}$ is closed under complementation, let $Z \in \mathcal{A}$ be expressed as

$$Z = \bigcup_{i=1}^N \bigcap_{j=1}^K A_{ij}$$

where $A_{ij} \in \mathcal{E}_c$. Therefore, writing $B_{ij} = A_{ij}^c \in \mathcal{E}_c$, we find that

$$Z^c = \bigcap_{i=1}^N \bigcup_{j=1}^K B_{ij} = \bigcup_{j_1,\dots,j_N=1}^K (B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}) \in \mathcal{A}$$

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}$ is a finite intersection of sets from $\mathcal{E}_c$. ∎

**Corollary 4.23.** *Let $\mathcal{E} \subset 2^\Omega$. If $\#(\mathcal{E}) < \infty$ then $\#(\mathcal{A}(\mathcal{E})) < \infty$ and $\sigma(\mathcal{E}) = \mathcal{A}(\mathcal{E})$.*

*Remark 4.24.* One might think that in general $\sigma(\mathcal{E})$ may be described as the countable unions of countable intersections of sets in $\mathcal{E}^c$. However this is in general **false,** since if

$$Z = \bigcup_{i=1}^\infty \bigcap_{j=1}^\infty A_{ij}$$

with $A_{ij} \in \mathcal{E}_c$, then

$$Z^c = \bigcup_{j_1=1,j_2=1,\dots j_N=1,\dots}^\infty \left(\bigcap_{\ell=1}^\infty A_{\ell,j_\ell}^c\right)$$

which is now an **uncountable** union. Thus the above description is not correct. In general it is complicated to explicitly describe $\sigma(\mathcal{E})$, see Proposition 1.23 on page 39 of Folland for details. Also see Proposition 4.20.

**Definition 4.25 (Topologies).** *A collection $\tau \subset 2^X$ is a said to be a topology on $X$ if*

1. *$\emptyset, X \in \tau$*
2. *if $V_1, V_2 \in \tau$, then $V_1 \cap V_2 \in \tau$, i.e. $\tau$ is closed under finite intersections.*
3. *If $\{V_\alpha\}_{\alpha\in A} \subset \tau$ then $\cup_{\alpha\in A} V_\alpha \in \tau$, i.e. $\tau$ is closed under arbitrary unions.*

*The sets $V \in \tau$ are called **open sets** while those sets $F \subset X$ such that $F^c \in \tau$ are said to be **closed sets.***

**Exercise 4.11.** Let $\tau$ be a topology on a set $X$ and $\mathcal{A} = \mathcal{A}(\tau)$ be the algebra generated by $\tau$. Show $\mathcal{A}$ is the collection of subsets of $X$ which may be written as finite union of sets of the form $F \cap V$ where $F$ is closed and $V$ is open.

**Definition 4.26.** *A set $\mathcal{S} \subset 2^X$ is said to be an **semialgebra or elementary class** provided that*

- *$\emptyset \in \mathcal{S}$*
- *$\mathcal{S}$ is closed under finite intersections*
- *if $E \in \mathcal{S}$, then $E^c$ is a finite disjoint union of sets from $\mathcal{S}$. (In particular $X = \emptyset^c$ is a finite disjoint union of elements from $\mathcal{S}$.)*

*We will typically denote semi-algebras or elementary classes by either $\mathcal{S}$ or $\mathcal{E}$.*

**Proposition 4.27.** *Suppose $\mathcal{S} \subset 2^X$ is a elementary class, then $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$.*

**Proof.** (Although it is possible to give a proof using Proposition 4.22, it is just as simple to give a direct proof.) Let $\mathcal{A}$ denote the collection of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$. Clearly $\mathcal{S} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{S})$ so it suffices to show $\mathcal{A}$ is an algebra since $\mathcal{A}(\mathcal{S})$ is the smallest algebra containing $\mathcal{S}$. By the properties of $\mathcal{S}$, we know that $\emptyset, X \in \mathcal{A}$. The following two steps now finish the proof.

1. ($\mathcal{A}$ is closed under finite intersections.) Suppose that $A_i = \sum_{F\in\Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \dots, n$, $\Lambda_i$ is a finite collection of disjoint sets from $\mathcal{S}$. Then

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \left( \sum_{F \in \Lambda_i} F \right) = \bigcup_{(F_1,,\dots,F_n) \in \Lambda_1 \times \cdots \times \Lambda_n} (F_1 \cap F_2 \cap \cdots \cap F_n)$$

and this is a disjoint (you check) union of elements from $\mathcal{S}$. Therefore $\mathcal{A}$ is closed under finite intersections.

2. ($\mathcal{A}$ is closed under complementation.) If $A = \sum_{F \in \Lambda} F$ with $\Lambda$ being a finite collection of disjoint sets from $\mathcal{S}$, then $A^c = \bigcap_{F \in \Lambda} F^c$. Since, by assumption, $F^c \in \mathcal{A}$ for all $F \in \Lambda \subset \mathcal{S}$ and $\mathcal{A}$ is closed under finite intersections by step 1., it follows that $A^c \in \mathcal{A}$. ■

*Example 4.28.* Let $X = \mathbb{R}$, then

$$\begin{aligned} \mathcal{S} &:= \left\{ (a, b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}} \right\} \\ &= \{(a, b] : a \in [-\infty, \infty) \text{ and } a < b < \infty\} \cup \{\emptyset, \mathbb{R}\} \end{aligned}$$

is a elementary class. The algebra, $\mathcal{A}(\mathcal{S})$, generated by $\mathcal{S}$ consists of finite disjoint unions of sets from $\mathcal{S}$. For example,

$$A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in \mathcal{A}(\mathcal{S}).$$

**Exercise 4.12.** Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be elementary class. Show the collection

$$\mathcal{S} := \mathcal{A} \dot{\times} \mathcal{B} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$$

is also a elementary class.

# Finitely Additive Measures / Integration

**Definition 5.1.** *Suppose that* $\mathcal{E} \subset 2^{\Omega}$ *is a collection of subsets of* $\Omega$ *and* $\mu : \mathcal{E} \to [0, \infty]$ *is a function. Then*

1. $\mu$ *is* ***additive or finitely additive on*** $\mathcal{E}$ *if*
$$\mu(E) = \sum_{i=1}^{n} \mu(E_i) \tag{5.1}$$
*whenever* $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ *with* $E_i \in \mathcal{E}$ *for* $i = 1, 2, \dots, n < \infty$.
2. $\mu$ *is* $\sigma$ – ***additive (or countable additive) on*** $\mathcal{E}$ *if Eq. (5.1) holds even when* $n = \infty$.
3. $\mu$ *is* ***sub-additive (finitely sub-additive) on*** $\mathcal{E}$ *if*
$$\mu(E) \le \sum_{i=1}^{n} \mu(E_i)$$
*whenever* $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ *with* $n \in \mathbb{N} \cup \{\infty\}$ $(n \in \mathbb{N})$.
4. $\mu$ *is a* ***finitely additive measure*** *if* $\mathcal{E} = \mathcal{A}$ *is an algebra,* $\mu(\emptyset) = 0$*, and* $\mu$ *is finitely additive on* $\mathcal{A}$.
5. $\mu$ *is a* ***premeasure*** *if* $\mu$ *is a finitely additive measure which is* $\sigma$ – *additive on* $\mathcal{A}$.
6. $\mu$ *is a* ***measure*** *if* $\mu$ *is a premeasure on a* $\sigma$ – *algebra. Furthermore if* $\mu(\Omega) = 1$*, we say* $\mu$ *is a* ***probability measure*** *on* $\Omega$.

**Proposition 5.2 (Basic properties of finitely additive measures).** *Suppose* $\mu$ *is a finitely additive measure on an algebra,* $\mathcal{A} \subset 2^{\Omega}$*,* $A, B \in \mathcal{A}$ *with* $A \subset B$ *and* $\{A_j\}_{j=1}^{n} \subset \mathcal{A}$*, then :*

1. *(*$\mu$ *is* ***monotone****)* $\mu(A) \le \mu(B)$ *if* $A \subset B$.
2. *For* $A, B \in \mathcal{A}$*, the following* ***strong additivity formula*** *holds;*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{5.2}$$
3. *(*$\mu$ *is* ***finitely subbadditive****)* $\mu(\cup_{j=1}^{n} A_j) \le \sum_{j=1}^{n} \mu(A_j)$.
4. $\mu$ *is sub-additive on* $\mathcal{A}$ *iff*
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \sum_{i=1}^{\infty} A_i \tag{5.3}$$
*where* $A \in \mathcal{A}$ *and* $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ *are pairwise disjoint sets.*
5. *(*$\mu$ *is* ***countably superadditive****) If* $A = \sum_{i=1}^{\infty} A_i$ *with* $A_i, A \in \mathcal{A}$*, then*
$$\mu\left(\sum_{i=1}^{\infty} A_i\right) \ge \sum_{i=1}^{\infty} \mu(A_i). \tag{5.4}$$
*(See Remark 5.9 for example where this inequality is strict.)*
6. *A finitely additive measure,* $\mu$*, is a premeasure iff* $\mu$ *is subadditive.*

**Proof.**

1. Since $B$ is the disjoint union of $A$ and $(B \setminus A)$ and $B \setminus A = B \cap A^c \in \mathcal{A}$ it follows that
$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$
2. Since
$$A \cup B = [A \setminus (A \cap B)] \sum [B \setminus (A \cap B)] \sum A \cap B,$$
$$\begin{aligned}\mu(A \cup B) &= \mu(A \cup B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B).\end{aligned}$$
Adding $\mu(A \cap B)$ to both sides of this equation proves Eq. (5.2).
3. Let $\{E_i\}_{i=1}^{n} \subset \mathcal{A}$ and set $\widetilde{E}_j = E_j \setminus (E_1 \cup \dots \cup E_{j-1})$ so that the $\tilde{E}_j$ 's are pair-wise disjoint and $E = \cup_{j=1}^{n} \widetilde{E}_j$. Since $\tilde{E}_j \subset E_j$ it follows from the monotonicity of $\mu$ that
$$\mu(E) = \sum_{j=1}^{n} \mu(\widetilde{E}_j) \le \sum_{j=1}^{n} \mu(E_j).$$
4. If $A = \bigcup_{i=1}^{\infty} B_i$ with $A \in \mathcal{A}$ and $B_i \in \mathcal{A}$, then $A = \sum_{i=1}^{\infty} A_i$ where $A_i := B_i \setminus (B_1 \cup \dots B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of $\mu$ and Eq. (5.3)
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \le \sum_{i=1}^{\infty} \mu(B_i).$$

5. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then $\sum_{i=1}^{n} A_i \subset A$ for all $n$ and so by the monotonicity and finite additivity of $\mu$, $\sum_{i=1}^{n} \mu(A_i) \le \mu(A)$. Letting $n \to \infty$ in this equation shows $\mu$ is superadditive.
6. This is a combination of items 5. and 6.

■

## 5.1 Examples of Measures

Most $\sigma$ − algebras and $\sigma$ -additive measures are somewhat difficult to describe and define. However, there are a few special cases where we can describe explicitly what is going on.

*Example 5.3.* Suppose that $\Omega$ is a finite set, $\mathcal{B} := 2^{\Omega}$, and $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Then

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \subset \Omega$$

defines a measure on $2^{\Omega}$.

*Example 5.4.* Suppose that $\Omega$ is any set and $\omega \in \Omega$ is a point. For $A \subset \Omega$, let

$$\delta_\omega(A) = \begin{cases} 1 \text{ if } & \omega \in A \\ 0 \text{ if } & \omega \notin A. \end{cases}$$

Then $\mu = \delta_\omega$ is a measure on $\Omega$ called the Dirac delta measure at $\omega$.

*Example 5.5.* Suppose $\mathcal{B} \subset 2^{\Omega}$ is a $\sigma$ algebra, $\mu$ is a measure on $\mathcal{B}$, and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on $\mathcal{B}$. Moreover, if $J$ is an index set and $\{\mu_j\}_{j\in J}$ are all measures on $\mathcal{B}$, then $\mu = \sum_{j=1}^{\infty} \mu_j$, i.e.

$$\mu(A) := \sum_{j=1}^{\infty} \mu_j(A) \text{ for all } A \in \mathcal{B},$$

defines another measure on $\mathcal{B}$. To prove this we must show that $\mu$ is countably additive. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{B}$, then (using Tonelli for sums, Proposition 1.9),

$$\begin{aligned} \mu(A) &= \sum_{j=1}^{\infty} \mu_j(A) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \mu_j(A_i) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu_j(A_i) = \sum_{i=1}^{\infty} \mu(A_i). \end{aligned}$$

*Example 5.6.* Suppose that $\Omega$ is a countable set and $\lambda : \Omega \to [0,\infty]$ is a function. Let $\Omega = \{\omega_n\}_{n=1}^{\infty}$ be an enumeration of $\Omega$ and then we may define a measure $\mu$ on $2^{\Omega}$ by,

$$\mu = \mu_\lambda := \sum_{n=1}^{\infty} \lambda(\omega_n)\delta_{\omega_n}.$$

We will now show this measure is independent of our choice of enumeration of $\Omega$ by showing,

$$\mu(A) = \sum_{\omega\in A} \lambda(\omega) := \sup_{\Lambda\subset_f A} \sum_{\omega\in\Lambda} \lambda(\omega) \;\; \forall\; A \subset \Omega. \tag{5.5}$$

Here we are using the notation, $\Lambda \subset_f A$ to indicate that $\Lambda$ is a finite subset of $A$.

To verify Eq. (5.5), let $M := \sup_{\Lambda\subset_f A} \sum_{\omega\in\Lambda} \lambda(\omega)$ and for each $N \in \mathbb{N}$ let

$$\Lambda_N := \{\omega_n : \omega_n \in A \text{ and } 1 \le n \le N\}.$$

Then by definition of $\mu$,

$$\begin{aligned} \mu(A) &= \sum_{n=1}^{\infty} \lambda(\omega_n)\delta_{\omega_n}(A) = \lim_{N\to\infty} \sum_{n=1}^{N} \lambda(\omega_n) 1_{\omega_n\in A} \\ &= \lim_{N\to\infty} \sum_{\omega\in\Lambda_N} \lambda(\omega) \le M. \end{aligned}$$

On the other hand if $\Lambda \subset_f A$, then

$$\sum_{\omega\in\Lambda} \lambda(\omega) = \sum_{n:\; \omega_n\in\Lambda} \lambda(\omega_n) = \mu(\Lambda) \le \mu(A)$$

from which it follows that $M \le \mu(A)$. This shows that $\mu$ is independent of how we enumerate $\Omega$.

The above example has a natural extension to the case where $\Omega$ is uncountable and $\lambda : \Omega \to [0,\infty]$ is any function. In this setting we simply may define $\mu : 2^{\Omega} \to [0,\infty]$ using Eq. (5.5). We leave it to the reader to verify that this is indeed a measure on $2^{\Omega}$.

We will construct many more measure in Chapter 6 below. The starting point of these constructions will be the construction of finitely additive measures using the next proposition.

**Proposition 5.7 (Construction of Finitely Additive Measures).** *Suppose $\mathcal{S} \subset 2^{\Omega}$ is a semi-algebra (see Definition 4.26) and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ is the algebra generated by $\mathcal{S}$. Then every additive function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure (which we still denote by $\mu$) on $\mathcal{A}$.*

**Proof.** Since (by Proposition 4.27) every element $A \in \mathcal{A}$ is of the form $A = \sum_i E_i$ for a finite collection of $E_i \in \mathcal{S}$, it is clear that if $\mu$ extends to a measure then the extension is unique and must be given by

$$\mu(A) = \sum_i \mu(E_i). \tag{5.6}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (5.6) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

$$\sum_i \mu(E_i) = \sum_j \mu(F_j). \tag{5.7}$$

But $E_i = \sum_j (E_i \cap F_j)$ and the additivity of $\mu$ on $\mathcal{S}$ implies $\mu(E_i) = \sum_j \mu(E_i \cap F_j)$ and hence

$$\sum_i \mu(E_i) = \sum_i \sum_j \mu(E_i \cap F_j) = \sum_{i,j} \mu(E_i \cap F_j).$$

Similarly,

$$\sum_j \mu(F_j) = \sum_{i,j} \mu(E_i \cap F_j)$$

which combined with the previous equation shows that Eq. (5.7) holds. It is now easy to verify that $\mu$ extended to $\mathcal{A}$ as in Eq. (5.6) is an additive measure on $\mathcal{A}$. ■

**Proposition 5.8.** *Let $\Omega = \mathbb{R}$, $\mathcal{S}$ be the semi-algebra,*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{5.8}$$

*and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{S}$, see Proposition 4.27. To each finitely additive probability measures $\mu : \mathcal{A} \to [0, \infty]$, there is a unique increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and*

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) \ \forall\ a \le b \text{ in } \bar{\mathbb{R}}. \tag{5.9}$$

*Conversely, given an increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ there is a unique finitely additive measure $\mu = \mu_F$ on $\mathcal{A}$ such that the relation in Eq. (5.9) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a \downarrow -\infty} F(a)$ and $F(\infty) = \lim_{b \uparrow \infty} F(b)$.)*

**Proof.** Given a finitely additive probability measure $\mu$, let

$$F(x) := \mu((-\infty, x] \cap \mathbb{R}) \text{ for all } x \in \bar{\mathbb{R}}.$$

Then $F(\infty) = 1$, $F(-\infty) = 0$ and for $b > a$,

$$F(b) - F(a) = \mu((-\infty, b] \cap \mathbb{R}) - \mu((-\infty, a]) = \mu((a, b] \cap \mathbb{R}).$$

Conversely, suppose $F : \bar{\mathbb{R}} \to [0, 1]$ as in the statement of the theorem is given. Define $\mu$ on $\mathcal{S}$ using the formula in Eq. (5.9). The argument will be completed by showing $\mu$ is additive on $\mathcal{S}$ and hence, by Proposition 5.7, has a unique extension to a finitely additive measure on $\mathcal{A}$. Suppose that

$$(a, b] = \sum_{i=1}^n (a_i, b_i].$$

By reordering $(a_i, b_i]$ if necessary, we may assume that

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_{n-1} = a_n < b_n = b.$$

Therefore, by the telescoping series argument,

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) = \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu((a_i, b_i] \cap \mathbb{R}).$$

■

*Remark 5.9.* Suppose that $F : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ is any non-decreasing function such that $F(\mathbb{R}) \subset \mathbb{R}$. Then the same methods used in the proof of Proposition 5.8 shows that there exists a unique finitely additive measure, $\mu = \mu_F$, on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ such that Eq. (5.9) holds. If $F(\infty) > \lim_{b \uparrow \infty} F(b)$ and $A_i = (i, i + 1]$ for $i \in \mathbb{N}$, then

$$\begin{aligned}\sum_{i=1}^\infty \mu_F(A_i) &= \sum_{i=1}^\infty (F(i+1) - F(i)) = \lim_{N \to \infty} \sum_{i=1}^N (F(i+1) - F(i)) \\ &= \lim_{N \to \infty} (F(N+1) - F(1)) < F(\infty) - F(1) = \mu_F(\cup_{i=1}^\infty A_i).\end{aligned}$$

This shows that strict inequality can hold in Eq. (5.4) and that $\mu_F$ is **not** a premeasure. Similarly one shows $\mu_F$ is **not** a premeasure if $F(-\infty) < \lim_{a \downarrow -\infty} F(a)$ or if $F$ is **not** right continuous at some point $a \in \mathbb{R}$. Indeed, in the latter case consider

$$(a, a+1] = \sum_{n=1}^\infty (a + \frac{1}{n+1}, a + \frac{1}{n}].$$

Working as above we find,

$$\sum_{n=1}^\infty \mu_F\left((a + \frac{1}{n+1}, a + \frac{1}{n}]\right) = F(a+1) - F(a+)$$

while $\mu_F((a, a+1]) = F(a+1) - F(a)$. We will eventually show in Chapter 6 below that $\mu_F$ extends uniquely to a $\sigma -$ additive measure on $\mathcal{B}_{\mathbb{R}}$ whenever $F$ is increasing, right continuous, and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$.

## 5.2 Measures on Product Spaces

In this section, let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be algebras.

**Definition 5.10.** *Let*

$$\mathcal{A}\dot{\times}\mathcal{B} := \{A \times B : (A, B) \in \mathcal{A} \times \mathcal{B}\}$$

*and* $\mathcal{A} \odot \mathcal{B}$ *be the sub-algebra of* $2^{X\times Y}$ *generated by* $\mathcal{A}\dot{\times}\mathcal{B}$.

As we have seen in Exercise 4.12, $\mathcal{A}\dot{\times}\mathcal{B}$ is a semi-algebra and therefore $\mathcal{A}\odot\mathcal{B}$ consists of subsets, $C \subset X \times Y$, which may be written as;

$$C = \sum_{i=1}^{n} A_i \times B_i \text{ with } (A_i, B_i) \in \mathcal{A} \times \mathcal{B}. \tag{5.10}$$

**Definition 5.11.** *A function,* $\rho : \mathcal{A}\dot{\times}\mathcal{B} \to \mathbb{C}$ *is* ***bi-additive*** *if for each* $A \in \mathcal{A}$, *the function*

$$B \in \mathcal{B} \to \rho(A \times B) \in \mathbb{C}$$

*is an additive measure on* $\mathcal{B}$ *and for each* $B \in \mathcal{B}$, *the function*

$$A \in \mathcal{A} \to \rho(A \times B) \in \mathbb{C}$$

*is an additive measure on* $\mathcal{A}$.

**Theorem 5.12.** *If* $\rho : \mathcal{A}\dot{\times}\mathcal{B} \to \mathbb{C}$ *is a bi-additive function, then* $\rho$ *extends uniquely to an additive measure on the product algebra* $\mathcal{A} \odot \mathcal{B}$. *[This theorem has an obvious generalization to multiple factors.]*

**Proof.** The collection $\mathcal{E} = \mathcal{A}\dot{\times}\mathcal{B}$ is an elementary family, see Exercise 4.12. Therefore, it suffices to show $\rho$ is additive on $\mathcal{E}$. To check this suppose that $A \times B \in \mathcal{E}$ and

$$A \times B = \sum_{k=1}^{n} (A_k \times B_k)$$

with $A_k \times B_k \in \mathcal{E}$. We wish to show

$$\rho(A \times B) = \sum_{k=1}^{n} \rho(A_k \times B_k).$$

For this consider the finite algebras $\mathcal{A}' \subset 2^A$ and $\mathcal{B}' \subset 2^B$ generated by $\{A_k\}_{k=1}^n$ and $\{B_k\}_{k=1}^n$ respectively. Let $\Pi_a \subset \mathcal{A}'$ and $\Pi_b \subset \mathcal{B}'$ be partitions of $A$ and $B$ which generate $\mathcal{A}'$ and $\mathcal{B}'$ respectively as described in Proposition 4.20. Then

$$\Pi_a \dot{\times} \Pi_b = \{\alpha \times \beta : (\alpha, \beta) \in \Pi_a \times \Pi_b\}$$

is a partition of $A \times B$. I now claim to each $(\alpha, \beta) \in \Pi_a \times \Pi_b$ there exists a unique $k$ such that $\alpha \times \beta \subset A_k \times B_k$. Indeed, choose an $x \in \alpha$ and $y \in \beta$, then there exists a unique $k$ such that $(x, y) \in A_k \times B_k$ and since $\alpha \cap A_k \neq \emptyset$ and $\beta \cap B_k \neq \emptyset$ we must have $\alpha \subset A_k$ and $\beta \subset B_k$. The consequence of this observation is that

$$A_k \times B_k = \sum_{\alpha\subset A_k,\ \beta\subset B_k} \alpha \times \beta \text{ for } 1 \le k \le n,$$

where we agree that sums involving $\alpha$ $(\beta)$ run through $\Pi_a$ $(\Pi_b)$.

By the construction of $\Pi_a$ and $\Pi_b$ we also have

$$A_k = \sum_{\alpha\subset A_k} \alpha, \text{ and } B_k = \sum_{\beta\subset B_k} \beta.$$

Using the bi-additivity of $\rho$ it then follows that

$$\begin{aligned} \rho(A_k \times B_k) &= \rho\left(A_k \times \sum_{\beta\subset B_k} \beta\right) = \sum_{\beta\subset B_k} \rho(A_k \times \beta) \\ &= \sum_{\beta\subset B_k} \rho\left(\sum_{\alpha\subset A_k} \alpha \times \beta\right) = \sum_{\beta\subset B_k}\sum_{\alpha\subset A_k} \rho(\alpha \times \beta) \\ &= \sum_{\alpha\times\beta\subset A_k\times B_k} \rho(\alpha \times \beta). \end{aligned} \tag{5.11}$$

By summing this equation on $k$, using the claim above, and then the bi-additivity of $\rho$ again we learn that

$$\begin{aligned} \sum_{k=1}^{n} \rho(A_k \times B_k) &= \sum_{(\alpha,\beta)\in\Pi_a\times\Pi_b} \rho(\alpha \times \beta) \\ &= \sum_{\alpha\in\Pi_a} \rho(\alpha \times B) = \rho(A \times B). \end{aligned}$$

■

*Example 5.13 (Product Measure).* If $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be are subalgebras and $\mu : \mathcal{A} \to [0, \infty]$ and $\nu : \mathcal{B} \to [0, \infty]$ are finitely additive measures then there exists a unique finitely additive measure, $\mu \odot \nu : \mathcal{A} \odot \mathcal{B} \to [0, \infty]$ such that $\mu \odot \nu(A \times B) = \mu(A) \cdot \nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. We refer to $\mu \odot \nu$ as a **product measure.** To verify this assertion one needs only apply Theorem 5.12 with $\rho(A \times B) := \mu(A) \cdot \nu(B)$.

Here is another interesting application of Theorem 5.12.

**Proposition 5.14.** *Suppose that* $\mathcal{A} \subset 2^X$ *is an algebra and for each* $t \in \mathbb{R}$, $\mu_t : \mathcal{A} \to \mathbb{C}$ *is a finitely additive measure. Let* $Y = (u, v] \subset \mathbb{R}$ *be a finite interval and* $\mathcal{B} \subset 2^Y$ *denote the algebra generated by* $\mathcal{E} := \{(a, b] : (a, b] \subset Y\}$. *Then there is a unique additive measure* $\mu$ *on* $\mathcal{C}$, *the algebra generated by* $\mathcal{A} \dot{\times} \mathcal{B} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$ *such that*

$$\mu(A \times (a, b]) = \mu_b(A) - \mu_a(A) \ \forall \ (a, b] \in \mathcal{E} \text{ and } A \in \mathcal{A}.$$

**Proof.** By Proposition 5.8 and Remark 5.9, for each $A \in \mathcal{A}$, the function $(a, b] \to \mu(A \times (a, b])$ extends to a unique measure on $\mathcal{B}$ which we continue to denote by $\mu$. Now if $B \in \mathcal{B}$, then $B = \sum_k I_k$ with $I_k \in \mathcal{E}$, then

$$\mu(A \times B) = \sum_k \mu(A \times I_k)$$

from which we learn that $A \to \mu(A \times B)$ is still finitely additive. The proof is complete with an application of Theorem 5.12. ■

### 5.2.1 Measures on simple product spaces

Let $S$ be a finite[1] (we refer to $S$ as **state space**), $\Omega := S^\infty := S^\mathbb{N}$ (think of $\mathbb{N}$ as time and $\Omega$ as **path space)**

$$\mathcal{A}_n := \{B \times \Omega : B \subset S^n\} \text{ for all } n \in \mathbb{N},$$

$\mathcal{A} := \cup_{n=1}^\infty \mathcal{A}_n$. We call the elements, $A \in \mathcal{A}$, the **cylinder subsets of** $\Omega$. Notice that $A \subset \Omega$ is a cylinder set iff there exists $n \in \mathbb{N}$ and $B \subset S^n$ such that

$$A = B \times \Omega := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Also observe that we may write $A$ as $A = B' \times \Omega$ where $B' = B \times S^k \subset S^{n+k}$ for any $k \geq 0$.

**Exercise 5.1 (Look at but do not hand in.).** Show;

1. $\mathcal{A}_n$ is an algebra for each $n \in \mathbb{N}$,
2. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all $n$, and
3. $\mathcal{A} \subset 2^\Omega$ is an algebra of subsets of $\Omega$. (In fact, you might show that $\mathcal{A} = \cup_{n=1}^\infty \mathcal{A}_n$ is an algebra whenever $\{\mathcal{A}_n\}_{n=1}^\infty$ is an increasing sequence of algebras.)

Let us now further suppose that $P : \mathcal{A} \to [0, 1]$ is a finitely additive probability measure and for $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, let

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \tag{5.12}$$

[1] One could take $S$ to be countable here but we take $S$ to be finite for simplicity.

**Exercise 5.2 (Consistency Conditions).** If $p_n$ is defined as above, show:

1. $\sum_{s \in S} p_1(s) = 1$ and
2. for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$,

$$p_n(s_1, \dots, s_n) = \sum_{s \in S} p_{n+1}(s_1, \dots, s_n, s).$$

**Exercise 5.3 (Converse to 5.2).** Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : S^n \to [0, 1]$ such that the consistency conditions in Exercise 5.2 hold. Then there exists a unique finitely additive probability measure, $P$ on $\mathcal{A}$ such that Eq. (5.12) holds for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$.

**Corollary 5.15.** *Suppose for each* $k \in \mathbb{N}$, $q_k : S \to [0, 1]$ *is a function such that* $\sum_{s \in S} q_k(s) = 1$. *Then there exists a unique finitely additive probability measure,* $P$ *on* $\mathcal{A}$ *such that*

$$P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}) = q_1(s_1) q_2(s_2) \cdots q_n(s_n)$$

*holds for all* $n \in \mathbb{N}$ *and* $(s_1, \dots, s_n) \in S^n$.

**Proof.** Let $p_n(s_1, \dots, s_n) := q_1(s_1) q_2(s_2) \cdots q_n(s_n)$ and observe that

$$\sum_{s \in S} p_1(s) = \sum_{s \in S} q_1(s) = 1$$

and

$$\begin{aligned}\sum_{\lambda \in S} p_{n+1}(s_1, \dots, s_n, \lambda) &= \sum_{\lambda \in S} q_1(s_1) q_2(s_2) \cdots q_n(s_n) q_{n+1}(\lambda) \\ &= q_1(s_1) q_2(s_2) \cdots q_n(s_n) \sum_{\lambda \in S} q_{n+1}(\lambda) \\ &= q_1(s_1) q_2(s_2) \cdots q_n(s_n) = p_n(s_1, \dots, s_n).\end{aligned}$$

Hence the result follows from Exercise 5.3. ■

## 5.3 Simple Random Variables

Before constructing $\sigma$ – additive measures (see Chapter 6 below), we are going to pause to discuss a preliminary notion of integration and develop some of its properties. Hopefully this will help the reader to develop the necessary intuition before heading to the general theory. First we need to describe the functions we are (currently) able to integrate.

**Definition 5.16 (Simple random variables).** *A function, $f:\Omega\to Y$ is said to be **simple** if $f(\Omega)\subset Y$ is a finite set. If $\mathcal{A}\subset 2^\Omega$ is an algebra, we say that a simple function $f:\Omega\to Y$ is **measurable** if $\{f=y\}:=f^{-1}(\{y\})\in\mathcal{A}$ for all $y\in Y$. A measurable simple function, $f:\Omega\to\mathbb{C}$, is called a **simple random variable** relative to $\mathcal{A}$.*

**Notation 5.17** *Given an algebra, $\mathcal{A}\subset 2^\Omega$, let $\mathbb{S}(\mathcal{A})$ denote the collection of simple random variables from $\Omega$ to $\mathbb{C}$. For example if $A\in\mathcal{A}$, then $1_A\in\mathbb{S}(\mathcal{A})$ is a measurable simple function.*

**Lemma 5.18.** *Let $\mathcal{A}\subset 2^\Omega$ be an algebra, then;*

1. *$\mathbb{S}(\mathcal{A})$ is a sub-algebra of all functions from $\Omega$ to $\mathbb{C}$.*
2. *$f:\Omega\to\mathbb{C}$, is a $\mathcal{A}$ – simple random variable iff there exists $\alpha_i\in\mathbb{C}$ and $A_i\in\mathcal{A}$ for $1\le i\le n$ for some $n\in\mathbb{N}$ such that*

$$f=\sum_{i=1}^n \alpha_i 1_{A_i}. \tag{5.13}$$

3. *For any function, $F:\mathbb{C}\to\mathbb{C}$, $F\circ f\in\mathbb{S}(\mathcal{A})$ for all $f\in\mathbb{S}(\mathcal{A})$. In particular, $|f|\in\mathbb{S}(\mathcal{A})$ if $f\in\mathbb{S}(\mathcal{A})$.*

**Proof.** 1. Let us observe that $1_\Omega=1$ and $1_\emptyset=0$ are in $\mathbb{S}(\mathcal{A})$. If $f,g\in\mathbb{S}(\mathcal{A})$ and $c\in\mathbb{C}\setminus\{0\}$, then

$$\{f+cg=\lambda\}=\bigcup_{a,b\in\mathbb{C}:a+cb=\lambda}(\{f=a\}\cap\{g=b\})\in\mathcal{A} \tag{5.14}$$

and

$$\{f\cdot g=\lambda\}=\bigcup_{a,b\in\mathbb{C}:a\cdot b=\lambda}(\{f=a\}\cap\{g=b\})\in\mathcal{A} \tag{5.15}$$

from which it follows that $f+cg$ and $f\cdot g$ are back in $\mathbb{S}(\mathcal{A})$.

2. Since $\mathbb{S}(\mathcal{A})$ is an algebra, every $f$ of the form in Eq. (5.13) is in $\mathbb{S}(\mathcal{A})$. Conversely if $f\in\mathbb{S}(\mathcal{A})$ it follows by definition that $f=\sum_{\alpha\in f(\Omega)}\alpha 1_{\{f=\alpha\}}$ which is of the form in Eq. (5.13).

3. If $F:\mathbb{C}\to\mathbb{C}$, then

$$F\circ f=\sum_{\alpha\in f(\Omega)}F(\alpha)\cdot 1_{\{f=\alpha\}}\in\mathbb{S}(\mathcal{A}).$$

■

**Exercise 5.4 ($\mathcal{A}$ – measurable simple functions).** As in Example 4.19, let $\mathcal{A}\subset 2^\Omega$ be a finite algebra and $\{B_1,\dots,B_k\}$ be the partition of $\Omega$ associated to $\mathcal{A}$. Show that a function, $f:\Omega\to\mathbb{C}$, is an $\mathcal{A}$ – simple function iff $f$ is constant on $B_i$ for each $i$. Thus any $\mathcal{A}$ – simple function is of the form,

$$f=\sum_{i=1}^k \alpha_i 1_{B_i} \tag{5.16}$$

for some $\alpha_i\in\mathbb{C}$.

**Corollary 5.19.** *Suppose that $\Lambda$ is a finite set and $Z:\Omega\to\Lambda$ is a function. Let*

$$\mathcal{A}:=\mathcal{A}(Z):=Z^{-1}\left(2^\Lambda\right):=\left\{Z^{-1}(E):E\subset\Lambda\right\}.$$

*Then $\mathcal{A}$ is an algebra and $f:\Omega\to\mathbb{C}$ is an $\mathcal{A}$ – simple function iff $f=F\circ Z$ for some function $F:\Lambda\to\mathbb{C}$.*

**Proof.** For $\lambda\in\Lambda$, let

$$A_\lambda:=\{Z=\lambda\}=\{\omega\in\Omega:Z(\omega)=\lambda\}.$$

The $\{A_\lambda\}_{\lambda\in\Lambda}$ is the partition of $\Omega$ determined by $\mathcal{A}$. Therefore $f$ is an $\mathcal{A}$ – simple function iff $f|_{A_\lambda}$ is constant for each $\lambda\in\Lambda$. Let us denote this constant value by $F(\lambda)$. As $Z=\lambda$ on $A_\lambda$, $F:\Lambda\to\mathbb{C}$ is a function such that $f=F\circ Z$.

Conversely if $F:\Lambda\to\mathbb{C}$ is a function and $f=F\circ Z$, then $f=F(\lambda)$ on $A_\lambda$, i.e. $f$ is an $\mathcal{A}$ – simple function. ■

### 5.3.1 The algebraic structure of simple functions*

**Definition 5.20.** *A **simple function algebra**, $\mathbb{S}$, is a subalgebra[2] of the bounded complex functions on $\Omega$ such that $1\in\mathbb{S}$ and each function in $\mathbb{S}$ is a simple function. If $\mathbb{S}$ is a simple function algebra, let*

$$\mathcal{A}(\mathbb{S}):=\{A\subset\Omega:1_A\in\mathbb{S}\}.$$

*(It is easily checked that $\mathcal{A}(\mathbb{S})$ is a sub-algebra of $2^\Omega$.)*

**Lemma 5.21.** *Suppose that $\mathbb{S}$ is a simple function algebra, $f\in\mathbb{S}$ and $\alpha\in f(\Omega)$ – the range of $f$. Then $\{f=\alpha\}\in\mathcal{A}(\mathbb{S})$.*

**Proof.** Let $\{\lambda_i\}_{i=0}^n$ be an enumeration of $f(\Omega)$ with $\lambda_0=\alpha$. Then

$$g:=\left[\prod_{i=1}^n(\alpha-\lambda_i)\right]^{-1}\prod_{i=1}^n(f-\lambda_i 1)\in\mathbb{S}.$$

Moreover, we see that $g=0$ on $\cup_{i=1}^n\{f=\lambda_i\}$ while $g=1$ on $\{f=\alpha\}$. So we have shown $g=1_{\{f=\alpha\}}\in\mathbb{S}$ and therefore that $\{f=\alpha\}\in\mathcal{A}(\mathbb{S})$. ■

[2] To be more explicit we are assuming that $\mathbb{S}$ is a linear subspace of bounded functions which is closed under pointwise multiplication.

**Exercise 5.5.** Continuing the notation introduced above:

1. Show $\mathcal{A}(\mathbb{S})$ is an algebra of sets.
2. Show $\mathbb{S}(\mathcal{A})$ is a simple function algebra.
3. Show that the map
$$\mathcal{A} \in \left\{\text{Algebras} \subset 2^{\Omega}\right\} \to \mathbb{S}(\mathcal{A}) \in \{\textbf{simple function algebras on } \Omega\}$$
is bijective and the map, $\mathbb{S} \to \mathcal{A}(\mathbb{S})$, is the inverse map.

## 5.4 Simple Integration

**Definition 5.22 (Simple Integral).** *Suppose now that $P$ is a finitely additive probability measure on an algebra $\mathcal{A} \subset 2^{\Omega}$. For $f \in \mathbb{S}(\mathcal{A})$ the **integral or expectation,** $\mathbb{E}(f) = \mathbb{E}_P(f)$, is defined by*
$$\mathbb{E}_P(f) = \int_{\Omega} f dP = \sum_{y\in\mathbb{C}} yP(f=y). \tag{5.17}$$

*Example 5.23.* Suppose that $A \in \mathcal{A}$, then
$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A). \tag{5.18}$$

*Remark 5.24.* Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (3.1) by
$$P(A) = \lim_{N\to\infty} \frac{1}{N} \sum 1_A(\omega(k))$$
where $\omega(k) \in \Omega$ was the result of the $k^{\text{th}}$ "independent" experiment. If we use this interpretation back in Eq. (5.17) we arrive at,
$$\begin{aligned}\mathbb{E}(f) = \sum_{y\in\mathbb{C}} yP(f=y) &= \sum_{y\in\mathbb{C}} y \cdot \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{y\in\mathbb{C}} y \sum_{k=1}^{N} 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \sum_{y\in\mathbb{C}} f(\omega(k)) \cdot 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(\omega(k)).\end{aligned}$$
Thus informally, $\mathbb{E}f$ should represent the limiting average of the values of $f$ over many "independent" experiments. We will come back to this later when we study the strong law of large numbers.

We now extend the above notion to general positively finite additive measures, $\mu$.

**Definition 5.25 (Simple Integral).** *Suppose now that $\mu$ is a finitely additive measure on an algebra $\mathcal{A} \subset 2^{\Omega}$ and let $\mathbb{S}_+(\mathcal{A})$ denote the $[0,\infty]$ – valued $\mathcal{A}$ – simple functions and for $f \in \mathbb{S}_+(\mathcal{A})$ we let*
$$\mathbb{E}_\mu(f) = \int_{\Omega} f d\mu = \sum_{y\in[0,\infty]} y\mu(f=y). \tag{5.19}$$
*[For $f \in \mathbb{S}_+(\mathcal{A})$, $\mathbb{E}_\mu(f) = \infty$ is allowed and we use the convention that $0 \cdot \infty = 0 = \infty \cdot 0$.] Further let*
$$\mathbb{S}_\mu(\mathcal{A}) := \{f \in \mathbb{S}(\mathcal{A}) : \mu(f \neq 0) < \infty\}$$
*and for $f \in \mathbb{S}_\mu(\mathcal{A})$ we let*
$$\mathbb{E}_\mu(f) = \int_{\Omega} f d\mu = \sum_{y\in\mathbb{C}\setminus\{0\}} y\mu(f=y).$$

It is easy to verify that $\mathbb{S}_\mu(\mathcal{A})$ is a subspace of $\mathbb{S}(\mathcal{A})$ and that $\mathbb{S}_+(\mathcal{A})$ is closed under addition and scalar multiplication by non-negative constants.

**Proposition 5.26.** *The expectation operator, $\mathbb{E} = \mathbb{E}_\mu : \mathbb{S}_+(\mathcal{A}) \to [0,\infty]$, satisfies:*

*1. If $f \in \mathbb{S}_+(\mathcal{A})$ and $\lambda \in [0,\infty]$, then*
$$\mathbb{E}(\lambda f) = \lambda\mathbb{E}(f). \tag{5.20}$$
*2. If $f, g \in \mathbb{S}_+(\mathcal{A})$, then $f + g \in \mathbb{S}_+(\mathcal{A})$ and*
$$\mathbb{E}(f+g) = \mathbb{E}(g) + \mathbb{E}(f). \tag{5.21}$$
*3. If $f, g \in \mathbb{S}_+(\mathcal{A})$ and $f \le g$, then $\mathbb{E}(f) \le \mathbb{E}(g)$.*

**Proof.**

1. If $\lambda = 0$ or $\lambda = \infty$ we have $\mathbb{E}(0 \cdot f) = 0 = 0 \cdot \mathbb{E}(f)$ and $\mathbb{E}[\infty \cdot f] = \infty$ iff $\mu(f \neq 0) > 0$ iff $\mathbb{E}f > 0$ iff $\infty \cdot \mathbb{E}f = \infty$ respectively. If $0 < \lambda < \infty$ and $f \in \mathbb{S}_+(\mathcal{A})$, then
$$\begin{aligned}\mathbb{E}(\lambda f) = \sum_{y\in[0,\infty]} y\ \mu(\lambda f = y) &= \sum_{y\in[0,\infty]} y\ \mu(f = y/\lambda) \\ &= \sum_{z\in[0,\infty]} \lambda z\ \mu(f=z) = \lambda\mathbb{E}(f).\end{aligned}$$

2. Writing $\{f=a,g=b\}$ for $f^{-1}(\{a\})\cap g^{-1}(\{b\})$, we have
$$\begin{aligned}\mathbb{E}(f+g) &= \sum_{z\in[0,\infty]} z\ \mu(f+g=z)\\ &= \sum_{z\in[0,\infty]} z\ \mu\left(\sum_{a+b=z}\{f=a,\ g=b\}\right)\\ &= \sum_{z\in[0,\infty]} z \sum_{a+b=z}\mu\left(\{f=a,\ g=b\}\right)\\ &= \sum_{z\in[0,\infty]}\sum_{a+b=z}(a+b)\,\mu\left(\{f=a,\ g=b\}\right)\\ &= \sum_{a,b}(a+b)\,\mu\left(\{f=a,\ g=b\}\right).\end{aligned}$$
But
$$\begin{aligned}\sum_{a,b} a\mu\left(\{f=a,\ g=b\}\right) &= \sum_a a\sum_b \mu\left(\{f=a,\ g=b\}\right)\\ &= \sum_a a\mu\left(\cup_b\{f=a,\ g=b\}\right)\\ &= \sum_a a\mu\left(\{f=a\}\right)=\mathbb{E}f\end{aligned}$$
and similarly,
$$\sum_{a,b} b\mu\left(\{f=a,\ g=b\}\right)=\mathbb{E}g.$$
Equation (5.21) is now a consequence of the last three displayed equations.
3. Let $h := g - f1_{\{f<\infty\}} \geq 0$. Notice that $\{h=\infty\}=\{g=\infty\}$ and for $x\in[0,\infty)$,
$$\{h=x\}=\cup_{a-b=x}\{g=a\}\cap\{f=b\}\in\mathcal{A}$$
so that $h\in\mathbb{S}_+(\mathcal{A})$. As (is easily verified) $g=f+h$ it follows that
$$\mathbb{E}_\mu g=\mathbb{E}_\mu f+\mathbb{E}_\mu h\geq\mathbb{E}_\mu f.$$
**Alternative proof.** If $\lambda\geq 0$, then
$$\begin{aligned}\lambda\cdot\mu(g=\lambda) &= \sum_{0\leq y\leq\lambda}\lambda\mu(g=\lambda, f=y)\\ &\geq \sum_{0\leq y\leq\lambda} y\mu(g=\lambda,f=y)\\ &= \sum_{0\leq y} y\mu(g=\lambda,f=y).\end{aligned}$$
Summing this inequality on $\lambda\geq 0$ then gives,
$$\begin{aligned}\mathbb{E}_\mu g &\geq \sum_{0\leq y,\lambda} y\mu(g=\lambda,f=y)\\ &= \sum_{0\leq y} y\sum_{0\leq\lambda}\mu(g=\lambda,f=y)\\ &= \sum_{0\leq y} y\mu(f=y)=\mathbb{E}_\mu f.\end{aligned}$$
■

**Proposition 5.27.** *The expectation operator, $\mathbb{E}=\mathbb{E}_\mu:\mathbb{S}_\mu(\mathcal{A})\to\mathbb{C}$, satisfies:*

1. *If $f\in\mathbb{S}_\mu(\mathcal{A})$ and $\lambda\in\mathbb{C}$, then*
$$\mathbb{E}(\lambda f)=\lambda\mathbb{E}(f). \tag{5.22}$$
2. *If $f,g\in\mathbb{S}_\mu(\mathcal{A})$, then*
$$\mathbb{E}(f+g)=\mathbb{E}(g)+\mathbb{E}(f). \tag{5.23}$$
*Items 1. and 2. say that $\mathbb{E}(\cdot)$ is a linear functional on $\mathbb{S}_\mu(\mathcal{A})$.*
3. *If $f=\sum_{j=1}^N\lambda_j 1_{A_j}$ for some $\lambda_j\in\mathbb{C}$ and some $A_j\in\mathcal{A}$ with $\mu(A_j)<\infty$, then*
$$\mathbb{E}(f)=\sum_{j=1}^N\lambda_j\mu(A_j). \tag{5.24}$$
4. *$\mathbb{E}$ is **positive**, i.e. $\mathbb{E}(f)\geq 0$ for all $0\leq f\in\mathbb{S}_\mu(\mathcal{A})$. More generally, if $f,g\in\mathbb{S}_\mu(\mathcal{A})$ and $f\leq g$, then $\mathbb{E}(f)\leq\mathbb{E}(g)$.*
5. *For all $f\in\mathbb{S}_\mu(\mathcal{A})$,*
$$|\mathbb{E}f|\leq\mathbb{E}|f|. \tag{5.25}$$

**Proof.**

1. If $\lambda\neq 0$, then
$$\begin{aligned}\mathbb{E}(\lambda f) &= \sum_{y\in\mathbb{C}} y\ \mu(\lambda f=y)=\sum_{y\in\mathbb{C}} y\ \mu(f=y/\lambda)\\ &= \sum_{z\in\mathbb{C}}\lambda z\ \mu(f=z)=\lambda\mathbb{E}(f).\end{aligned}$$
The case $\lambda=0$ is trivial.

2. Writing $\{f=a,g=b\}$ for $f^{-1}(\{a\})\cap g^{-1}(\{b\})$, then

$$\begin{aligned}\mathbb{E}(f+g)&=\sum_{z\in\mathbb{C}} z\ \mu(f+g=z)\\&=\sum_{z\in\mathbb{C}} z\ \mu\left(\sum_{a+b=z}\{f=a,\ g=b\}\right)\\&=\sum_{z\in\mathbb{C}} z\sum_{a+b=z}\mu\left(\{f=a,\ g=b\}\right)\\&=\sum_{z\in\mathbb{C}}\sum_{a+b=z}(a+b)\,\mu\left(\{f=a,\ g=b\}\right)\\&=\sum_{a,b}(a+b)\,\mu\left(\{f=a,\ g=b\}\right).\end{aligned}$$

But

$$\begin{aligned}\sum_{a,b} a\mu\left(\{f=a,\ g=b\}\right)&=\sum_a a\sum_b\mu\left(\{f=a,\ g=b\}\right)\\&=\sum_a a\mu\left(\cup_b\{f=a,\ g=b\}\right)\\&=\sum_a a\mu\left(\{f=a\}\right)=\mathbb{E}f\end{aligned}$$

and similarly,

$$\sum_{a,b} b\mu\left(\{f=a,\ g=b\}\right)=\mathbb{E}g.$$

Equation (5.23) is now a consequence of the last three displayed equations.

3. If $f=\sum_{j=1}^N\lambda_j 1_{A_j}$, then

$$\mathbb{E}f=\mathbb{E}\left[\sum_{j=1}^N\lambda_j 1_{A_j}\right]=\sum_{j=1}^N\lambda_j\mathbb{E}1_{A_j}=\sum_{j=1}^N\lambda_j\mu\left(A_j\right).$$

4. If $f\geq 0$ then

$$\mathbb{E}(f)=\sum_{a\geq 0} a\mu(f=a)\geq 0$$

and if $f\leq g$, then $g-f\geq 0$ so that

$$\mathbb{E}\left(g\right)-\mathbb{E}\left(f\right)=\mathbb{E}\left(g-f\right)\geq 0.$$

5. By the triangle inequality,

$$|\mathbb{E}f|=\left|\sum_{\lambda\in\mathbb{C}}\lambda\mu\left(f=\lambda\right)\right|\leq\sum_{\lambda\in\mathbb{C}}|\lambda|\,\mu\left(f=\lambda\right)=\mathbb{E}\left|f\right|,$$

wherein the last equality we have used Eq. (5.24) and the fact that $|f|=\sum_{\lambda\in\mathbb{C}}|\lambda|\,1_{f=\lambda}$.

∎

*Remark 5.28.* If $\Omega$ is a finite set and $\mathcal{A}=2^\Omega$, then

$$f\left(\cdot\right)=\sum_{\omega\in\Omega} f\left(\omega\right)1_{\{\omega\}}$$

and hence

$$\mathbb{E}_P f=\sum_{\omega\in\Omega} f\left(\omega\right)P\left(\{\omega\}\right).$$

**Exercise 5.6.** Let $P$ is a finitely additive probability measure on an algebra $\mathcal{A}\subset 2^\Omega$ and for $A,B\in\mathcal{A}$ let $\rho\left(A,B\right):=P\left(A\bigtriangleup B\right)$ where $A\bigtriangleup B=\left(A\setminus B\right)\cup\left(B\setminus A\right)$. Show;

1. $\rho\left(A,B\right)=\mathbb{E}\left|1_A-1_B\right|$ and then use this (or not) to show
2. $\rho\left(A,C\right)\leq\rho\left(A,B\right)+\rho\left(B,C\right)$ for all $A,B,C\in\mathcal{A}$.

Remark: it is now easy to see that $\rho:\mathcal{A}\times\mathcal{A}\to[0,1]$ satisfies the axioms of a metric except for the condition that $\rho\left(A,B\right)=0$ does not imply that $A=B$ but only that $A=B$ modulo a set of probability zero.

**Lemma 5.29 (Chebyshev's Inequality).** *Suppose that $f\in\mathbb{S}(\mathcal{A})$, $\varepsilon>0$, and $p>0$, then*

$$P\left(\{|f|\geq\varepsilon\}\right)=\mathbb{E}\left[1_{|f|\geq\varepsilon}\right]\leq\mathbb{E}\left[\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}\right]\leq\varepsilon^{-p}\mathbb{E}\left|f\right|^p. \tag{5.26}$$

**Proof.** First observe that

$$|f|^p=\sum_{\lambda\in\mathbb{C}}|\lambda|^p\,1_{\{f=\lambda\}}$$

is a simple random variable and $\{|f|\geq\varepsilon\}=\sum_{|\lambda|\geq\varepsilon}\{f=\lambda\}\in\mathcal{A}$. Therefore $\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}$ is a simple random variable and since,

$$1_{|f|\geq\varepsilon}\leq\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}\leq\varepsilon^{-p}\left|f\right|^p,$$

the estimates in Eq. (5.26) follow from item 4. of Proposition 5.27. ∎

**Lemma 5.30 (Inclusion Exclusion Formula).** *If $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots, M$ such that $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, then*

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right). \tag{5.27}$$

**Proof.** This may be proved inductively from Eq. (5.2). We will give a different and perhaps more illuminating proof here. Let $A := \cup_{n=1}^M A_n$.

Since $A^c = \left(\cup_{n=1}^M A_n\right)^c = \cap_{n=1}^M A_n^c$, we have

$$\begin{aligned} 1 - 1_A = 1_{A^c} &= \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M \left(1 - 1_{A_n}\right) \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1}} \cdots 1_{A_{n_k}} \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}} \end{aligned}$$

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}. \tag{5.28}$$

Integrating this identity with respect to $\mu$ gives Eq. (5.27). ■

*Remark 5.31.* The following identity holds even when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$,

$$\begin{aligned} \mu\left(\cup_{n=1}^M A_n\right) &+ \sum_{k=2 \text{ \& } k \text{ even}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right) \\ &= \sum_{k=1 \text{ \& } k \text{ odd}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right). \end{aligned} \tag{5.29}$$

This can be proved by moving every term with a negative sign on the right side of Eq. (5.28) to the left side and then integrate the resulting identity. Alternatively, Eq. (5.29) follows directly from Eq. (5.27) if $\mu\left(\cup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (5.29) are infinite.

To better understand Eq. (5.28), consider the case $M = 3$ where,

$$\begin{aligned} 1 - 1_A &= \left(1 - 1_{A_1}\right)\left(1 - 1_{A_2}\right)\left(1 - 1_{A_3}\right) \\ &= 1 - \left(1_{A_1} + 1_{A_2} + 1_{A_3}\right) \\ &\quad + 1_{A_1}1_{A_2} + 1_{A_1}1_{A_3} + 1_{A_2}1_{A_3} - 1_{A_1}1_{A_2}1_{A_3} \end{aligned}$$

so that

$$1_{A_1 \cup A_2 \cup A_3} = 1_{A_1} + 1_{A_2} + 1_{A_3} - \left(1_{A_1 \cap A_2} + 1_{A_1 \cap A_3} + 1_{A_2 \cap A_3}\right) + 1_{A_1 \cap A_2 \cap A_3}$$

Here is an alternate proof of Eq. (5.28). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1 \cap \cdots \cap A_m$ and $\omega \notin A_{m+1} \cup \cdots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (5.28) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

$$\begin{aligned} &\sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le m} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} \\ &= 1 - \sum_{k=0}^m (-1)^k (1)^{n-k} \binom{m}{k} \\ &= 1 - (1-1)^m = 1. \end{aligned}$$

This verifies Eq. (5.28) since $1_{\cup_{n=1}^M A_n}(\omega) = 1$.

*Example 5.32 (Coincidences).* Let $\Omega$ be the set of permutations (think of card shuffling), $\omega : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, and define $P(A) := \frac{\#(A)}{n!}$ to be the uniform distribution (Haar measure) on $\Omega$. We wish to compute the probability of the event, $B$, that a random permutation fixes some index $i$. To do this, let $A_i := \{\omega \in \Omega : \omega(i) = i\}$ and observe that $B = \cup_{i=1}^n A_i$. So by the Inclusion Exclusion Formula, we have

$$P(B) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < i_2 < i_3 < \cdots < i_k \le n} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right).$$

Since

$$\begin{aligned} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) &= P\left(\{\omega \in \Omega : \omega(i_1) = i_1, \ldots, \omega(i_k) = i_k\}\right) \\ &= \frac{(n-k)!}{n!} \end{aligned}$$

and

$$\#\{1 \le i_1 < i_2 < i_3 < \cdots < i_k \le n\} = \binom{n}{k},$$

we find

$$P(B) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}. \tag{5.30}$$

For large $n$ this gives,

$$P(B) = -\sum_{k=1}^{n} \frac{1}{k!} (-1)^k \cong 1 - \sum_{k=0}^{\infty} \frac{1}{k!} (-1)^k = 1 - e^{-1} \cong 0.632.$$

*Example 5.33 (Expected number of coincidences).* Continue the notation in Example 5.32. We now wish to compute the expected number of fixed points of a random permutation, $\omega$, i.e. how many cards in the shuffled stack have not moved on average. To this end, let

$$X_i = 1_{A_i}$$

and observe that

$$N(\omega) = \sum_{i=1}^{n} X_i(\omega) = \sum_{i=1}^{n} 1_{\omega(i)=i} = \#\{i : \omega(i) = i\}.$$

denote the number of fixed points of $\omega$. Hence we have

$$\mathbb{E}N = \sum_{i=1}^{n} \mathbb{E}X_i = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = 1.$$

Let us check the above formulas when $n = 3$. In this case we have

| $\omega$ | $N(\omega)$ |
|---|---|
| 1 2 3 | 3 |
| 1 3 2 | 1 |
| 2 1 3 | 1 |
| 2 3 1 | 0 |
| 3 1 2 | 0 |
| 3 2 1 | 1 |

and so

$$P(\exists \text{ a fixed point}) = \frac{4}{6} = \frac{2}{3} \cong 0.67 \cong 0.632$$

while

$$\sum_{k=1}^{3} (-1)^{k+1} \frac{1}{k!} = 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

and

$$\mathbb{E}N = \frac{1}{6}(3 + 1 + 1 + 0 + 0 + 1) = 1.$$

The next three problems generalize the results above. The following notation will be used throughout these exercises.

1. $(\Omega, \mathcal{A}, P)$ is a finitely additive probability space, so $P(\Omega) = 1$,
2. $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots, n$,
3. $N(\omega) := \sum_{i=1}^{n} 1_{A_i}(\omega) = \#\{i : \omega \in A_i\}$, and
4. $\{S_k\}_{k=1}^{n}$ are given by

$$S_k := \sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \sum_{\Lambda \subset \{1,2,\ldots,n\} \ni |\Lambda| = k} P(\cap_{i \in \Lambda} A_i).$$

**Exercise 5.7.** For $1 \le k \le n$, show;

1. (as functions on $\Omega$) that
$$\binom{N}{k} = \sum_{\Lambda \subset \{1,2,\ldots,n\} \ni |\Lambda| = k} 1_{\cap_{i \in \Lambda} A_i}, \tag{5.31}$$
where by definition
$$\binom{m}{k} = \begin{cases} 0 & \text{if } \quad k > m \\ \frac{m!}{k! \cdot (m-k)!} & \text{if } 1 \le k \le m \\ 1 & \text{if } \quad k = 0 \end{cases}. \tag{5.32}$$
2. Conclude from Eq. (5.31) that for all $z \in \mathbb{C}$,
$$(1+z)^N = 1 + \sum_{k=1}^{n} z^k \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}} \tag{5.33}$$
provided $(1+z)^0 = 1$ even when $z = -1$.
3. Conclude from Eq. (5.31) that $S_k = \mathbb{E}_P \binom{N}{k}$.

**Exercise 5.8.** Taking expectations of Eq. (5.33) implies,

$$\mathbb{E}\left[(1+z)^N\right] = 1 + \sum_{k=1}^{n} S_k z^k. \tag{5.34}$$

Show that setting $z = -1$ in Eq. (5.34) gives another proof of the inclusion exclusion formula. **Hint:** use the definition of the expectation to write out $\mathbb{E}\left[(1+z)^N\right]$ explicitly.

**Exercise 5.9.** Let $1 \le m \le n$. In this problem you are asked to compute the probability that there are exactly $m$ – coincidences. Namely you should show,

$$\begin{aligned} P(N=m) &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} S_k \\ &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \sum_{1 \le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned}$$

**Hint:** differentiate Eq. (5.34) $m$ times with respect to $z$ and then evaluate the result at $z=-1$. In order to do this you will find it useful to derive formulas for;

$$\frac{d^m}{dz^m}|_{z=-1} (1+z)^n \text{ and } \frac{d^m}{dz^m}|_{z=-1} z^k.$$

*Example 5.34.* Let us again go back to Example 5.33 where we computed,

$$S_k = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Therefore it follows from Exercise 5.9 that

$$\begin{aligned} P(\exists \text{ exactly m fixed points}) &= P(N=m) \\ &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \frac{1}{m!} \sum_{k=m}^{n} (-1)^{k-m} \frac{1}{(k-m)!}. \end{aligned}$$

So if $n$ is much bigger than $m$ we may conclude that

$$P(\exists \text{ exactly m fixed points}) \cong \frac{1}{m!} e^{-1}.$$

Let us check our results are consistent with Eq. (5.30);

$$\begin{aligned} P(\exists \text{ a fixed point}) &= \sum_{m=1}^{n} P(N=m) \\ &= \sum_{m=1}^{n} \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{1 \le m \le k \le n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^{n} \left[ \sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} - (-1)^k \right] \frac{1}{k!} \\ &= -\sum_{k=1}^{n} (-1)^k \frac{1}{k!} \end{aligned}$$

wherein we have used,

$$\sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} = (1-1)^k = 0.$$

### 5.4.1 Appendix: Bonferroni Inequalities

In this appendix (see Feller Volume 1., p. 106-111 for more) we want to discuss what happens if we truncate the sums in the inclusion exclusion formula of Lemma 5.30. In order to do this we will need the following lemma whose combinatorial meaning was explained to me by Jeff Remmel.

**Lemma 5.35.** *Let $n \in \mathbb{N}_0$ and $0 \le k \le n$, then*

$$\sum_{l=0}^{k} (-1)^l \binom{n}{l} = (-1)^k \binom{n-1}{k} 1_{n>0} + 1_{n=0}. \tag{5.35}$$

**Proof.** The case $n=0$ is trivial. We give two proofs for when $n \in \mathbb{N}$.

**First proof.** Just use induction on $k$. When $k=0$, Eq. (5.35) holds since $1=1$. The induction step is as follows,

$$\sum_{l=0}^{k+1}(-1)^l\binom{n}{l}=(-1)^k\binom{n-1}{k}+\binom{n}{k+1}$$
$$=\frac{(-1)^{k+1}}{(k+1)!}\left[n(n-1)\dots(n-k)-(k+1)(n-1)\dots(n-k)\right]$$
$$=\frac{(-1)^{k+1}}{(k+1)!}\left[(n-1)\dots(n-k)(n-(k+1))\right]=(-1)^{k+1}\binom{n-1}{k+1}.$$

**Second proof.** Let $\Omega=\{1,2,\dots,n\}$ and observe that

$$m_k:=\sum_{l=0}^{k}(-1)^l\binom{n}{l}=\sum_{l=0}^{k}(-1)^l\cdot\#\left(\Lambda\in 2^\Omega:\#(\Lambda)=l\right)$$
$$=\sum_{\Lambda\in 2^\Omega:\ \#(\Lambda)\le k}(-1)^{\#(\Lambda)} \tag{5.36}$$

Define $T:2^\Omega\to 2^\Omega$ by

$$T(S)=\begin{cases}S\cup\{1\} \text{ if } 1\notin S\\ S\setminus\{1\} \text{ if } 1\in S\end{cases}.$$

Observe that $T$ is a bijection of $2^\Omega$ such that $T$ takes even cardinality sets to odd cardinality sets and visa versa. Moreover, if we let

$$\Gamma_k:=\left\{\Lambda\in 2^\Omega:\#(\Lambda)\le k \text{ and } 1\in\Lambda \text{ if } \#(\Lambda)=k\right\},$$

then $T(\Gamma_k)=\Gamma_k$ for all $1\le k\le n$. Since

$$\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(T(\Lambda))}=\sum_{\Lambda\in\Gamma_k}-(-1)^{\#(\Lambda)}$$

we see that $\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=0$. Using this observation with Eq. (5.36) implies

$$m_k=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}+\sum_{\#(\Lambda)=k\ \&\ 1\notin\Lambda}(-1)^{\#(\Lambda)}=0+(-1)^k\binom{n-1}{k}.$$

■

**Corollary 5.36 (Bonferroni Inequalitites).** *Let $\mu:\mathcal{A}\to[0,\mu(\Omega)]$ be a finitely additive finite measure on $\mathcal{A}\subset 2^\Omega$, $A_n\in\mathcal{A}$ for $n=1,2,\dots,M$, $N:=\sum_{n=1}^M 1_{A_n}$, and*

$$S_k:=\sum_{1\le i_1<\dots<i_k\le M}\mu\left(A_{i_1}\cap\dots\cap A_{i_k}\right)=\mathbb{E}_\mu\left[\binom{N}{k}\right].$$



*Then for $1\le k\le M$,*

$$\mu\left(\cup_{n=1}^M A_n\right)=\sum_{l=1}^{k}(-1)^{l+1}S_l+(-1)^k\mathbb{E}_\mu\left[\binom{N-1}{k}\right]. \tag{5.37}$$

*This leads to the Bonferroni inequalities;*

$$\mu\left(\cup_{n=1}^M A_n\right)\le\sum_{l=1}^{k}(-1)^{l+1}S_l \text{ if } k \text{ is odd}$$
*and*
$$\mu\left(\cup_{n=1}^M A_n\right)\ge\sum_{l=1}^{k}(-1)^{l+1}S_l \text{ if } k \text{ is even.}$$

**Proof.** By Lemma 5.35,

$$\sum_{l=0}^{k}(-1)^l\binom{N}{l}=(-1)^k\binom{N-1}{k}1_{N>0}+1_{N=0}.$$

Therefore integrating this equation with respect to $\mu$ gives,

$$\mu(\Omega)+\sum_{l=1}^{k}(-1)^l S_l=\mu(N=0)+(-1)^k\mathbb{E}_\mu\binom{N-1}{k}$$

and therefore,

$$\mu\left(\cup_{n=1}^M A_n\right)=\mu(N>0)=\mu(\Omega)-\mu(N=0)$$
$$=-\sum_{l=1}^{k}(-1)^l S_l+(-1)^k\mathbb{E}_\mu\binom{N-1}{k}.$$

The Bonferroni inequalities are a simple consequence of Eq. (5.37) and the fact that

$$\binom{N-1}{k}\ge 0 \implies \mathbb{E}_\mu\binom{N-1}{k}\ge 0.$$

■

## 5.5 Finitely Additive Measures on $\mathbb{R}^d$ and $[0,1]^d$

(Riesz Markov theorem is in Section 9.5 below.) Let us begin by describing finite measure on $\mathcal{A}=\mathcal{A}\left(\mathbb{R}^d\right)$ – the subalgebra of $2^{\mathbb{R}^d}$ generated by $\mathcal{E}\dot{\times}\dots\dot{\times}\mathcal{E}$ where $\mathcal{E}$ is the semi-algebra of subsets of $\mathbb{R}$ defined by

$$\mathcal{E} := \{\mathbb{R} \cap (a,b] : -\infty \le a < b \le \infty\} \subset 2^{\mathbb{R}}.$$

For $\mathbf{a}, \mathbf{b} \in \bar{\mathbb{R}}^d$ we let $(\mathbf{a}, \mathbf{b}] := ((a_1, b_1] \times \cdots \times (a_d, b_d]) \cap \mathbb{R}^d$ and so with this notation,

$$\mathcal{E} \dot{\times} \ldots \dot{\times} \mathcal{E} = \left\{ (\mathbf{a}, \mathbf{b}] : \mathbf{a}, \mathbf{b} \in \bar{\mathbb{R}}^d \text{ with } \mathbf{a} \le \mathbf{b} \right\}.$$

**Notation 5.37** *If $a, b \in \bar{\mathbb{R}}^d$ and $\gamma \subset \{1, 2, \ldots, d\}$, let $a_\gamma \times b_{\gamma^c}$ be the point in $\bar{\mathbb{R}}^d$ defined by*

$$(a_\gamma \times b_{\gamma^c})_j := \begin{cases} a_j \text{ if } j \in \gamma \\ b_j \text{ if } j \in \gamma^c \end{cases}.$$

**Lemma 5.38.** *For all $a, b \in \bar{\mathbb{R}}^d$ with $a \le b$ we have,*

$$1_{(a,b]} = \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} 1_{(-\infty, a_\gamma \times b_{\gamma^c}]}. \tag{5.38}$$

**Proof.** If $x \in \bar{\mathbb{R}}^d$, then

$$\begin{aligned} 1_{(a,b]}(x) &= \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = \prod_{i=1}^d \left[ 1_{(-\infty,b_i]}(x_i) - 1_{(-\infty,a_i]}(x_i) \right] \\ &= \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} \prod_{i \in \gamma} 1_{(-\infty,a_i]}(x_i) \cdot \prod_{j \in \gamma^c} 1_{(-\infty,b_j]}(x_j) \\ &= \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} 1_{(-\infty, a_\gamma \times b_{\gamma^c}]}(x). \end{aligned}$$

■

**Corollary 5.39.** *Suppose that $V$ is a vector space, $\mu : \mathcal{A}(\mathbb{R}^d) \to V$ is a finitely additive measure, and for $b \in \bar{\mathbb{R}}^d$ let $F(b) := \mu((-\infty, b])$. Then for all $a, b \in \bar{\mathbb{R}}^d$ with $a \le b$,*

$$\mu((a,b]) = \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}). \tag{5.39}$$

**Proof.** The result follows directly by integrating Eq. (5.38) relative to $\mu$ while using $\mu((-\infty, a_\gamma \times b_{\gamma^c}]) = F(a_\gamma \times b_{\gamma^c})$ for all $\gamma \subset \{1, 2, \ldots, d\}$. ■

*Remark 5.40.* Corollary 5.39 may be understood using the inclusion exclusion formula. For example, when $d = 2$ we have

$$(a,b] = S_b \setminus \left[ S_{(a_1,b_2)} \cup S_{(b_1,a_2)} \right]$$

and hence

$$\begin{aligned} \mu((a,b]) &= \mu(S_b) - \mu\left(S_{(a_1,b_2)} \cup S_{(b_1,a_2)}\right) \\ &= \mu(S_b) - \left[ \mu\left(S_{(a_1,b_2)}\right) + \mu\left(S_{(b_1,a_2)}\right) - \mu\left(S_{(a_1,b_2)} \cap S_{(b_1,a_2)}\right) \right] \\ &= \mu(S_b) - \mu\left(S_{(a_1,b_2)}\right) - \mu\left(S_{(b_1,a_2)}\right) + \mu(S_a) \\ &= F(b) - F(a_1, b_1) - F(b_1, a_2) + F(a), \end{aligned}$$

wherein the third equality we have used $S_{(a_1,b_2)} \cap S_{(b_1,a_2)} = S_a$).

We will give a converse to this corollary in Proposition 5.41 below. To help motivate the proof of Proposition 5.41, observe if $\mu$ is a finitely additive measure on $\mathcal{A}(\mathbb{R}^d)$, then to each $b \in \bar{\mathbb{R}}$, $\mu_b(C) := \mu(C \times (-\infty, b])$ for $C \in \mathcal{A}(\mathbb{R}^{d-1})$ defines a finitely additive measure on $\mathcal{A}(\mathbb{R}^{d-1})$. Moreover, if $a, b \in \bar{\mathbb{R}}$ with $a \le b$ we further have $\mu(C \times (a,b]) = \mu_b(C) - \mu_a(C)$.

**Proposition 5.41.** *To every function, $F : \bar{\mathbb{R}}^d \to V$, there exists a unique finitely additive measure $(\mu_F)$ on $\mathcal{A}(\mathbb{R}^d)$ such that*

$$\mu_F((a,b]) = \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}) \tag{5.40}$$

*for all $a, b \in \bar{\mathbb{R}}^d$ with $a \le b$.*

**Proof.** The proof for $d = 1$ is completely analogous to the proof of Proposition 5.8 and so will be omitted. To each $a = (a_1, \ldots, a_d) \in \bar{\mathbb{R}}^d$, let $a' := (a_1, \ldots, a_{d-1})$ so that $a = (a', a_d)$. With this notation in hand we proceed to the induction step.

It wo
to incl
explic
leavin

Suppose that $d \ge 2$ and we have proved the proposition when $d$ is replaced by $d-1$. Then for $c \in \bar{\mathbb{R}}$ let $\mu_c = \mu_{F(\cdot,c)}$ be the unique finitely additive measure on $\mathcal{A}(\mathbb{R}^{d-1})$ (guaranteed to exist by the induction hypothesis) such that for all $a', b' \in \bar{\mathbb{R}}^{d-1}$ with $a' \le b'$ we have

$$\mu_c((a',b']) = \sum_{\gamma \subset \{1,2,\ldots,d-1\}} (-1)^{|\gamma|} F(a'_\gamma \times b'_{\gamma^c}, c). \tag{5.41}$$

By Proposition 5.14, there is a unique finitely additive measure $(\mu_F)$ on $\mathcal{A}(\mathbb{R}^d) = \mathcal{A}(\mathbb{R}^{d-1}) \otimes \mathcal{A}(\mathbb{R})$ such that

$$\mu_F(C \times (a_d, b_d]) = \mu_{b_d}(C) - \mu_{a_d}(C) \ \forall\, C \in \mathcal{A}(\mathbb{R}^{d-1}) \text{ and } -\infty \le a_d \le b_d \le \infty.$$

Letting $a \le b$ with $a, b \in \bar{\mathbb{R}}^d$ and using Eq. (5.41) it is not to hard to show,

$$\begin{aligned} \mu_F((a,b]) &= \sum_{\gamma \subset \{1,2,\ldots,d-1\}} (-1)^{|\gamma|} \left[ F(a'_\gamma \times b'_{\gamma^c}, b) - F(a'_\gamma \times b'_{\gamma^c}, a) \right] \\ &= \sum_{\gamma \subset \{1,2,\ldots,d\}} (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}). \end{aligned}$$

The point is that the subsets, $\gamma \subset \{1, 2, \dots, d\}$, split into two types; those which contain $d$ and those which do not. ■

We now wish to prove the analogous result with $\mathbb{R}^d$ replaced by $[0,1]^d$. We begin with some needed additional notation.

**Notation 5.42** *For $a, b \in [0,1]^d$ with $a \le b$, let*

$$\{a, b] := \{_1 a_1, b_1] \times \cdots \times \{_d a_d, b_d]$$

*where $\{_i = ($ if $a_i > 0$ and is either $($ or $[$ when $a_i = 0$.*

*Remark 5.43.* The point of this notation is that $[0,1]^d$ is the product of $[0,1]$ with itself $d$ – times and if $\mathcal{E} \subset 2^{[0,1]}$ are the sets of the form $[0, b]$ with $0 \le b \le 1$ or $(a, b]$ with $0 \le a < b \le 1$, then $\mathcal{A}\left([0,1]^d\right)$ is generated by the semi-algebra, $\mathcal{E} \dot{\times} \dots \dot{\times} \mathcal{E}$, consisting of sets of the form $\{a, b]$ with $\mathbf{0} \le a \le b \le \mathbf{1}$.

**Lemma 5.44.** *For all $a, b \in [0,1]^d$ with $a \le b$ we have,*

$$1_{\{a,b]} = \sum_{\gamma \subset \{1,2,\dots,d\}} w_\gamma (-1)^{|\gamma|} 1_{(-\infty, a_\gamma \times b_{\gamma^c}]} \tag{5.42}$$

*where*

$$w_\gamma = \begin{cases} 1 & \text{if} \quad \{_i = ( \text{ for all } i \in \gamma \\ 0 & \text{otherwise.} \end{cases}$$

*In particular,*

$$\sum_{|\gamma| \text{ even}} w_\gamma 1_{[0, a_\gamma \times b_{\gamma^c}]} = \sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0, a_\gamma \times b_{\gamma^c}]} + 1_{\{a,b]} \ge \sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0, a_\gamma \times b_{\gamma^c}]}. \tag{5.43}$$

**Proof.** Let

$$w_i := \begin{cases} 1 & \text{if} \quad \{_i = ( \\ 0 & \text{otherwise.} \end{cases}$$

and observe that $w_\gamma = \prod_{i \in \gamma} w_i$. Hence if $x \in [0,1]^d$, then

$$\begin{aligned} 1_{\{a,b]}(x) &= \prod_{i=1}^d 1_{\{_i a_i, b_i]}(x_i) = \prod_{i=1}^d \left[1_{[0,b_i]}(x_i) - w_i 1_{[0,a_i]}(x_i)\right] \\ &= \sum_{\gamma \subset \{1,2,\dots,d\}} (-1)^{|\gamma|} \prod_{i \in \gamma} w_i 1_{[0,a_i]}(x_i) \cdot \prod_{j \in \gamma^c} 1_{(-\infty, b_j]}(x_j) \\ &= \sum_{\gamma \subset \{1,2,\dots,d\}} (-1)^{|\gamma|} w_\gamma 1_{[0, a_\gamma \times b_{\gamma^c}]}(x). \end{aligned}$$

■



**Corollary 5.45.** *Suppose that $V$ is a vector space, $\mu : \mathcal{A}\left([0,1]^d\right) \to V$ is a finitely additive measure, and for $b \in [0,1]^d$ let $F(b) := \mu([0,b])$. Then for all $a, b \in [0,1]^d$ with $a \le b$,*

$$\mu(\{a, b]) = \sum_{\gamma \subset \{1,2,\dots,d\}} w_\gamma (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}). \tag{5.44}$$

**Proof. First proof.** The result follows directly by integrating Eq. (5.42) relative to $\mu$ while using $\mu([0, a_\gamma \times b_{\gamma^c}]) = F(a_\gamma \times b_{\gamma^c})$ for all $\gamma \subset \{1, 2, \dots, d\}$.

**Second proof.** We carry this proof out only in the case $d = 2$. We have

$$\begin{aligned} \mu(\{a, b]) &= \mu(\{_1 a_1, b_1] \times \{_2 a_2, b_2]) = \mu(\{_1 a_1, b_1] \times [0, b_2]) - w_2 \mu(\{_1 a_1, b_1] \times [0, a_2]) \\ &= \mu([0, b_1]] \times [0, b_2]) - w_1 \mu([0, a_1] \times [0, b_2]) \\ &\quad - w_2 [\mu([0, b_1] \times [0, a_2]) - w_1 \mu([0, a_1] \times [0, a_2])] \\ &= F(b) - w_1 F(a_1, b_2) - w_2 F(b_1, a_2) - w_1 w_2 F(a). \end{aligned}$$

The general scheme of this proof would then follow by induction. ■

The next result is the converse of this corollary.

**Proposition 5.46.** *To every function, $F : [0,1]^d \to V$, there exists a unique finitely additive measure $(\mu_F)$ on $\mathcal{A}([0,1]^d)$ such that Eq. (5.44) holds for all $a, b \in [0,1]^d$ with $a \le b$.*

**Proof.** The proof for $d = 1$ is completely analogous to the proof of Proposition 5.8 and so will be omitted. Now suppose that $d \ge 2$ and we have proved the Proposition when $d$ is replaced by $d - 1$. Then for $c \in [0,1]$ let $\mu_c = \mu_{F(\cdot, c)}$ be the unique finitely additive measure on $\mathcal{A}\left([0,1]^{d-1}\right)$ (guaranteed to exist by the induction hypothesis) such that for all $a', b' \in [0,1]^{d-1}$ with $a' \le b'$ we have

$$\mu_c(\{a', b']) = \sum_{\gamma \subset \{1,2,\dots,d-1\}} w_\gamma (-1)^{|\gamma|} F(a'_\gamma \times b'_{\gamma^c}, c). \tag{5.45}$$

By Proposition 5.14, there is a unique finitely additive measure $(\mu_F)$ on $\mathcal{A}([0,1]^d) = \mathcal{A}\left([0,1]^{d-1}\right) \otimes \mathcal{A}([0,1])$ such that

$$\mu_F(C \times \{a_d, b_d]) = \mu_{b_d}(C) - w_d \mu_{a_d}(C) \ \forall\, C \in \mathcal{A}\left([0,1]^{d-1}\right) \ \& \ 0 \le a_d \le b_d \le 1.$$

Letting $a \le b$ with $a, b \in [0,1]^d$ and using Eq. (5.41) it is not to hard to show,

$$\begin{aligned} \mu_F(\{a, b]) &= \sum_{\gamma \subset \{1,2,\dots,d-1\}} w_\gamma (-1)^{|\gamma|} \left[F(a'_\gamma \times b'_{\gamma^c}, b) - w_d F(a'_\gamma \times b'_{\gamma^c}, a)\right] \\ &= \sum_{\gamma \subset \{1,2,\dots,d\}} w_\gamma (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}). \end{aligned}$$

The point is that the subsets, $\gamma \subset \{1, 2, \ldots, d\}$, split into two types; those which contain $d$ and those which do not. ■

## 5.6 Appendix: Riemann Stieljtes integral

In this subsection, let $\Omega$ be a set, $\mathcal{A} \subset 2^{\Omega}$ be an algebra of sets, and $P := \mu : \mathcal{A} \to [0, \infty)$ be a finitely additive measure with $\mu(\Omega) < \infty$. As above let

$$\mathbb{E}_{\mu} f := \int_{\Omega} f d\mu := \sum_{\lambda \in \mathbb{C}} \lambda \mu(f = \lambda) \ \forall \ f \in \mathbb{S}(\mathcal{A}). \tag{5.46}$$

**Notation 5.47** *For any function, $f : \Omega \to \mathbb{C}$ let $\|f\|_u := \sup_{\omega \in \Omega} |f(\omega)|$. Further, let $\bar{\mathbb{S}} := \overline{\mathbb{S}(\mathcal{A})}$ denote those functions, $f : \Omega \to \mathbb{C}$ such that there exists $f_n \in \mathbb{S}(\mathcal{A})$ such that $\lim_{n \to \infty} \|f - f_n\|_u = 0$.*

**Exercise 5.10 (Do not hand in).** Prove the following statements.

1. For all $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_{\mu} f| \le \mu(\Omega) \|f\|_u. \tag{5.47}$$
2. If $f \in \bar{\mathbb{S}}$ and $f_n \in \mathbb{S} := \mathbb{S}(\mathcal{A})$ such that $\lim_{n \to \infty} \|f - f_n\|_u = 0$, show $\lim_{n \to \infty} \mathbb{E}_{\mu} f_n$ exists. Also show that defining $\mathbb{E}_{\mu} f := \lim_{n \to \infty} \mathbb{E}_{\mu} f_n$ is well defined, i.e. you must show that $\lim_{n \to \infty} \mathbb{E}_{\mu} f_n = \lim_{n \to \infty} \mathbb{E}_{\mu} g_n$ if $g_n \in \mathbb{S}$ such that $\lim_{n \to \infty} \|f - g_n\|_u = 0$.
3. Show $\mathbb{E}_{\mu} : \bar{\mathbb{S}} \to \mathbb{C}$ is still linear and still satisfies Eq. (5.47).
4. Show $|f| \in \bar{\mathbb{S}}$ if $f \in \bar{\mathbb{S}}$ and that Eq. (5.25) is still valid, i.e. $|\mathbb{E}_{\mu} f| \le \mathbb{E}_{\mu} |f|$ for all $f \in \bar{\mathbb{S}}$.

Let us now specialize the above results to the case where $\Omega = [0, T]$ for some $T < \infty$. Let $\mathcal{S} := \{(a, b] : 0 \le a \le b \le T\} \cup \{0\}$ which is easily seen to be a semi-algebra. The following proposition is fairly straightforward and will be left to the reader.

**Proposition 5.48 (Riemann Stieljtes integral).** *Let $F : [0, T] \to \mathbb{R}$ be an increasing function, then;*

1. *there exists a unique finitely additive measure, $\mu_F$, on $\mathcal{A} := \mathcal{A}(\mathcal{S})$ such that $\mu_F((a, b]) = F(b) - F(a)$ for all $0 \le a \le b \le T$ and $\mu_F(\{0\}) = 0$. (In fact one could allow for $\mu_F(\{0\}) = \lambda$ for any $\lambda \ge 0$, but we would then have to write $\mu_{F,\lambda}$ rather than $\mu_F$.)*
2. *Show $C([0, 1], \mathbb{C}) \subset \overline{\mathbb{S}(\mathcal{A})}$. More precisely, suppose $\pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of $[0, T]$ and $c = (c_1, \ldots, c_n) \in [0, T]^n$ with $t_{i-1} \le c_i \le t_i$ for each $i$. Then for $f \in C([0, 1], \mathbb{C})$, let*
$$f_{\pi,c} := f(0) 1_{\{0\}} + \sum_{i=1}^{n} f(c_i) 1_{(t_{i-1}, t_i]}. \tag{5.48}$$
*Show that $\|f - f_{\pi,c}\|_u$ is small provided, $|\pi| := \max\{|t_i - t_{i-1}| : i = 1, 2, \ldots, n\}$ is small.*
3. *Using the above results, show*
$$\int_{[0,T]} f d\mu_F = \lim_{|\pi| \to 0} \sum_{i=1}^{n} f(c_i) (F(t_i) - F(t_{i-1}))$$
*where the $c_i$ may be chosen arbitrarily subject to the constraint that $t_{i-1} \le c_i \le t_i$.*

It is customary to write $\int_0^T f dF$ for $\int_{[0,T]} f d\mu_F$. This integral satisfies the estimates,

$$\left| \int_{[0,T]} f d\mu_F \right| \le \int_{[0,T]} |f| \, d\mu_F \le \|f\|_u (F(T) - F(0)) \ \forall \ f \in \overline{\mathbb{S}(\mathcal{A})}.$$

When $F(t) = t$,

$$\int_0^T f dF = \int_0^T f(t) \, dt,$$

is the usual Riemann integral.

**Exercise 5.11.** Let $a \in (0, T)$, $\lambda > 0$, and

$$G(x) = \lambda \cdot 1_{x \ge a} = \begin{cases} \lambda \text{ if } x \ge a \\ 0 \text{ if } x < a \end{cases}.$$

1. Explicitly compute $\int_{[0,T]} f d\mu_G$ for all $f \in C([0, 1], \mathbb{C})$.
2. If $F(x) = x + \lambda \cdot 1_{x \ge a}$ describe $\int_{[0,T]} f d\mu_F$ for all $f \in C([0, 1], \mathbb{C})$. **Hint:** if $F(x) = G(x) + H(x)$ where $G$ and $H$ are two increasing functions on $[0, T]$, show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G + \int_{[0,T]} f d\mu_H.$$

**Exercise 5.12.** Suppose that $F, G : [0, T] \to \mathbb{R}$ are two increasing functions such that $F(0) = G(0)$, $F(T) = G(T)$, and $F(x) \ne G(x)$ for at most countably many points, $x \in (0, T)$. Show

$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G \text{ for all } f \in C([0, 1], \mathbb{C}). \tag{5.49}$$

**Note well**, given $F(0) = G(0)$, $\mu_F = \mu_G$ on $\mathcal{A}$ iff $F = G$.

One of the points of the previous exercise is to show that Eq. (5.49) holds when $G(x) := F(x+)$ – the right continuous version of $F$. The exercise applies since and increasing function can have at most countably many jumps, see Remark 23.18. So if we only want to integrate continuous functions, we may always assume that $F : [0,T] \to \mathbb{R}$ is right continuous.

## 5.7 Tonelli and Fubini's Theorem I

In the last part of this section we will extend some of the above ideas to more general "finitely additive measure spaces." A **finitely additive measure space** is a triple, $(X, \mathcal{A}, \mu)$, where $X$ is a set, $\mathcal{A} \subset 2^X$ is an algebra, and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Let $(Y, \mathcal{B}, \nu)$ be another finitely additive measure space. Further let $\mu \odot \nu$ be the product measure on $\mathcal{A} \odot \mathcal{B}$ as described in Example

**Theorem 5.49 (Tonelli's Theorem).** *If $f \in \mathbb{S}_+(\mathcal{A} \odot \mathcal{B})$ then for each $x \in X$, $f(x,\cdot) \in \mathbb{S}_+(\mathcal{B})$ and $X \ni x \to \int_Y f(x,y)\,d\nu(y)$ is in $\mathbb{S}_+(\mathcal{A})$ and moreover,*

$$\int_{X\times Y} f(x,y)\,d(\mu\odot\nu)(x,y) = \int_X \left[\int_Y f(x,y)\,d\nu(y)\right] d\mu(x).$$

*Similarly, for each $y \in Y$, $f(\cdot,y) \in \mathbb{S}_+(\mathcal{A})$ and $Y \ni y \to \int_X f(x,y)\,d\mu(x)$ is in $\mathbb{S}_+(\mathcal{B})$ and moreover,*

$$\int_{X\times Y} f(x,y)\,d(\mu\odot\nu)(x,y) = \int_Y \left[\int_X f(x,y)\,d\mu(x)\right] d\nu(y).$$

**Proof.** By the usual arguments it suffices to assume $f = 1_{A\times B}$ for some $(A,B) \in \mathcal{A}\times\mathcal{B}$ in which case the above results are trivial. ∎

**Theorem 5.50 (Product Measure and Fubini's Theorem).** *Assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ for simplicity. Then there is a unique finitely additive measure, $\mu\odot\nu$, on $\mathcal{A}\odot\mathcal{B}$ such that $\mu\odot\nu(A\times B) = \mu(A)\,\nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover if $f \in \mathbb{S}(\mathcal{A}\odot\mathcal{B})$ then;*

1. *$y \to f(x,y)$ is in $\mathbb{S}(\mathcal{B})$ for all $x \in X$ and $x \to f(x,y)$ is in $\mathbb{S}(\mathcal{A})$ for all $y \in Y$.*
2. *$x \to \int_Y f(x,y)\,d\nu(y)$ is in $\mathbb{S}(\mathcal{A})$ and $y \to \int_X f(x,y)\,d\mu(x)$ is in $\mathbb{S}(\mathcal{B})$.*
3. *we have,*

$$\int_X \left[\int_Y f(x,y)\,d\nu(y)\right] d\mu(x) = \int_{X\times Y} f(x,y)\,d(\mu\odot\nu)(x,y) = \int_Y \left[\int_X f(x,y)\,d\mu(x)\right] d\nu(y).$$

*We will refer to $\mu\odot\nu$ as the **product measure** of $\mu$ and $\nu$.*

**Proof.** According to Eq. (5.10),

$$1_C(x,y) = \sum_{i=1}^n 1_{A_i\times B_i}(x,y) = \sum_{i=1}^n 1_{A_i}(x)\,1_{B_i}(y)$$

from which it follows that $1_C(x,\cdot) \in \mathbb{S}(\mathcal{B})$ for each $x \in X$ and

$$\int_Y 1_C(x,y)\,d\nu(y) = \sum_{i=1}^n 1_{A_i}(x)\,\nu(B_i).$$

It now follows from this equation that $x \to \int_Y 1_C(x,y)\,d\nu(y) \in \mathbb{S}(\mathcal{A})$ and that

$$\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

Similarly one shows that

$$\int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

In particular this shows that we may define

$$(\mu\odot\nu)(C) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i)$$

and with this definition we have,

$$\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) = (\mu\odot\nu)(C) = \int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y).$$

From either of these representations it is easily seen that $\mu \odot \nu$ is a finitely additive measure on $\mathcal{A} \odot \mathcal{B}$ with the desired properties. Moreover, we have already verified the Theorem in the special case where $f = 1_C$ with $C \in \mathcal{A} \odot \mathcal{B}$. Since the general element, $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$, is a linear combination of such functions, it is easy to verify using the linearity of the integral and the fact that $\mathbb{S}(\mathcal{A})$ and $\mathbb{S}(\mathcal{B})$ are vector spaces that the theorem is true in general. ■

*Example 5.51.* Suppose that $f \in \mathbb{S}(\mathcal{A})$ and $g \in \mathbb{S}(\mathcal{B})$. Let $f \otimes g(x,y) := f(x)g(y)$. Since we have,

$$\begin{aligned} f \otimes g(x,y) &= \left(\sum_a a 1_{f=a}(x)\right)\left(\sum_b b 1_{g=b}(y)\right) \\ &= \sum_{a,b} ab 1_{\{f=a\}\times\{g=b\}}(x,y) \end{aligned}$$

it follows that $f \otimes g \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$. Moreover, using Fubini's Theorem 5.50 it follows that

$$\int_{X\times Y} f \otimes g \, d(\mu \odot \nu) = \left[\int_X f \, d\mu\right]\left[\int_Y g \, d\nu\right].$$

*Remark 5.52.* We can at this point now use the obvious generalizations of these results to prove the classical Weierstrass approximation theorems. [See Theorem 5.70 and more generally Section 5.11.]

## 5.8 Bernstein Polynomials and the Classical Weierstrass Approximation Theorem

**Theorem 5.53 (Weierstrass Approximation Theorem via Bernstein's Polynomials.).** *Suppose that $f \in C([0,1], \mathbb{C})$ and*

$$p_n(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

*Then*

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - p_n(x)| = 0.$$

**Proof.** Let $\Lambda = \{0,1\}$ and for $x \in [0,1]$ let $\mu_x : 2^\Lambda \to [0,1]$ be the probability measure,

$$\mu_x = x\delta_1 + (1-x)\delta_0.$$

We then set $\Omega = \Lambda^n$, $X_i(\omega_1,\dots,\omega_n) = \omega_i$ for all $\omega \in \Omega$ and $1 \le i \le n$, and $P_x = \mu_x^{\odot n} : 2^\Omega \to [0,1]$ be the $n$ – fold product measure of $\mu_x$ with itself. If we let $S := X_1 + \dots + X_n$, then using $\{\omega\} = \{\omega_1\} \times \dots \times \{\omega_n\}$ for all $\omega \in \Omega$, we find

$$P_x(\{\omega\}) = \mu_x(\{\omega_1\})\dots\mu_x(\{\omega_n\}) = \prod_{j=1}^n x^{\omega_j} \cdot (1-x)^{1-\omega_j} = x^{S(\omega)} \cdot (1-x)^{1-S(\omega)}.$$

If $0 \le k \le n$, then $\#\{\omega \in \Omega : S(\omega) = k\} = \binom{n}{k}$, it follows that

$$P_x(S = k) = \binom{n}{k} x^k (1-x)^{n-k}$$

and therefore (writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$) we have

$$\mathbb{E}_x\left[f\left(\frac{S}{n}\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} = p_n(x). \tag{5.50}$$

Given any $\varepsilon > 0$, using Eq. (5.50) we find,

$$\begin{aligned} |p_n(x) - f(x)| &= \left|\mathbb{E}_x f\left(\frac{S}{n}\right) - f(x)\right| = \left|\mathbb{E}_x\left[f\left(\frac{S}{n}\right) - f(x)\right]\right| \\ &\le \mathbb{E}_x\left|f\left(\frac{S}{n}\right) - f(x)\right| \\ &= \mathbb{E}_x\left[\left|f\left(\frac{S}{n}\right) - f(x)\right| : |S - x| \ge \varepsilon\right] \\ &\quad + \mathbb{E}_x\left[\left|f\left(\frac{S}{n}\right) - f(x)\right| : |S - x| < \varepsilon\right] \\ &\le 2M \cdot P_x\left(\left|\frac{S}{n} - x\right| \ge \varepsilon\right) + \delta(\varepsilon) \end{aligned} \tag{5.51}$$

where

$$\begin{aligned} M &:= \max_{y\in[0,1]} |f(y)| \text{ and} \\ \delta(\varepsilon) &:= \sup\{|f(y) - f(x)| : x,y \in [0,1] \text{ and } |y - x| \le \varepsilon\} \end{aligned}$$

is the modulus of continuity of $f$.

By Chebyshev's inequality,

$$P_x\left(\left|\frac{S}{n} - x\right| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left|\frac{S}{n} - x\right|^2 = \frac{1}{\varepsilon^2}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n (X_j - x)\right|^2.$$

For $i \ne j$ we have by Fubini's theorem that

$$\mathbb{E}_x\left[(X_i - x)(X_j - x)\right] = \left[\int_\Lambda (\lambda - x)\, d\mu_x(\lambda)\right]^2 = 0$$

since

$$\int_\Lambda \lambda d\mu_x(\lambda) = 1 \cdot x + 0 \cdot (1 - x) = x.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n (X_j - x)\right|^2 &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}_x\left[(X_i - x)(X_j - x)\right] \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_x\left[(X_i - x)^2\right] \\ &= \frac{1}{n}\cdot\int_\Lambda (\lambda - x)^2\, d\mu_x(\lambda) \\ &= \frac{1}{n}\left[(1-x)^2 x + (0-x)^2(1-x)\right] = \frac{1}{n}x(1-x) \le \frac{1}{4n}.\end{aligned}$$

Combining these results shows proves

$$P_x\left(\left|\frac{S}{n} - x\right| \ge \varepsilon\right) \le \frac{1}{4n\varepsilon^2}$$

which is a simple version of the weak law of large numbers. Using this estimate back in Eq. (5.51) gives

$$\limsup_{n\to\infty}\max_{x\in[0,1]}|p_n(x) - f(x)| \le \limsup_{n\to\infty}\left[2M\cdot\frac{1}{4n\varepsilon^2} + \delta(\varepsilon)\right] = \delta(\varepsilon).$$

This completes the proof, since by uniform continuity of $f$, $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. ■

The previous theorem has the following multi-dimensional generalization.

**Theorem 5.54.** *Let $f \in C\left([0,1]^d, \mathbb{C}\right)$ and for $n \in \mathbb{N}$ let $p_n(\mathbf{x})$ be the polynomial function defined by*

$$p_n(\mathbf{x}) = \sum_{\mathbf{k}\in\{0,1,\dots,n\}^d} f\left(\frac{\mathbf{k}}{n}\right)\prod_{j=1}^d \binom{n}{k_j} x_j^{k_j}(1 - x_j)^{n-k_j} \;\; \forall\, \mathbf{x} = (x_1, \dots, x_d) \in [0,1]^d.$$

*Then $p_n \to f$ uniformly, i.e.*

$$\lim_{n\to\infty}\max_{\mathbf{x}\in[0,1]^d}|p_n(\mathbf{x}) - f(\mathbf{x})| = 0.$$

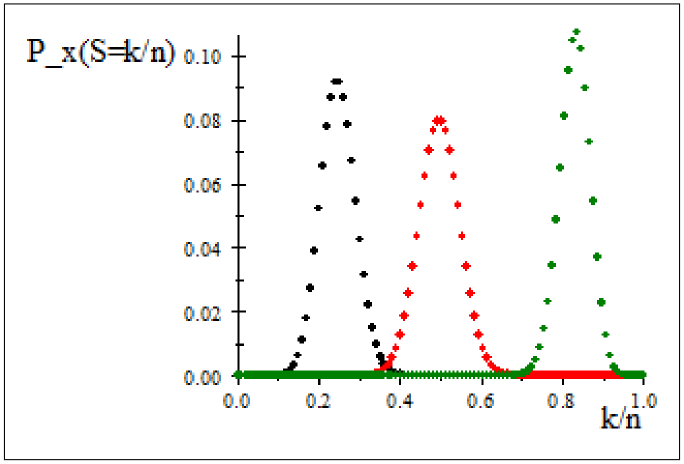



**Fig. 5.1.** Plots of $P_x(S_n = k/n)$ versus $k/n$ for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

**Proof.** Let $\mathbf{x} \in [0,1]^d$ and set $\mathbb{P}_\mathbf{x} := P_{x_1} \odot \cdots \odot P_{x_d}$ which is a probability measure on $\Omega^d$ where $\Omega = \Lambda^n$ as above. Further let $\pi_j : \Omega^d \to \Omega$ be projection onto the $j^{th}$ factor and set $S_j := S \circ \pi_j$ for $1 \le j \le d$ and $\mathbf{S} := (S_1, \dots, S_d) : \Omega^d \to \{0, 1, \dots, n\}^d$. For $\mathbf{k} \in \{0, 1, \dots, n\}^d$ we have by the definition of product measure that,

$$\begin{aligned}\mathbb{P}_\mathbf{x}(\mathbf{S} = \mathbf{k}) &= \mathbb{P}_\mathbf{x}(\{S = k_1\} \times \cdots \times \{S = k_d\}) \\ &= \prod_{j=1}^d P_{x_j}(S = k_j) = \prod_{j=1}^d \binom{n}{k_j} x_j^{k_j}(1 - x_j)^{n-k_j}\end{aligned}$$

and therefore,

$$\mathbb{E}_\mathbf{x} f\left(\frac{\mathbf{S}}{n}\right) = \sum_{\mathbf{k}\in\{0,1,\dots,n\}^d} f\left(\frac{\mathbf{k}}{n}\right)\prod_{j=1}^d \binom{n}{k_j} x_j^{k_j}(1 - x_j)^{n-k_j} = p_n(\mathbf{x}).$$

Given $\varepsilon > 0$ we further have,

$$\begin{aligned}\mathbb{P}_\mathbf{x}\left(\left\|\frac{\mathbf{S}}{n} - x\right\| \ge \varepsilon\right) &\le \frac{1}{\varepsilon^2}\mathbb{E}_\mathbf{x}\left\|\frac{\mathbf{S}}{n} - \mathbf{x}\right\|^2 \\ &= \frac{1}{\varepsilon^2}\mathbb{E}_\mathbf{x}\sum_{i=1}^d\left(\frac{S_i}{n} - x_i\right)^2 \\ &= \frac{1}{n\varepsilon^2}\sum_{i=1}^d x_i(1 - x_i) \le \frac{d}{n4\varepsilon^2}.\end{aligned}$$

Thus working as in the proof of Theorem 5.53, we find,

$$\begin{aligned}|p_n(\mathbf{x}) - f(\mathbf{x})| &\le 2M \cdot \mathbb{P}_{\mathbf{x}}\left(\left\|\frac{\mathbf{S}}{n} - x\right\| \ge \varepsilon\right) + \delta(\varepsilon) \\ &\le \frac{dM}{2\varepsilon^2 n} + \delta(\varepsilon)\end{aligned}$$

where

$$M := \max_{\mathbf{y}\in[0,1]^d} |f(\mathbf{y})| \text{ and}$$
$$\delta(\varepsilon) := \sup\left\{|f(\mathbf{y}) - f(\mathbf{x})| : \mathbf{x}, \mathbf{y} \in [0,1]^2 \text{ and } \|\mathbf{y} - \mathbf{x}\| \le \varepsilon\right\}$$

is the modulus of continuity of $f$. Therefore,

$$\limsup_{n\to\infty} \max_{\mathbf{x}\in[0,1]^d} |p_n(\mathbf{x}) - f(\mathbf{x})| \le \limsup_{n\to\infty}\left[\frac{dM}{2\varepsilon^2 n} + \delta(\varepsilon)\right] = \delta(\varepsilon) \to 0 \text{ as } \varepsilon \downarrow 0.$$

■

**Corollary 5.55 (Weierstrass Approximation Theorem).** *Suppose that $K = [a_1, b_1] \times \dots [a_d, b_d]$ with $-\infty < a_i < b_i < \infty$ is a compact rectangle in $\mathbb{R}^d$. Then for every $f \in C(K, \mathbb{C})$, there exists polynomials $q_n$ on $\mathbb{R}^d$ such that $q_n \to f$ uniformly on $K$.*

**Proof.** Let $T : [0,1]^d \to K$ be the invertible affine transformation,

$$T(\mathbf{x}) = (a_1 + x_1(b_1 - a_1), \dots, a_d + x_d(b_d - a_d)).$$

By Theorem 5.54, there exists polynomial functions, $p_n$ on $[0,1]^d$ such that $p_n \to f \circ T$ uniformly on $[0,1]^d$. It is now a simple matter to verify that $q_n := p_n \circ T^{-1}$ are polynomial function on $K$ such that $q_n \to f$ uniformly on $K$. ■

Here is a version of the complex Weierstrass approximation theorem.

**Theorem 5.56 (Complex Weierstrass Approximation Theorem).** *Suppose that $K \subset \mathbb{C}$ is a compact rectangle. Then there exists polynomials in $(z = x + iy, \bar{z} = x - iy)$, $p_n(z, \bar{z})$ for $z \in \mathbb{C}$, such that $\sup_{z\in K} |q_n(z, \bar{z}) - f(z)| \to 0$ as $n \to \infty$ for every $f \in C(K, \mathbb{C})$.*

**Proof.** The mapping $(x, y) \in \mathbb{R} \times \mathbb{R} \to z = x + iy \in \mathbb{C}$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z+\bar{z}}{2}$ and $y = \frac{z-\bar{z}}{2i}$. Therefore under this identification any polynomial $p(x, y)$ on $\mathbb{R} \times \mathbb{R}$ may be written as a polynomial $q$ in $(z, \bar{z})$, namely

$$q(z, \bar{z}) = p\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right).$$

Conversely a polynomial $q$ in $(z, \bar{z})$ may be thought of as a polynomial $p$ in $(x, y)$, namely $p(x, y) = q(x + iy, x - iy)$. Hence the result now follows from Theorem 5.71. ■

*Example 5.57.* Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z, \bar{z})$ restricted to $S^1$. Then $\mathcal{A}$ is dense in $C(S^1)$. To prove this first observe if $f \in C(S^1)$ then $F(z) = |z| f\left(\frac{z}{|z|}\right)$ for $z \neq 0$ and $F(0) = 0$ defines $F \in C(\mathbb{C})$ such that $F|_{S^1} = f$. By applying Theorem 5.56 to $F$ restricted to a compact rectangle containing $S^1$ we may find $q_n(z, \bar{z})$ converging uniformly to $F$ on $K$ and hence on $S^1$. Since $\bar{z} = z^{-1}$ on $S^1$, we have shown polynomials in $z$ and $z^{-1}$ are dense in $C(S^1)$.

**Theorem 5.58 (Density of Trigonometric Polynomials).** *Any $2\pi$ – periodic continuous function, $f : \mathbb{R} \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form*

$$p(x) = \sum_{\lambda\in\Lambda} a_\lambda e^{i\lambda\cdot x}$$

*where $\Lambda$ is a finite subset of $\mathbb{Z}$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$.*

**Proof.** For $z \in S^1$, define $F(z) := f(\theta)$ where $\theta \in \mathbb{R}$ is chosen so that $z = e^{i\theta}$. Since $f$ is $2\pi$ – periodic, $F$ is well defined since if $\theta$ solves $e^{i\theta} = z$ then all other solutions are of the form $\{\theta + 2\pi n : n \in \mathbb{Z}\}$. Since the map $\theta \to e^{i\theta}$ is a local homeomorphism, i.e. for any $J = (a, b)$ with $b - a < 2\pi$, the map $\theta \in J \stackrel{\phi}{\to} \tilde{J} := \{e^{i\theta} : \theta \in J\} \subset S^1$ is a homeomorphism, it follows that $F(z) = f \circ \phi^{-1}(z)$ for $z \in \tilde{J}$. This shows $F$ is continuous when restricted to $\tilde{J}$. Since such sets cover $S^1$, it follows that $F$ is continuous.

By Example 5.57, the polynomials in $z$ and $\bar{z} = z^{-1}$ are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z, \bar{z}) = \sum_{0\le m,n\le N} a_{m,n} z^m \bar{z}^n$$

such that $|F(z) - p(z, \bar{z})| \le \varepsilon$ for all $z \in S^1$. Taking $z = e^{i\theta}$ then implies

$$\sup_\theta \left|f(\theta) - p\left(e^{i\theta}, e^{-i\theta}\right)\right| \le \varepsilon$$

where

$$p\left(e^{i\theta}, e^{-i\theta}\right) = \sum_{0\le m,n\le N} a_{m,n} e^{i(m-n)\theta}$$

is the desired trigonometry polynomial. ■

**Exercise 5.13.** Use Example 5.60 to show that any $2\pi$ – periodic continuous function, $g : \mathbb{R}^d \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

$$p(x) = \sum_{\lambda \in \Lambda} a_\lambda e^{i\lambda \cdot x}$$

where $\Lambda$ is a finite subset of $\mathbb{Z}^d$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$. **Hint:** start by showing there exists a unique continuous function, $f : (S^1)^d \to \mathbb{C}$ such that $f(e^{ix_1}, \dots, e^{ix_d}) = F(x)$ for all $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

**Exercise 5.14.** Suppose $f \in C(\mathbb{R}, \mathbb{C})$ is a $2\pi$ – periodic function (i.e. $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$) and

$$\int_0^{2\pi} f(x) e^{inx} dx = 0 \text{ for all } n \in \mathbb{Z},$$

show again that $f \equiv 0$. **Hint**: Use Exercise 5.13.

Theorem 5.56 has the following multi-dimensional generalization.

**Theorem 5.59 (Complex Weierstrass Approximation Theorem).** *Suppose that $K \subset \mathbb{C}^d \cong \mathbb{R}^d \times \mathbb{R}^d$ is a compact rectangle. Then there exists polynomials in $(z = x + iy, \bar{z} = x - iy)$, $p_n(z, \bar{z})$ for $z \in \mathbb{C}^d$, such that* $\sup_{z \in K} |q_n(z, \bar{z}) - f(z)| \to 0$ *as $n \to \infty$ for every $f \in C(K, \mathbb{C})$.*

**Proof.** The mapping $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \to z = x + iy \in \mathbb{C}^d$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z + \bar{z}}{2}$ and $y = \frac{z - \bar{z}}{2i}$. Therefore under this identification any polynomial $p(x, y)$ on $\mathbb{R}^d \times \mathbb{R}^d$ may be written as a polynomial $q$ in $(z, \bar{z})$, namely

$$q(z, \bar{z}) = p(\frac{z + \bar{z}}{2}, \frac{z - \bar{z}}{2i}).$$

Conversely a polynomial $q$ in $(z, \bar{z})$ may be thought of as a polynomial $p$ in $(x, y)$, namely $p(x, y) = q(x + iy, x - iy)$. Hence the result now follows from Theorem 5.79. ∎

*Example 5.60.* Let

$$(S^1)^d = \{\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{C}^d : |z_j| = 1 \; 1 \le j \le d\} \subset \mathbb{C}^d$$

$\mathcal{A}$ be the set of polynomials in $(\mathbf{z}, \bar{\mathbf{z}})$ restricted to $(S^1)^d$. Then $\mathcal{A}$ is dense in $C\left((S^1)^d\right)$. To prove this first observe if $f \in C\left((S^1)^d\right)$ then

$$F(\mathbf{z}) = |z_1| \dots |z_d| \, f\left(\frac{z_1}{|z_1|}, \dots, \frac{z_d}{|z_d|}\right) \text{ when } z_i \neq 0 \text{ for all } i$$

and $F(\mathbf{z}) = 0$ if $z_i = 0$ for some $i$ defines a continuous function on $\mathbb{C}^d$ such that $F|_K = f$. By applying Theorem 5.59 to $F$ restricted to a compact rectangle $K$ containing $(S^1)^d$ we may find $q_n(z, \bar{z})$ converging uniformly to $F$ on $K$ and hence on $(S^1)^d$. Since $\bar{z} = z^{-1}$ on $(S^1)^d$, we have also shown polynomials in $\mathbf{z}$ and $\mathbf{z}^{-1} := (z_1^{-1}, \dots, z_d^{-1})$ are dense in $C\left((S^1)^d\right)$.

## 5.9 Simple Independence and the Weak Law of Large Numbers

Informally, we say two experiments are independent if knowing the outcome of one of the experiments in no way influences the second experiment. As an example of independent experiments, suppose that one experiment is the outcome of spinning a roulette wheel and the second is the outcome of rolling a dice. We expect these two experiments will be independent. As an example of dependent experiments, suppose that dice roller now has two dice – one red and one black. The person rolling dice throws his black or red dice after the roulette ball has stopped and landed on either black or red respectively. If the black and the red dice are weighted differently, we expect that these two experiments are no longer independent. We begin with the notion of conditional probabilities which tries to capture the dependency of one event on another.

Throughout this section let $(\Omega, \mathcal{A}, P)$ be a finitely additive probability space. For events $A$ and $B \in \mathcal{A}$ we wish to know how likely the event $A$ is given that we know that $B$ has occurred. Informally (in the spirit of Chapter 3) we want $P(A|B) := \lim_{N\to\infty} P_N(A|B)$ where

$$\begin{aligned} P_N(A|B) &= \frac{\#\{k : 1 \le k \le N \text{ and } \omega(k) \in A \cap B\}}{\#\{k : 1 \le k \le N \text{ and } \omega(k) \in B\}} \\ &= \frac{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega(k) \in A \cap B\}}{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega(k) \in B\}} \\ &= \frac{P_N(A \cap B)}{P_N(B)}, \end{aligned}$$

which represents the frequency in the first $N$ trials that $A$ occurs given that we know that $B$ has occurred. [This is only defined when $P_N(B) > 0$, i.e. where $B$ has occurred during the first $N$ – trials.] As we explained, we expect $P(A \cap B) = \lim_{N\to\infty} P_N(A \cap B)$ and $P(B) = \lim_{N\to\infty} P_N(B)$ and all of this together leads to the following definition.

**Definition 5.61.** *If $B$ is a non-null event, i.e. $P(B) > 0$, define the* ***conditional probability of*** *$A$* ***given*** *$B$ by,*

$$P(A|B) := \frac{P(A\cap B)}{P(B)}.$$

*We further say that $A$ and $B$ are $(P)$ **independent events** if either $P(B) = 0$ or $P(B) > 0$ and $P(A|B) = P(A)$. Alternatively put, $A$ and $B$ are independent provided $P(A\cap B) = P(A)P(B)$.*

**Definition 5.62.** *Suppose that $S$ and $T$ are sets and $X : \Omega \to S$ and $Y : \Omega \to T$ are $\mathcal{A}$ – simple functions, i.e. $X$ and $Y$ have finite range and $\{X = s\}, \{Y = t\} \in \mathcal{A}$ for all $s \in S$ and $t \in T$. We say that $X$ and $Y$ are $(P)$ independent if $\{X = s\}$ and $\{Y = t\}$ are independent events for all $s \in S$ and $t \in T$.*

**Proposition 5.63.** *Suppose that $S$ and $T$ are sets and $X : \Omega \to S$ and $Y : \Omega \to T$ are $\mathcal{A}$ – simple functions. The following are equivalent;*

1. *$X$ and $Y$ are independent*
2. *$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ for all functions, $f : S \to \mathbb{R}$ and $g : T \to \mathbb{R}$.*
3. *$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all $A \subset S$ and $B \subset T$.*

**Proof.** We will prove 1. $\implies$ 2. $\implies$ 3. $\implies$ 1.

(1. $\implies$ 2.) We have $f(X) = \sum_{s\in S} f(s) 1_{X=s}$ and $g(Y) = \sum_{t\in T} g(t) 1_{Y=t}$ and therefore,

$$f(X)g(Y) = \sum_{s\in S}\sum_{t\in T} f(s)g(t)1_{X=s}\cdot 1_{Y=t} = \sum_{s\in S}\sum_{t\in T} f(s)g(t)1_{X=s,Y=t}$$

and therefore,

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \sum_{s\in S}\sum_{t\in T} f(s)g(t)P(X=s,Y=t)\\ &= \sum_{s\in S}\sum_{t\in T} f(s)P(X=s)g(t)P(Y=t) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].\end{aligned}$$

(2. $\implies$ 3.) For $A \subset S$ and $B \subset T$ take $f = 1_A$ and $g = 1_B$ in item 2. to learn

$$\begin{aligned}P(X\in A, Y\in B) &= \mathbb{E}[1_{X\in A}1_{Y\in B}] = \mathbb{E}[1_A(X)1_B(Y)]\\ &= \mathbb{E}[1_A(X)]\cdot\mathbb{E}[1_B(Y)] = P(X\in A)\cdot P(Y\in B).\end{aligned}$$

(3. $\implies$ 1.) For $s \in S$ and $t \in T$, let $A = \{s\}$ and $B = \{t\}$ in item 3. to learn

$$P(\{X=s\}\cap\{Y=t\}) = P(\{X=s\})P(\{Y=t\}),$$

i.e. $\{X = s\}$ and $\{Y = t\}$ are independent events for all $s \in S$ and $t \in T$. ■

**Exercise 5.15.** Suppose that $S$, $T$, $U$ are sets and $X : \Omega \to S$, $Y : \Omega \to T$, and $Z : \Omega \to U$ are $\mathcal{A}$ – simple functions such that $Z$ is independent of $(X,Y) : \Omega \to S\times T$. Show that $Z$ is independent of $X$. [In words, if knowing values of both $(X,Y)$ does not change the likelihood that $Z = u$ then knowing the value of just $X$ does not change the likelihood that $Z = u$ either.]

*Example 5.64.* It is **not true** in general that if $Z$ is independent of $X$ and $Z$ is independent of $Y$ then $Z$ is independent of $(X,Y)$. For example let $X, Y : \Omega \to \{\pm 1\}$ be independent random variables (i.e. $\mathcal{A}$ – simple functions) such that $P(X = \pm 1) = \frac{1}{2} = P(Y = \pm 1)$. Then take $Z = XY$. It is now fairly easy to verify that $X$ and $Z$ are independent and $Y$ and $Z$ are independent while $Z = XY$ is not independent of $(X,Y)$.

**Method 1.** Let $f, g : \{\pm 1\} \to \mathbb{R}$ be two functions, then (by doing the $x$ – sum first)

$$\mathbb{E}[f(Z)] = \frac{1}{4}\sum_{x,y\in\{\pm 1\}} f(xy) = \frac{1}{2}\sum_{y\in\{\pm 1\}}\mathbb{E}[f(X)] = \mathbb{E}f(X),$$

$$\begin{aligned}\mathbb{E}[f(Z)g(Y)] = \mathbb{E}[f(XY)g(Y)] &= \frac{1}{4}\sum_{x,y\in\{\pm 1\}} f(xy)g(y)\\ &= \frac{1}{2}\sum_{y\in\{\pm 1\}}[\mathbb{E}f(X)]g(y) = \mathbb{E}f(X)\cdot\mathbb{E}g(Y)\\ &= \mathbb{E}[f(Z)]\cdot\mathbb{E}g(Y).\end{aligned}$$

This shows $Y$ and $Z$ are independent and a similar computation shows $X$ and $Z$ are independent. On the other hand $Z = XY$ is not independent of $(X,Y)$. For example

$$0 = P(Z = 1, (X,Y) = (1,-1))$$

while

$$P(Z=1)\cdot P((X,Y) = (1,-1)) = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8} \neq 0.$$

**Method 2.** The reader should verify that for $\varepsilon, \delta \in \{\pm 1\}$ that $P(Z = \delta) = \frac{1}{2}$,

$$P(X=\varepsilon, Z=\delta) = \frac{1}{4} = P(X=\varepsilon)P(Z=\delta)$$

and same with $X$ replaced by $Y$. This shows that $X$ and $Z$ are independent and $Y$ and $Z$ are independent while $Z = XY$ is not independent of $(X,Y)$. For example

$$0 = P(Z = 1, (X,Y) = (1,-1))$$

while

$$P(Z=1)\cdot P((X,Y) = (1,-1)) = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8} \neq 0.$$

*Remark 5.65.* The above example easily generalizes as follows. Let $G$ be any finite group and suppose that $X, Y : \Omega \to G$ are independent random functions such that $P(X = g) = P(Y = g) = \frac{1}{\#(G)}$ for all $g \in G$, i.e. $X$ and $Y$ are uniformly distributed on $G$. Then $Z = XY$ will be independent of $X$ and of $Y$ separately but not independent of $(X, Y) : \Omega \to G \times G$. Example 5.64 is the special case where $G = \{\pm 1\}$. Later we may see "continuous" versions of this example as well where $G$ is replaced by any compact group.

**Definition 5.66.** *Suppose that $\{S_i\}_{i=1}^n$ are sets and $X_i : \Omega \to S_i$ is an $\mathcal{A}$ – simple function for each $1 \le i \le n$. We say $\{X_i\}_{i=1}^n$ are independent iff for each $1 \le i \le n$, $X_i$ is independent of $X^{(i)} : \Omega \to \prod_{j \ne i} S_j$ where $X^{(i)}(\omega) := \left(X_1(\omega), \dots, \hat{X}_i(\omega), \dots, X_n(\omega)\right)$ where the hat over a term means that term is to be omitted from the list.*

**Exercise 5.16.** Suppose that $X_i : \Omega \to S_i$ is an $\mathcal{A}$ – simple function for each $1 \le i \le n$. Show $\{X_i\}_{i=1}^n$ are independent iff

$$P\left(\cap_{i=1}^n \{X_i = s_i\}\right) = \prod_{i=1}^n P(X_i = s_i) \tag{5.52}$$

for all $(s_1, \dots, s_n) \in S_1 \times \dots \times S_n$.

**Exercise 5.17.** Suppose that $X_i : \Omega \to S_i$ is an $\mathcal{A}$ – simple function for each $1 \le i \le n$. Show the following are equivalent;

1. $\{X_i\}_{i=1}^n$ are independent.
2. For all choices of functions, $f_i : S_i \to \mathbb{R}$ with $1 \le i \le n$,
$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(X_i)\right].$$
3. For all choices of subsets $A_i \subset S_i$ with $1 \le i \le n$,
$$P\left(\cap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(\{X_i \in A_i\}).$$

*Remark 5.67.* If $\{X, Y, Z\}$ are as in Example 5.64, then $X$ is independent of $Y$, $X$ is independent of $Z$, and $Y$ is independent of $Z$, yet $\{X, Y, Z\}$ is not an independent collection of random variables. Thus independence of three or more random variables can not be verified by checking independence of all pairs of these random variables.

**Exercise 5.18.** Suppose now that $S$ is a finite set, $\Omega = S^n$, $\mathcal{A} = 2^\Omega$, and $X_i : \Omega \to S$is projection onto the $i^{th}$ factor of $\Omega$, i.e.

$$X_i(\omega) = \omega_i \in S \text{ for all } \omega = (\omega_1, \dots, \omega_n) \in \Omega.$$

Let $P$ be a probability measure on $(\Omega, \mathcal{A})$. Show $\{X_i\}_{i=1}^n$ are $P$ – independent iff $P$ there are functions, $q_i : S \to [0,1]$ (for $1 \le i \le n$) such that $\sum_{s \in S} q_i(s) = 1$ and

$$P(\{s\}) = \prod_{i=1}^n q_i(s_i) \text{ for all } s \in \Omega. \tag{5.53}$$

and define $q_i(s) := P(X_i = s)$ for all $s \in S$ and $1 \le i \le n$.

**Exercise 5.19 (A Weak Law of Large Numbers).** Suppose that $\Lambda \subset \mathbb{R}$ is a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, $p(\omega) = \prod_{i=1}^n q(\omega_i)$ where $q : \Lambda \to [0,1]$ such that $\sum_{\lambda \in \Lambda} q(\lambda) = 1$, and let $P : 2^\Omega \to [0,1]$ be the probability measure defined as in Eq. (5.53) with $S$ replaced by $\Lambda$. Further let $X_i(\omega) = \omega_i$ for $i = 1, 2, \dots, n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

$$S_n = \frac{1}{n}(X_1 + \dots + X_n).$$

1. Show, $\xi = \sum_{\lambda \in \Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda \in \Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda \in \Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{5.54}$$
2. Show, $\mathbb{E}S_n = \xi$.
3. Let $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \ne j$. Show
$$\mathbb{E}\left[(X_i - \xi)(X_j - \xi)\right] = \delta_{ij}\sigma^2.$$
4. Using $S_n - \xi$ may be expressed as, $\frac{1}{n}\sum_{i=1}^n (X_i - \xi)$, show
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{5.55}$$
5. Conclude using Eq. (5.55) and Remark 5.29 that
$$P(|S_n - \xi| \ge \varepsilon) \le \frac{1}{n\varepsilon^2}\sigma^2. \tag{5.56}$$

So for large $n$, $S_n$ is concentrated near $\xi = \mathbb{E}X_i$ with probability approaching 1 for $n$ large. This is a version of the weak law of large numbers.

**Definition 5.68 (Covariance).** *Let $(\Omega, \mathcal{B}, P)$ is a finitely additive probability. The* ***covariance,*** $\operatorname{Cov}(X,Y)$, *of $X, Y \in \mathbb{S}(\mathcal{B})$ is defined by*

$$\operatorname{Cov}(X,Y) = \mathbb{E}\left[(X-\xi_X)(Y-\xi_Y)\right] = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

*where $\xi_X := \mathbb{E}X$ and $\xi_Y := \mathbb{E}Y$. The variance of $X$,*

$$\operatorname{Var}(X) := \operatorname{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$$

*We say that $X$ and $Y$ are* ***uncorrelated*** *if $\operatorname{Cov}(X,Y)=0$, i.e. $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n \subset \mathbb{S}(\mathcal{B})$ are uncorrelated iff $\operatorname{Cov}(X_i, X_j) = 0$ for all $i \neq j$.*

*Remark 5.69.* 1. Observe that $X$ and $Y$ are independent iff $f(X)$ and $g(Y)$ are uncorrelated for all functions, $f$ and $g$ on the range of $X$ and $Y$ respectively. In particular if $X$ and $Y$ are independent then $\operatorname{Cov}(X,Y)=0$.

2. If you look at your proof of the weak law of large numbers in Exercise 5.19 you will see that it suffices to assume that $\{X_i\}_{i=1}^n$ are uncorrelated rather than the stronger condition of being independent.

**Exercise 5.20 (Bernoulli Random Variables).** Let $\Lambda = \{0,1\}$, $X : \Lambda \to \mathbb{R}$ be defined by $X(0) = 0$ and $X(1) = 1$, $x \in [0,1]$, and define $Q = x\delta_1 + (1-x)\delta_0$, i.e. $Q(\{0\}) = 1-x$ and $Q(\{1\}) = x$. Verify,

$$\xi(x) := \mathbb{E}_Q X = x \text{ and}$$
$$\sigma^2(x) := \mathbb{E}_Q (X-x)^2 = (1-x)x \le 1/4.$$

**Theorem 5.70 (Weierstrass Approximation Theorem via Bernstein's Polynomials.).** *Suppose that $f \in C([0,1], \mathbb{C})$ and*

$$p_n(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

*Then*

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - p_n(x)| = 0.$$

**Proof.** Let $x \in [0,1]$, $\Lambda = \{0,1\}$, $q(0) = 1-x$, $q(1) = x$, $\Omega = \Lambda^n$, and

$$P_x(\{\omega\}) = q(\omega_1)\dots q(\omega_n) = x^{\sum_{i=1}^n \omega_i} \cdot (1-x)^{1-\sum_{i=1}^n \omega_i}.$$

As above, let $S_n = \frac{1}{n}(X_1 + \dots + X_n)$, where $X_i(\omega) = \omega_i$ and observe that

$$P_x\left(S_n = \frac{k}{n}\right) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Therefore, writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$, we have

$$\mathbb{E}_x[f(S_n)] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} = p_n(x).$$

Hence we find

$$\begin{aligned}
|p_n(x) - f(x)| &= |\mathbb{E}_x f(S_n) - f(x)| = |\mathbb{E}_x[f(S_n) - f(x)]| \\
&\le \mathbb{E}_x |f(S_n) - f(x)| \\
&= \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| \ge \varepsilon] \\
&\quad + \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| < \varepsilon] \\
&\le 2M \cdot P_x(|S_n - x| \ge \varepsilon) + \delta(\varepsilon)
\end{aligned}$$

where

$$M := \max_{y\in[0,1]} |f(y)| \text{ and}$$
$$\delta(\varepsilon) := \sup\{|f(y) - f(x)| : x, y \in [0,1] \text{ and } |y-x| \le \varepsilon\}$$

is the modulus of continuity of $f$. Now by the above exercises,

$$P_x(|S_n - x| \ge \varepsilon) \le \frac{1}{4n\varepsilon^2} \quad \text{(see Figure 5.2)} \tag{5.57}$$

and hence we may conclude that

$$\max_{x\in[0,1]} |p_n(x) - f(x)| \le \frac{M}{2n\varepsilon^2} + \delta(\varepsilon)$$

and therefore, that

$$\limsup_{n\to\infty} \max_{x\in[0,1]} |p_n(x) - f(x)| \le \delta(\varepsilon).$$

This completes the proof, since by uniform continuity of $f$, $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. ■

### 5.9.1 Complex Weierstrass Approximation Theorem

The main goal of this subsection is to prove Theorem 5.58 which states that any continuous $2\pi$ – periodic function on $\mathbb{R}$ may be well approximated by trigonometric polynomials. The main ingredient is the following two dimensional generalization of Theorem 5.70. All of the results in this section have natural generalization to higher dimensions as well , see Theorem 5.79.

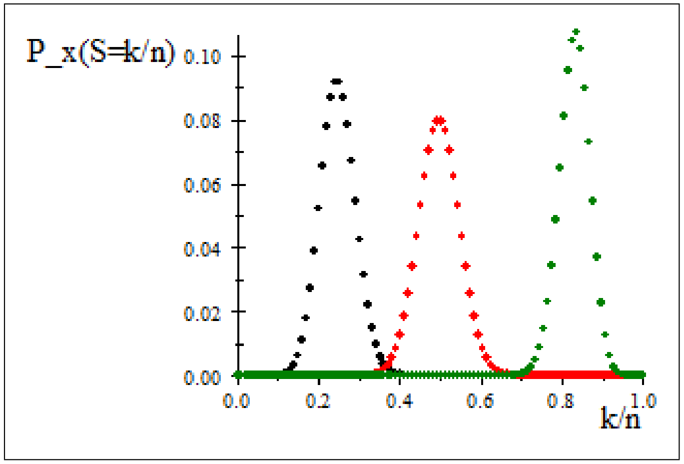



**Fig. 5.2.** Plots of $P_x(S_n = k/n)$ versus $k/n$ for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

**Theorem 5.71 (Weierstrass Approximation Theorem).** *Suppose that* $K = [0,1]^2$, $f \in C(K, \mathbb{C})$, *and*

$$p_n(x,y) := \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) \binom{n}{k} \binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}. \tag{5.58}$$

*Then* $p_n \to f$ *uniformly on* $K$.

**Proof.** We are going to follow the argument given in the proof of Theorem 5.70. By considering the real and imaginary parts of $f$ separately, it suffices to assume $f \in C([0,1]^2, \mathbb{R})$. For $(x,y) \in K$ and $n \in \mathbb{N}$ we may choose a collection of independent Bernoulli simple random variables $\{X_i, Y_i\}_{i=1}^n$ such that $P(X_i = 1) = x$ and $P(Y_i = 1) = y$ for all $1 \le i \le n$. Then letting $S_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, we have

$$\mathbb{E}[f(S_n, T_n)] = \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) P(n \cdot S_n = k, n \cdot T_n = l) = p_n(x,y)$$

where $p_n(x,y)$ is the polynomial given in Eq. (5.58) wherein the assumed independence is needed to show,

$$P(n \cdot S_n = k, n \cdot T_n = l) = \binom{n}{k} \binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}.$$

Thus if $M = \sup\{|f(x,y)| : (x,y) \in K\}$, $\varepsilon > 0$,

$$\delta_\varepsilon = \sup\{|f(x',y') - f(x,y)| : (x,y), (x',y') \in K \text{ and } \|(x',y') - (x,y)\| \le \varepsilon\},$$

and

$$A := \{\|(S_n, T_n) - (x,y)\| > \varepsilon\},$$

we have,

$$\begin{aligned} |f(x,y) - p_n(x,y)| &= |\mathbb{E}(f(x,y) - f((S_n,T_n)))| \\ &\le \mathbb{E}|f(x,y) - f((S_n,T_n))| \\ &= \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A] \\ &\quad + \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A^c] \\ &\le 2M \cdot P(A) + \delta_\varepsilon \cdot P(A^c) \\ &\le 2M \cdot P(A) + \delta_\varepsilon. \end{aligned} \tag{5.59}$$

To estimate $P(A)$, observe that if

$$\|(S_n,T_n) - (x,y)\|^2 = (S_n - x)^2 + (T_n - y)^2 > \varepsilon^2,$$

then either,

$$(S_n - x)^2 > \varepsilon^2/2 \text{ or } (T_n - y)^2 > \varepsilon^2/2$$

and therefore by sub-additivity and Eq. (5.57) we know

$$\begin{aligned} P(A) &\le P\left(|S_n - x| > \varepsilon/\sqrt{2}\right) + P\left(|T_n - y| > \varepsilon/\sqrt{2}\right) \\ &\le \frac{1}{2n\varepsilon^2} + \frac{1}{2n\varepsilon^2} = \frac{1}{n\varepsilon^2}. \end{aligned} \tag{5.60}$$

Using this estimate in Eq. (5.59) gives,

$$|f(x,y) - p_n(x,y)| \le 2M \cdot \frac{1}{n\varepsilon^2} + \delta_\varepsilon$$

and as right is independent of $(x,y) \in K$ we may conclude,

$$\limsup_{n\to\infty} \sup_{(x,y)\in K} |f(x,y) - p_n(x,y)| \le \delta_\varepsilon$$

which completes the proof since $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$ because $f$ is uniformly continuous on $K$. ■

*Remark 5.72.* We can easily improve our estimate on $P(A)$ in Eq. (5.60) by a factor of two as follows. As in the proof of Theorem 5.70,

$$\begin{aligned} \mathbb{E}\left[\|(S_n,T_n) - (x,y)\|^2\right] &= \mathbb{E}\left[(S_n - x)^2 + (T_n - y)^2\right] \\ &= \mathrm{Var}(S_n) + \mathrm{Var}(T_n) \\ &= \frac{1}{n}x(1-x) + y(1-y) \le \frac{1}{2n}. \end{aligned}$$

Therefore by Chebyshev's inequality,

$$P(A) = P(\|(S_n, T_n) - (x,y)\| > \varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|(S_n,T_n) - (x,y)\|^2 \le \frac{1}{2n\varepsilon^2}.$$

**Corollary 5.73.** *Suppose that $K = [a,b]\times[c,d]$ is any compact rectangle in $\mathbb{R}^2$. Then every function, $f \in C(K,\mathbb{C})$, may be uniformly approximated by polynomial functions in $(x,y) \in \mathbb{R}^2$.*

**Proof.** Let $F(x,y) := f(a + x(b-a), c + y(d-c))$ – a continuous function of $(x,y) \in [0,1]^2$. Given $\varepsilon > 0$, we may use Theorem Theorem 5.71 to find a polynomial, $p(x,y)$, such that $\sup_{(x,y)\in[0,1]^2} |F(x,y) - p(x,y)| \le \varepsilon$. Letting $\xi = a + x(b-a)$ and $\eta := c + y(d-c)$, it now follows that

$$\sup_{(\xi.\eta)\in K}\left| f(\xi,\eta) - p\left(\frac{\xi - a}{b-a}, \frac{\eta - c}{d-c}\right)\right| \le \varepsilon$$

which completes the proof since $p\left(\frac{\xi-a}{b-a}, \frac{\eta-c}{d-c}\right)$ is a polynomial in $(\xi,\eta)$. ■

## 5.10 Simple Conditional Expectation

In this section, $\mathcal{B}$ is a sub-algebra of $2^{\Omega}$, $P : \mathcal{B} \to [0,1]$ is a finitely additive probability measure, $\mathcal{A} \subset \mathcal{B}$ is a finite sub-algebra, and $X : \Omega \to \mathbb{R}$ is a $\mathcal{B}$ – simple function. As in Example 4.19, for each $\omega \in \Omega$, let $A_\omega := \cap\{A \in \mathcal{A} : \omega \in A\}$ and recall that either $A_\omega = A_{\omega'}$ or $A_\omega \cap A_{\omega'} = \emptyset$ for all $\omega, \omega' \in \Omega$. Let $\{B_1, \dots, B_m\} \subset \mathcal{A}$ be an enumeration of $\{A_\omega : \omega \in \Omega\}$ in which case $\{B_1, \dots, B_m\}$ is a partition of $\Omega$ and $\mathcal{A} = \mathcal{A}(\{B_1, \dots, B_m\})$. As usual we heuristically think that for $B \in \mathcal{B}$,

$$P(B) := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_B(\omega_n)$$

where $\{\omega_n\}_{n=1}^{\infty} \subset \Omega$ are the outcomes of a sequence of identical and independent experiments.

We now imagine that our measuring devices can only tell us which $B_i$ that the $\omega_n$ is in and hence we get a report from the experiment consisting of $\{(i(\omega_n), X(\omega_n))\}_{n=1}^{\infty} \subset \{1, \dots, m\} \times \mathbb{R}$ where $i(\omega) = j$ iff $\omega \in B_j$, i.e. $i := \sum_{j=1}^{m} j 1_{B_j}$. Our manager now asks us to make sense out this data, i.e. she wants to know what $X(\omega)$ will be if we know $\omega \in B_j$. However the data we are given is inconsistent with giving such an answer so we do the best we can and give her the average of the $X(\omega_n)$ for which $\omega_n \in B_j$. Thus we "define" for $\omega \in B_j$,

$$\begin{aligned}\bar{X}(\omega) &= \lim_{N\to\infty} \frac{1}{\sum_{n=1}^{N} 1_{B_j}(\omega_n)} \sum_{n=1}^{N} X(\omega_n) 1_{B_j}(\omega_n) \\ &= \lim_{N\to\infty} \frac{N}{\sum_{n=1}^{N} 1_{B_j}(\omega_n)} \frac{1}{N}\sum_{n=1}^{N} X(\omega_n) 1_{B_j}(\omega_n) \\ &= \frac{1}{P(B_j)}\mathbb{E}\left[X \cdot 1_{B_j}\right] = \mathbb{E}_{P(\cdot|B_j)}[X].\end{aligned}$$

In summary, if $\omega \in B_j$, then $\bar{X}(\omega)$ is the limiting average of $X(\omega_n)$ over those experiments $\omega_n$ which happen to lie in $B_j$. This represents our "best guess" what $X(\omega)$ will be given $\omega \in B_j$. We now remove the heuristics and formalize this notion in the following definition.

**Definition 5.74 (Conditional expectation).** *Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{B}$ – simple random variable, i.e. $X \in \mathbb{S}(\mathcal{B})$, and*

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)}\mathbb{E}[1_{A_\omega} X] \text{ for all } \omega \in \Omega, \tag{5.61}$$

*where by convention let us set $\bar{X}(\omega) = 0$ if $P(A_\omega) = 0$. We will denote $\bar{X}$ by $\mathbb{E}[X|\mathcal{A}]$ or $\mathbb{E}_{\mathcal{A}}X$ and call it the* ***conditional expectation*** *of $X$ given $\mathcal{A}$. Alternatively we may write $\bar{X}$ as*

$$\bar{X} = \sum_{j=1}^{m} \frac{\mathbb{E}\left[1_{B_j} X\right]}{P(B_j)} 1_{B_j} = \sum_{j=1}^{m} \mathbb{E}_{P(\cdot|B_j)}[X] \cdot 1_{B_j}, \tag{5.62}$$

*again with the convention that $\mathbb{E}\left[1_{B_j}X\right]/P(B_j) = 0$ if $P(B_j) = 0$. [It should be noted, from Exercise 5.4, that $\bar{X} = \mathbb{E}_{\mathcal{A}}X \in \mathbb{S}(\mathcal{A})$.]*

**Exercise 5.21 (Simple conditional expectation).** Let $X \in \mathbb{S}(\mathcal{B})$ and, for simplicity, assume all functions are real valued. Prove the following assertions;

1. **(Orthogonal Projection Property 1.)** If $Z \in \mathbb{S}(\mathcal{A})$, then

$$\mathbb{E}[XZ] = \mathbb{E}\left[\bar{X}Z\right] = \mathbb{E}[\mathbb{E}_{\mathcal{A}}X \cdot Z] \tag{5.63}$$

and

$$(\mathbb{E}_{\mathcal{A}}Z)(\omega) = \begin{cases} Z(\omega) & \text{if } P(A_\omega) > 0 \\ 0 & \text{if } P(A_\omega) = 0 \end{cases}. \tag{5.64}$$

In particular, $\mathbb{E}_{\mathcal{A}}[\mathbb{E}_{\mathcal{A}}Z] = \mathbb{E}_{\mathcal{A}}Z$.
This basically says that $\mathbb{E}_{\mathcal{A}}$ is orthogonal projection from $\mathbb{S}(\mathcal{B})$ onto $\mathbb{S}(\mathcal{A})$ relative to the inner product

$$(f,g) = \mathbb{E}[fg] \text{ for all } f,g \in \mathbb{S}(\mathcal{B}).$$

2. **(Orthogonal Projection Property 2.)** If $Y \in \mathbb{S}(\mathcal{A})$ satisfies, $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all $Z \in \mathbb{S}(\mathcal{A})$, then $Y(\omega) = \bar{X}(\omega)$ whenever $P(A_\omega) > 0$. In particular, $P\left(Y \neq \bar{X}\right) = 0$. **Hint:** use item 1. to compute $\mathbb{E}\left[\left(\bar{X} - Y\right)^2\right]$.
3. **(Best Approximation Property.)** For any $Y \in \mathbb{S}(\mathcal{A})$,
$$\mathbb{E}\left[\left(X - \bar{X}\right)^2\right] \leq \mathbb{E}\left[(X - Y)^2\right] \tag{5.65}$$
with equality iff $\bar{X} = Y$ almost surely (a.s. for short), where $\bar{X} = Y$ a.s. iff $P\left(\bar{X} \neq Y\right) = 0$. In words, $\bar{X} = \mathbb{E}_\mathcal{A} X$ is the best ("$L^2$") approximation to $X$ by an $\mathcal{A}$ – measurable random variable.
4. **(Contraction Property.)** $\mathbb{E}\left|\bar{X}\right| \leq \mathbb{E}|X|$. (It is typically **not** true that $\left|\bar{X}(\omega)\right| \leq |X(\omega)|$ for all $\omega$.)
5. (**Pull Out Property.**) If $Z \in \mathbb{S}(\mathcal{A})$, then
$$\mathbb{E}_\mathcal{A}[ZX] = Z\mathbb{E}_\mathcal{A} X.$$

*Remark 5.75.* The cleanest way to see that $\mathbb{E}_\mathcal{A}$ in Eq. (5.62) is an orthogonal projection is to let

$$(X, Y) := \mathbb{E}[XY] \text{ for all } X, Y \in \mathbb{S}(\mathcal{B})$$

and observe that $(\cdot, \cdot)$ satisfies the axioms of inner product except for possibly the axiom that $(X, X) = 0$ implies $X = 0$. What is true is that if $(X, X) = 0$, then $X = 0$ a.s., i.e. $P(X \neq 0) = 0$. To avoid technicalities associate with these "null" sets, let us suppose that $P(B_i) > 0$ for each $i$. In this case $\left\{\frac{1_{B_i}}{\sqrt{P(B_i)}}\right\}_i$ is an orthonormal basis for the subspace $\mathbb{S}(\mathcal{A}) \subset \mathbb{S}(\mathcal{B})$. Therefore orthogonal projection from $\mathbb{S}(\mathcal{B})$ onto $\mathbb{S}(\mathcal{A})$ is given by

$$X \to \sum_i \left(X, \frac{1_{B_i}}{\sqrt{P(B_i)}}\right) \frac{1_{B_i}}{\sqrt{P(B_i)}} = \sum_i \frac{\mathbb{E}[X 1_{B_i}]}{P(B_i)} 1_{B_i}$$

which is precisely the formula for $\mathbb{E}_\mathcal{A} X$.

*Example 5.76 (Heuristics of independence and conditional expectations).* Let us suppose that we have an experiment consisting of spinning a spinner with values in $\Lambda_1 = \{1, 2, \dots, 10\}$ and rolling a die with values in $\Lambda_2 = \{1, 2, 3, 4, 5, 6\}$. So the outcome of an experiment is represented by a point, $\omega = (x, y) \in \Omega = \Lambda_1 \times \Lambda_2$. Let $X(x, y) = x$, $Y(x, y) = y$, $\mathcal{B} = 2^\Omega$, and

$$\mathcal{A} = \mathcal{A}(X) = X^{-1}\left(2^{\Lambda_1}\right) = \left\{X^{-1}(A) : A \subset \Lambda_1\right\} \subset \mathcal{B},$$

so that $\mathcal{A}$ is the smallest algebra of subsets of $\Omega$ such that $\{X = x\} \in \mathcal{A}$ for all $x \in \Lambda_1$. Notice that the partition associated to $\mathcal{A}$ is precisely

$$\{\{X = 1\}, \{X = 2\}, \dots, \{X = 10\}\}.$$

Let us now suppose that the spins of the spinner are "empirically independent" of the throws of the dice. As usual let us run the experiment repeatedly to produce a sequence of results, $\omega_n = (x_n, y_n)$ for all $n \in \mathbb{N}$. If $g : \Lambda_2 \to \mathbb{R}$ is a function, we have (heuristically) that

$$\begin{aligned}\mathbb{E}_\mathcal{A}[g(Y)](x, y) &= \lim_{N\to\infty} \frac{\sum_{n=1}^N g(Y(\omega(n))) 1_{X(\omega(n))=x}}{\sum_{n=1}^N 1_{X(\omega(n))=x}} \\ &= \lim_{N\to\infty} \frac{\sum_{n=1}^N g(y_n) 1_{x_n=x}}{\sum_{n=1}^N 1_{x_n=x}}.\end{aligned}$$

As the $\{y_n\}$ sequence of results are independent of the $\{x_n\}$ sequence, we should expect by the usual mantra[3] that

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N g(y_n) 1_{x_n=x}}{\sum_{n=1}^N 1_{x_n=x}} = \lim_{N\to\infty} \frac{1}{M(N)} \sum_{n=1}^{M(N)} g(\bar{y}_n) = \mathbb{E}[g(Y)],$$

where $M(N) = \sum_{n=1}^N 1_{x_n=x}$ and $(\bar{y}_1, \bar{y}_2, \dots) = \{y_l : 1_{x_l=x}\}$. (We are also assuming here that $P(X = x) > 0$ so that we expect, $M(N) \sim P(X = x) N$ for $N$ large, in particular $M(N) \to \infty$.) Thus under the assumption that $X$ and $Y$ are describing "independent" experiments we have heuristically deduced that $\mathbb{E}_\mathcal{A}[g(Y)] : \Omega \to \mathbb{R}$ is the constant function;

$$\mathbb{E}_\mathcal{A}[g(Y)](x, y) = \mathbb{E}[g(Y)] \text{ for all } (x, y) \in \Omega. \tag{5.66}$$

Let us further observe that if $f : \Lambda_1 \to \mathbb{R}$ is any other function, then $f(X)$ is an $\mathcal{A}$ – simple function and therefore by Eq. (5.66) and Exercise 5.21

$$\mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)] = \mathbb{E}[f(X) \cdot \mathbb{E}[g(Y)]] = \mathbb{E}[f(X) \cdot \mathbb{E}_\mathcal{A}[g(Y)]] = \mathbb{E}[f(X) \cdot g(Y)].$$

**Lemma 5.77 (Conditional Expectation and Independence).** *Let $\Omega = \Lambda_1 \times \Lambda_2$, $X$, $Y$, $\mathcal{B} = 2^\Omega$, and $\mathcal{A} = X^{-1}\left(2^{\Lambda_1}\right)$, be as in Example 5.76 above. Assume that $P : \mathcal{B} \to [0, 1]$ is a probability measure. If $X$ and $Y$ are $P$ – independent, then Eq. (5.66) holds.*

**Proof.** From the definitions of conditional expectation and of independence we have,

$$\mathbb{E}_\mathcal{A}[g(Y)](x, y) = \frac{\mathbb{E}[1_{X=x} \cdot g(Y)]}{P(X = x)} = \frac{\mathbb{E}[1_{X=x}] \cdot \mathbb{E}[g(Y)]}{P(X = x)} = \mathbb{E}[g(Y)].$$

[3] That is it should not matter which sequence of independent experiments are used to compute the time averages.

■

The following theorem summarizes much of what we (i.e. you) have shown regarding the underlying notion of independence of a pair of simple functions.

**Theorem 5.78 (Independence result summary).** *Let $(\Omega, \mathcal{B}, P)$ be a finitely additive probability space, $\Lambda$ be a finite set, and $X, Y : \Omega \to \Lambda$ be two $\mathcal{B}$ – measurable simple functions, i.e. $\{X = x\} \in \mathcal{B}$ and $\{Y = y\} \in \mathcal{B}$ for all $x, y \in \Lambda$. Further let $\mathcal{A} = \mathcal{A}(X) := \mathcal{A}(\{X = x\} : x \in \Lambda)$. Then the following are equivalent;*

1. $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ *for all* $x \in \Lambda$ *and* $y \in \Lambda$,
2. $\mathbb{E}[f(X)\, g(Y)] = \mathbb{E}[f(X)]\, \mathbb{E}[g(Y)]$ *for all functions,* $f : \Lambda \to \mathbb{R}$ *and* $g : \Lambda \to \mathbb{R}$,
3. $\mathbb{E}_{\mathcal{A}(X)}[g(Y)] = \mathbb{E}[g(Y)]$ *for all* $g : \Lambda \to \mathbb{R}$, *and*
4. $\mathbb{E}_{\mathcal{A}(Y)}[f(X)] = \mathbb{E}[f(X)]$ *for all* $f : \Lambda \to \mathbb{R}$.

*We say that $X$ and $Y$ are $P$ – independent if any one (and hence all) of the above conditions holds.*

## 5.11 Appendix: A Multi-dimensional Weirstrass Approximation Theorem

The following theorem is the multi-dimensional generalization of Theorem 5.70.

**Theorem 5.79 (Weierstrass Approximation Theorem).** *Suppose that $K = [a_1, b_1] \times \ldots [a_d, b_d]$ with $-\infty < a_i < b_i < \infty$ is a compact rectangle in $\mathbb{R}^d$. Then for every $f \in C(K, \mathbb{C})$, there exists polynomials $p_n$ on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on $K$.*

**Proof.** By a simple scaling and translation of the arguments of $f$ we may assume without loss of generality that $K = [0,1]^d$. By considering the real and imaginary parts of $f$ separately, it suffices to assume $f \in C([\mathbf{0}, \mathbf{1}], \mathbb{R})$.

Given $x \in K$, let $\left\{X_n = \left(X_n^1, \ldots, X_n^d\right)\right\}_{n=1}^\infty$ be i.i.d. random vectors with values in $\mathbb{R}^d$ such that

$$P(X_n = \eta) = \prod_{i=1}^d (1 - x_i)^{1-\eta_i} x_i^{\eta_i}$$

for all $\eta = (\eta_1, \ldots, \eta_d) \in \{0,1\}^d$. Since each $X_n^j$ is a Bernoulli random variable with $P\left(X_n^j = 1\right) = x_j$, we know that

$$\mathbb{E}X_n = x \text{ and } \operatorname{Var}\left(X_n^j\right) = x_j - x_j^2 = x_j(1 - x_j).$$

As usual let $S_n := X_1 + \cdots + X_n \in \mathbb{R}^d$, then

$$\mathbb{E}\left[\frac{S_n}{n}\right] = x \text{ and}$$

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{S_n}{n} - x\right\|^2\right] &= \sum_{j=1}^d \mathbb{E}\left(\frac{S_n^j}{n} - x_j\right)^2 = \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n} - x_j\right) \\ &= \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n}\right) = \frac{1}{n^2} \cdot \sum_{j=1}^d \sum_{k=1}^n \operatorname{Var}\left(X_k^j\right) \\ &= \frac{1}{n} \sum_{j=1}^d x_j(1 - x_j) \le \frac{d}{4n}.\end{aligned}$$

This shows $S_n/n \to x$ in $L^2(P)$ and hence by Chebyshev's inequality, $S_n/n \xrightarrow{P} x$ and in fact for any $\varepsilon > 0$,

$$P\left(\left\|\frac{S_n}{n} - x\right\| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2} \frac{d}{4n}.$$

Observe that

$$\begin{aligned}p_n(x) &:= \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \\ &= \sum_{\eta:\{1,2,\ldots,n\} \to \{0,1\}^d} f\left(\frac{\eta(1) + \cdots + \eta(n)}{n}\right) P(X_1 = \eta(1), \ldots, X_n = \eta(n)) \\ &= \sum_{\eta:\{1,2,\ldots,n\} \to \{0,1\}^d} f\left(\frac{\eta(1) + \cdots + \eta(n)}{n}\right) \prod_{k=1}^n \prod_{i=1}^d (1 - x_i)^{1-\eta_i(k)} x_i^{\eta_i(k)}\end{aligned} \tag{5.67}$$

is a polynomial in $x$ of degree $nd$ at most. If we further let $M = \sup\{|f(x)| : x \in K\}$, and

$$\delta_\varepsilon = \sup\{|f(y) - f(x)| : x, y \in K \text{ and } \|y - x\| \le \varepsilon\},$$

then

$$\begin{aligned}|f(x) - p_n(x)| &= \left|\mathbb{E}\left(f(x) - f\left(\frac{S_n}{n}\right)\right)\right| \le \mathbb{E}\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \\ &\le \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right| : \|S_n - x\| > \varepsilon\right] \\ &\quad + \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right| : \|S_n - x\| \le \varepsilon\right] \\ &\le 2MP(\|S_n - x\| > \varepsilon) + \delta_\varepsilon \le \frac{2dM}{n\varepsilon^2} + \delta_\varepsilon.\end{aligned} \tag{5.68}$$

Therefore,

$$\limsup_{n\to\infty} \max_{x\in K} |f(x) - p_n(x)| \le \delta_\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0$$

wherein we have used the uniform continuity of $f$ to guarantee $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$.
■

*Remark 5.80.* We can write out the expression in Eq. (5.67) more explicitly by observing that $\left\{S_n^1, \dots, S_n^d\right\}$ are independent random variables with

$$\mathbb{P}_x\left(S_n^i = k\right) = \binom{n}{k} x_i^k \left(1 - x_i\right)^{n-k} \text{ for } 0 \le k \le n$$

and therefore

$$p_n(x) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \mathbb{E}\left[f\left(\frac{\left(S_n^1, \dots, S_n^d\right)}{n}\right)\right]$$
$$= \sum_{\mathbf{k}\in\{0,\dots,n\}^d} f\left(\frac{\mathbf{k}}{n}\right) \prod_{i=1}^d \binom{n}{k_i} x_i^{k_i} \left(1 - x_i\right)^{n-k_i}.$$

Here is a version of the complex Weierstrass approximation theorem.

# Countably Additive Measures

Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra and $\mu : \mathcal{A} \to [0, \infty]$ be a finitely additive measure. Recall that $\mu$ is a **premeasure** on $\mathcal{A}$ if $\mu$ is $\sigma$ – additive on $\mathcal{A}$.

**Definition 6.1.** *A* ***premeasure space*** *is a triple, $(\Omega, \mathcal{A}, \mu)$, where $\Omega$ is a non-empty set, $\mathcal{A}$ is a sub-algebra of $2^{\Omega}$, and $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure on $\mathcal{A}$. If in addition $\mathcal{A}$ is a $\sigma$ – algebra (Definition 4.12), we say $(\Omega, \mathcal{A})$ is a* ***measurable space*** *and $(\Omega, \mathcal{A}, \mu)$ is a* ***measure space.***

**Definition 6.2.** *Let $(\Omega, \mathcal{B})$ be a measurable space. We say that $P : \mathcal{B} \to [0, 1]$ is a* ***probability measure on*** *$(\Omega, \mathcal{B})$ if $P$ is a measure on $\mathcal{B}$ such that $P(\Omega) = 1$ and in this case we say $(\Omega, \mathcal{B}, P)$ is a* ***probability space.***

## 6.1 Overview

The goal of this chapter is develop methods for proving the existence of probability measures with desirable properties. The main results of this chapter may are summarized in the following theorem. Throughout this chapter $\mathcal{A}$ will be a sub-algebra of $2^{\Omega}$ and $\mu$ will be a finitely additive measure on $\mathcal{A}$ with $\mu(\Omega) < \infty$.

**Theorem 6.3.** *A finitely additive finite measure $\mu$ on an algebra, $\mathcal{A} \subset 2^{\Omega}$, extends to $\sigma$ – additive measure on $\sigma(\mathcal{A})$ iff $\mu$ is a premeasure on $\mathcal{A}$. If the extension exists it is unique.*

**Proof.** The uniqueness assertion is proved Proposition 6.18 below. The existence assertion of the theorem is contained inTheorem 6.44 (also restated in Theorem 6.13 below). ∎

In order to use this theorem it is necessary to determine when a finitely additive probability measure in is in fact a premeasure. The following proposition (which may be omitted until needed) is sometimes useful in this regard.

**Proposition 6.4 (Equivalent premeasure conditions).** *Suppose that $\mu$ is a finitely additive probability measure on an algebra, $\mathcal{A} \subset 2^{\Omega}$. Then the following are equivalent:*

1. *$\mu$ is subadditive on $\mathcal{A}$.*
2. *$\mu$ is a premeasure on $\mathcal{A}$, i.e. $\mu$ is $\sigma$ – additive on $\mathcal{A}$.*
3. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, $\mu(A_n) \uparrow \mu(A)$.*
4. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, $\mu(A_n) \downarrow \mu(A)$.*
5. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow \Omega$, $\mu(A_n) \uparrow \mu(\Omega)$.*
6. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow \emptyset$, $\mu(A_n) \downarrow 0$.*

**Proof.** The equivalence of 1 and 2 has already been shown in Proposition 5.2. We will next show $2 \iff 3 \iff 4$.

$2. \implies 3.$ Suppose $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Let $A_n' := A_n \setminus A_{n-1}$ with $A_0 := \emptyset$. Then $\{A_n'\}_{n=1}^{\infty}$ are disjoint, $A_n = \cup_{k=1}^{n} A_k'$ and $A = \cup_{k=1}^{\infty} A_k'$. Therefore,

$$\mu(A) = \sum_{k=1}^{\infty} \mu(A_k') = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k') = \lim_{n\to\infty} \mu\left(\cup_{k=1}^{n} A_k'\right) = \lim_{n\to\infty} \mu(A_n).$$

$3. \implies 2.$ If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ are disjoint and $A := \cup_{n=1}^{\infty} A_n \in \mathcal{A}$, then $\cup_{n=1}^{N} A_n \uparrow A$. Therefore,

$$\mu(A) = \lim_{N\to\infty} \mu\left(\cup_{n=1}^{N} A_n\right) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

$3. \implies 4.$ If $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, then $A_n^c \uparrow A^c$ and therefore,

$$\lim_{n\to\infty} (\mu(\Omega) - \mu(A_n)) = \lim_{n\to\infty} \mu(A_n^c) = \mu(A^c) = \mu(\Omega) - \mu(A).$$

$4. \implies 3.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A_n^c \downarrow A^c$ and therefore we again have,

$$\lim_{n\to\infty} (\mu(\Omega) - \mu(A_n)) = \lim_{n\to\infty} \mu(A_n^c) = \mu(A^c) = \mu(\Omega) - \mu(A).$$

The same proof used for $3. \iff 4.$ shows $5. \iff 6.$ and it is clear that $4. \implies 6.$ To finish the proof we will show $6. \implies 3.$

$6. \implies 3.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A \setminus A_n \downarrow \emptyset$ and therefore

$$\lim_{n\to\infty} [\mu(A) - \mu(A_n)] = \lim_{n\to\infty} \mu(A \setminus A_n) = 0.$$

∎

*Remark 6.5.* Observe that the equivalence of items 1., 2., and 3. in Proposition 6.4 hold even if $\mu(\Omega) = \infty$.

### 6.1.1 Examples of measures on $\mathbb{R}$

Let us now specialize to the case where $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{A}\left(\{(a,b]\cap\mathbb{R} : -\infty \le a \le b \le \infty\}\right)$. In this case we will describe probability measures, $P$, on $\mathcal{B}_\mathbb{R}$ by their "cumulative distribution functions."

**Definition 6.6.** *Given a probability measure, $P$ on $\mathcal{B}_\mathbb{R}$, the* ***cumulative distribution function*** *(CDF) of $P$ is defined as the function, $F = F_P : \mathbb{R} \to [0,1]$ given as*

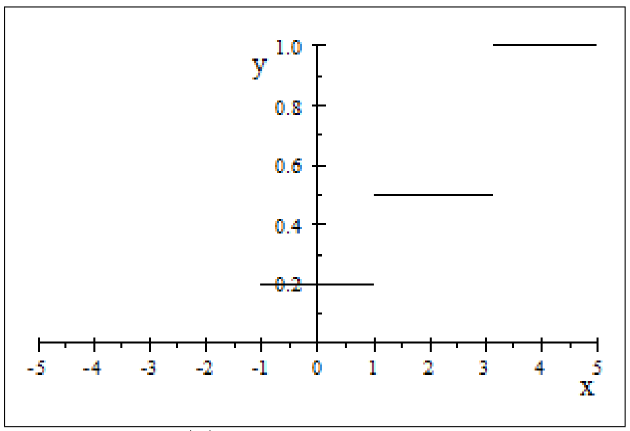
$$F(x) := P((-\infty, x]). \tag{6.1}$$

*Example 6.7.* Suppose that

$$P = p\delta_{-1} + q\delta_1 + r\delta_\pi$$

with $p,q,r > 0$ and $p+q+r = 1$. In this case,

$$F(x) = \begin{cases} 0 & \text{for } \quad x < -1 \\ p & \text{for } -1 \le x < 1 \\ p+q & \text{for } \ 1 \le x < \pi \\ 1 & \text{for } \pi \le x < \infty \end{cases}.$$




A plot of $F(x)$ with $p = .2$, $q = .3$, and $r = .5$.

**Lemma 6.8.** *If $F = F_P : \mathbb{R} \to [0,1]$ is a distribution function for a probability measure, $P$, on $\mathcal{B}_\mathbb{R}$, then:*

1. *$F$ is non-decreasing,*
2. *$F$ is right continuous,*
3. *$F(-\infty) := \lim_{x\to-\infty} F(x) = 0,$ and $F(\infty) := \lim_{x\to\infty} F(x) = 1.$*

**Proof.** The monotonicity of $P$ shows that $F(x)$ in Eq. (6.1) is non-decreasing. For $b \in \mathbb{R}$ let $A_n = (-\infty, b_n]$ with $b_n \downarrow b$ as $n \to \infty$. The continuity of $P$ implies

$$F(b_n) = P((-\infty, b_n]) \downarrow \mu((-\infty, b]) = F(b).$$

Since $\{b_n\}_{n=1}^\infty$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $F(b+) := \lim_{y\downarrow b} F(y) = F(b)$. This show that $F$ is right continuous. Similar arguments show that $F(\infty) = 1$ and $F(-\infty) = 0$. ■

It turns out that Lemma 6.8 has the following important converse.

**Theorem 6.9.** *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 3.. in Lemma 6.8, there exists a unique probability measure, $P_F$, on $\mathcal{B}_\mathbb{R}$ such that*

$$P_F((a,b]) = F(b) - F(a) \ \text{ for all } -\infty < a \le b < \infty.$$

**Proof.** The uniqueness assertion is proved in Corollary 6.20 below or see Exercises 6.2 and 6.3 below. The existence portion of the theorem is a special case of Theorem 6.58 below. ■

*Example 6.10 (Uniform Distribution).* The function,

$$F(x) := \begin{cases} 0 \text{ for } \quad x \le 0 \\ x \text{ for } \ 0 \le x < 1 \\ 1 \text{ for } 1 \le x < \infty \end{cases},$$

is the distribution function for a measure, $m$ on $\mathcal{B}_\mathbb{R}$ which is concentrated on $(0,1]$. The measure, $m$ is called the **uniform distribution** or **Lebesgue measure** on $(0,1]$.

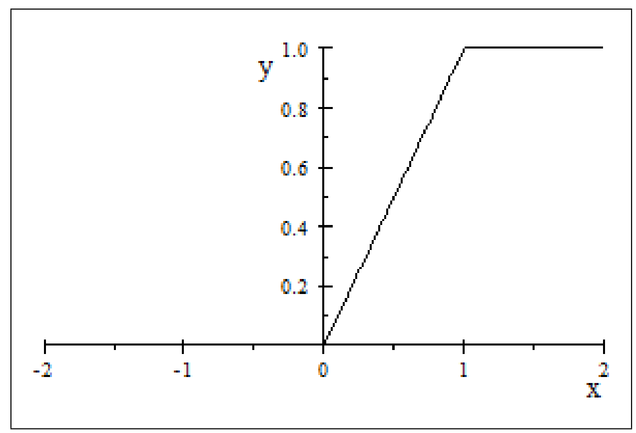



**Fig. 6.1.** The cumulative distribution function for the uniform distribution.

### 6.1.2 An Extension of Measure Strategy

Let us end this overview by briefly explaining the strategy we will use below for extending measures.

*Example 6.11.* It is easy to verify that to every finitely additive measure, $\mu : \mathcal{A} \to [0, \infty)$, the function $d_\mu : \mathcal{A} \times \mathcal{A} \to [0, \infty)$ defined by

$$d_\mu (A, B) := \mu (A \triangle B) = \int_\Omega |1_B - 1_A| \, d\mu$$

is a pseudo metric on $\mathcal{A}$. In general $d_\mu (A, B) = 0$ does not imply $A = B$ but only that $A = B$ modulo sets of $\mu$ – measure zero.

**Lemma 6.12.** *If $\mu : \mathcal{A} \to [0, \infty)$ is a finitely additive measure, then $\mu$ is Lip-1 relative to $d_\mu$, i.e.*

$$|\mu (B) - \mu (A)| \le d_\mu (A, B) \;\; \forall \; A, B \in \mathcal{A}.$$

**Proof. First proof.** By the basic properties of the simple integral we find,

$$|\mu (B) - \mu (A)| = \left| \int_\Omega (1_B - 1_A) \, d\mu \right| \le \int_\Omega |1_B - 1_A| \, d\mu = \mu (B \triangle A) = d_\mu (A, B) .$$

**Second proof.** Using the basic properties of measures we have,

$$\begin{aligned} \mu (B) - \mu (A) &= \mu (B \setminus A) + \mu (B \cap A) - [\mu (A \setminus B) + \mu (B \cap A)] \\ &= \mu (B \setminus A) - \mu (A \setminus B) . \end{aligned}$$

Basic inequalities then give

$$|\mu (B) - \mu (A)| \le \mu (B \setminus A) + \mu (A \setminus B) = \mu (B \triangle A) = d_\mu (A, B) .$$

■

Our proof strategy for constructing $\sigma$ – additive measures is now as follows. 1) we look for a pseudo-metric $d$ on $2^\Omega$ such that $d = d_\mu$ on $\mathcal{A}$ and then 2) we extend $\mu$ to a Lip –1 function ($\bar{\mu}$) on $\bar{\mathcal{A}}^d$ by continuity using Lemma 2.28. It will turn out that if we choose $d$ sufficiently carefully (i.e. sufficiently "small"), then $\mathcal{B} := \bar{\mathcal{A}}^d$ will be a $\sigma$ – algebra and $\bar{\mu}$ will be a measure on $\mathcal{B}$. The outcome of this startegy is summarized in the next theorem.

**Theorem 6.13 (Finite premeasure extension theorem II.).** *Let $(\Omega, \mathcal{A}, \mu)$ be a premeasure space with $\mu (\Omega) < \infty$ and define $\mu^* : 2^\Omega \to [0, \mu (\Omega)]$ by*

$$\mu^* (B) = \inf \left\{ \sum_{n=1}^\infty \mu (A_n) : A_n \in \mathcal{A} \text{ with } B \subset \cup_{n=1}^\infty A_n \right\} .$$

*Further let $\mathcal{B} \subset 2^\Omega$ consist of those $B \subset \Omega$ such that there exists $\{A_n\} \subset \mathcal{A}$ such that $\lim_{n\to\infty} \mu^* (A_n \triangle B) = 0$. Then;*

1. *$\mathcal{B}$ is a $\sigma$ -algebra containing $\mathcal{A}$,*
2. *if $B \in \mathcal{B}$ and $\{A_n\} \subset \mathcal{A}$ with and $B \subset \Omega$ such that $\lim_{n\to\infty} \mu^* (A_n \triangle B) = 0$, then $\bar{\mu} (B) := \lim_{n\to\infty} \mu (A_n)$ exists and gives a well defined $\sigma$ – additive measure, $\bar{\mu} : \mathcal{B} \to [0, \infty)$ extending $\mu$ on $\mathcal{A}$.*

We will begin the proof of this result in section 6.4 below and it will be completed with the proof of Theorem 6.44. Before doing so we pause for a couple of optional sections pertaining to the uniqueness of the extensions and their continuity properties.

## 6.2 * The $\pi - \lambda$ and monotone class theorems

This section may and probably should be omitted on first reading except for the two Exercises 6.12 and 6.10 which can be done independently of the material in this section. Later in Chapter 10 we will come back to a function theoretic variant of the results in this chapter which we will tend to use throughout the book.

### 6.2.1 The $\pi - \lambda$ theorem.

The goal of this section is to immediately discuss the uniqueness statement in Theorem 6.3. [We will nevertheless prove the most important uniqueness assertions by other means at the same time we carry out the construction of extensions.] Recall that a collection, $\mathcal{P} \subset 2^\Omega$, is a $\pi$ – **class** or $\pi$ – system if it is closed under finite intersections. We also need the notion of a $\lambda$ –system.

**Definition 6.14 ($\lambda$ – system).** *A collection of sets, $\mathcal{L} \subset 2^\Omega$, is $\lambda$ – **class** or $\lambda$ – **system** if*

a. *$\Omega \in \mathcal{L}$*
b. *If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$. (Closed under proper differences.)*
c. *If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$. (Closed under countable increasing unions.)*

*Remark 6.15.* If $\mathcal{L}$ is a collection of subsets of $\Omega$ which is both a $\lambda$ – class and a $\pi$ – system then $\mathcal{L}$ is a $\sigma$ – algebra. Indeed, since $A^c = \Omega \setminus A$, we see that any $\lambda$ - system is closed under complementation. If $\mathcal{L}$ is also a $\pi$ – system, it is closed under intersections and therefore $\mathcal{L}$ is an algebra. Since $\mathcal{L}$ is also closed under increasing unions, $\mathcal{L}$ is a $\sigma$ – algebra.

**Lemma 6.16 (Alternate Axioms for a $\lambda$ – System*).** *Suppose that $\mathcal{L} \subset 2^\Omega$ is a collection of subsets $\Omega$. Then $\mathcal{L}$ is a $\lambda$ – class iff $\lambda$ satisfies the following postulates:*

1. $\Omega \in \mathcal{L}$
2. $A \in \mathcal{L}$ *implies* $A^c \in \mathcal{L}$. *(Closed under complementation.)*
3. *If* $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ *are disjoint, then* $\sum_{n=1}^\infty A_n \in \mathcal{L}$. *(Closed under disjoint unions.)*

**Proof.** Suppose that $\mathcal{L}$ satisfies a. – c. above. Clearly then postulates 1. and 2. hold. Suppose that $A, B \in \mathcal{L}$ such that $A \cap B = \emptyset$, then $A \subset B^c$ and

$$A^c \cap B^c = B^c \setminus A \in \mathcal{L}.$$

Taking complements of this result shows $A \cup B \in \mathcal{L}$ as well. So by induction, $B_m := \sum_{n=1}^m A_n \in \mathcal{L}$. Since $B_m \uparrow \sum_{n=1}^\infty A_n$ it follows from postulate c. that $\sum_{n=1}^\infty A_n \in \mathcal{L}$.

Now suppose that $\mathcal{L}$ satisfies postulates 1. – 3. above. Notice that $\emptyset \in \mathcal{L}$ and by postulate 3., $\mathcal{L}$ is closed under finite disjoint unions. Therefore if $A, B \in \mathcal{L}$ with $A \subset B$, then $B^c \in \mathcal{L}$ and $A \cap B^c = \emptyset$ allows us to conclude that $A \cup B^c \in \mathcal{L}$. Taking complements of this result shows $B \setminus A = A^c \cap B \in \mathcal{L}$ as well, i.e. postulate *b.* holds. If $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $B_n := A_n \setminus A_{n-1} \in \mathcal{L}$ for all $n$, where by convention $A_0 = \emptyset$. Hence it follows by postulate 3 that $\cup_{n=1}^\infty A_n = \sum_{n=1}^\infty B_n \in \mathcal{L}$. ∎

**Theorem 6.17 (Dynkin's $\pi - \lambda$ Theorem).** *If* $\mathcal{L}$ *is a* $\lambda$ *class which contains a* $\pi$ *– class,* $\mathcal{P}$*, then* $\sigma(\mathcal{P}) \subset \mathcal{L}$.

**Proof.** We start by proving the following assertion; for any element $C \in \mathcal{L}$, the collection of sets,

$$\mathcal{L}^C := \{D \in \mathcal{L} : C \cap D \in \mathcal{L}\},$$

is a $\lambda$ – system. To prove this claim, observe that: a. $\Omega \in \mathcal{L}^C$, b. if $A \subset B$ with $A, B \in \mathcal{L}^C$, then $A \cap C$, $B \cap C \in \mathcal{L}$ with $A \cap C \subset B \cap C$ and therefore,

$$(B \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.$$

This shows that $\mathcal{L}^C$ is closed under proper differences. c. If $A_n \in \mathcal{L}^C$ with $A_n \uparrow A$, then $A_n \cap C \in \mathcal{L}$ and $A_n \cap C \uparrow A \cap C \in \mathcal{L}$, i.e. $A \in \mathcal{L}^C$. Hence we have verified $\mathcal{L}^C$ is still a $\lambda$ – system.

For the rest of the proof, we may assume without loss of generality that $\mathcal{L}$ is the smallest $\lambda$ – class containing $\mathcal{P}$ – if not just replace $\mathcal{L}$ by the intersection of all $\lambda$ – classes containing $\mathcal{P}$. Then for $C \in \mathcal{P}$ we know that $\mathcal{L}^C \subset \mathcal{L}$ is a $\lambda$ - class containing $\mathcal{P}$ and hence $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{P}$ was arbitrary, we have shown, $C \cap D \in \mathcal{L}$ for all $C \in \mathcal{P}$ and $D \in \mathcal{L}$. We may now conclude that if $C \in \mathcal{L}$, then $\mathcal{P} \subset \mathcal{L}^C \subset \mathcal{L}$ and hence again $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{L}$ is arbitrary, we have shown $C \cap D \in \mathcal{L}$ for all $C, D \in \mathcal{L}$, i.e. $\mathcal{L}$ is a $\pi$ – system. So by Remark 6.15, $\mathcal{L}$ is a $\sigma$ algebra. Since $\sigma(\mathcal{P})$ is the smallest $\sigma$ – algebra containing $\mathcal{P}$ it follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$. ∎

As an immediate corollary, we have the following uniqueness result.

**Proposition 6.18.** *Suppose that* $\mathcal{P} \subset 2^\Omega$ *is a* $\pi$ *– system. If* $P$ *and* $Q$ *are two probability*[1] *measures on* $\sigma(\mathcal{P})$ *such that* $P = Q$ *on* $\mathcal{P}$*, then* $P = Q$ *on* $\sigma(\mathcal{P})$.

**Proof.** Let $\mathcal{L} := \{A \in \sigma(\mathcal{P}) : P(A) = Q(A)\}$. One easily shows $\mathcal{L}$ is a $\lambda$ – class which contains $\mathcal{P}$ by assumption. Indeed, $\Omega \in \mathcal{P} \subset \mathcal{L}$, if $A, B \in \mathcal{L}$ with $A \subset B$, then

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$$

so that $B \setminus A \in \mathcal{L}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} Q(A_n) = Q(A)$ which shows $A \in \mathcal{L}$. Therefore $\sigma(\mathcal{P}) \subset \mathcal{L} = \sigma(\mathcal{P})$ and the proof is complete. ∎

*Example 6.19.* Let $\Omega := \{a, b, c, d\}$ and let $\mu$ and $\nu$ be the probability measure on $2^\Omega$ determined by, $\mu(\{x\}) = \frac{1}{4}$ for all $x \in \Omega$ and $\nu(\{a\}) = \nu(\{d\}) = \frac{1}{8}$ and $\nu(\{b\}) = \nu(\{c\}) = 3/8$. In this example,

$$\mathcal{L} := \left\{A \in 2^\Omega : P(A) = Q(A)\right\}$$

is $\lambda$ – system which is not an algebra. Indeed, $A = \{a, b\}$ and $B = \{a, c\}$ are in $\mathcal{L}$ but $A \cap B \notin \mathcal{L}$.

**Exercise 6.1.** Suppose that $\mu$ and $\nu$ are two measures (not assumed to be finite) on a measure space, $(\Omega, \mathcal{B})$ such that $\mu = \nu$ on a $\pi$ – system, $\mathcal{P}$. Further assume $\mathcal{B} = \sigma(\mathcal{P})$ and there exists $\Omega_n \in \mathcal{P}$ such that; i) $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ for all $n$ and ii) $\Omega_n \uparrow \Omega$ as $n \uparrow \infty$. Show $\mu = \nu$ on $\mathcal{B}$.

**Hint:** Consider the measures, $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$.

**Corollary 6.20.** *A probability measure,* $P$*, on* $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ *is uniquely determined by its cumulative distribution function,*

$$F(x) := P((-\infty, x]).$$

**Proof.** This follows from Proposition 6.18 wherein we use the fact that $\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}$ is a $\pi$ – system such that $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{P})$. ∎

*Remark 6.21.* Corollary 6.20 generalizes to $\mathbb{R}^n$. Namely a probability measure, $P$, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is uniquely determined by its CDF,

$$F(x) := P((-\infty, x]) \text{ for all } x \in \mathbb{R}^n$$

where now

$$(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n].$$

[1] More generally, $P$ and $Q$ could be two measures such that $P(\Omega) = Q(\Omega) < \infty$.

**Exercise 6.2 (Density of $\mathcal{A}$ in $\sigma(\mathcal{A})$).** Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(\mathcal{A})$, and $P$ is a probability measure on $\mathcal{B}$. Let $\rho(A, B) := P(A \triangle B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $\mathcal{A}$ is dense in $\mathcal{B}$ relative to the "metric," $\rho$. More precisely you are to show using the following outline that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \triangle B) < \varepsilon$.

1. Recall from Exercise 5.6 that $\rho(a, B) = P(A \triangle B) = \mathbb{E}|1_A - 1_B|$.
2. Observe; if $B = \cup B_i$ and $A = \cup_i A_i$, then
$$\begin{aligned} B \setminus A &= \cup_i [B_i \setminus A] \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and} \\ A \setminus B &= \cup_i [A_i \setminus B] \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i \end{aligned}$$
so that
$$A \triangle B \subset \cup_i (A_i \triangle B_i).$$
3. We also have
$$\begin{aligned} (B_2 \setminus B_1) \setminus (A_2 \setminus A_1) &= B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c \\ &= B_2 \cap B_1^c \cap (A_2 \cap A_1^c)^c \\ &= B_2 \cap B_1^c \cap (A_2^c \cup A_1) \\ &= [B_2 \cap B_1^c \cap A_2^c] \cup [B_2 \cap B_1^c \cap A_1] \\ &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \end{aligned}$$
and similarly,
$$(A_2 \setminus A_1) \setminus (B_2 \setminus B_1) \subset (A_2 \setminus B_2) \cup (B_1 \setminus A_1)$$
so that
$$\begin{aligned} (A_2 \setminus A_1) \triangle (B_2 \setminus B_1) &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1) \\ &= (A_1 \triangle B_1) \cup (A_2 \triangle B_2). \end{aligned}$$
4. Observe that $A_n \in \mathcal{B}$ and $A_n \uparrow A$, then
$$\begin{aligned} P(B \triangle A_n) &= P(B \setminus A_n) + P(A_n \setminus B) \\ &\to P(B \setminus A) + P(A \setminus B) = P(A \triangle B). \end{aligned}$$
5. Let $\mathcal{L}$ be the collection of sets $B \in \mathcal{B}$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a $\lambda -$ system which contains $\mathcal{A}$.

### 6.2.2 ** Monotone Class Theorems

In this subsection we record a theorem which is a cousin of the $\pi - \lambda$ theorem.

**Definition 6.22 (Montone Class).** *$\mathcal{C} \subset 2^\Omega$ is a **monotone class** if it is closed under countable increasing unions and countable decreasing intersections.*

**Lemma 6.23 (Monotone Class Theorem*).** *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$. Then $\mathcal{C} = \sigma(\mathcal{A})$.*

**Proof.** For $C \in \mathcal{C}$ let
$$\mathcal{C}(C) = \{B \in \mathcal{C} : C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}\},$$
then $\mathcal{C}(C)$ is a monotone class. Indeed, if $B_n \in \mathcal{C}(C)$ and $B_n \uparrow B$, then $B_n^c \downarrow B^c$ and so
$$\begin{aligned} &\mathcal{C} \ni C \cap B_n \uparrow C \cap B \\ &\mathcal{C} \ni C \cap B_n^c \downarrow C \cap B^c \text{ and} \\ &\mathcal{C} \ni B_n \cap C^c \uparrow B \cap C^c. \end{aligned}$$
Since $\mathcal{C}$ is a monotone class, it follows that $C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}$, i.e. $B \in \mathcal{C}(C)$. This shows that $\mathcal{C}(C)$ is closed under increasing limits and a similar argument shows that $\mathcal{C}(C)$ is closed under decreasing limits. Thus we have shown that $\mathcal{C}(C)$ is a monotone class for all $C \in \mathcal{C}$. If $A \in \mathcal{A} \subset \mathcal{C}$, then $A \cap B, A \cap B^c, B \cap A^c \in \mathcal{A} \subset \mathcal{C}$ for all $B \in \mathcal{A}$ and hence it follows that $\mathcal{A} \subset \mathcal{C}(A) \subset \mathcal{C}$. Since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$, we conclude that $\mathcal{C}(A) = \mathcal{C}$ for any $A \in \mathcal{A}$. Let $B \in \mathcal{C}$ and notice that $A \in \mathcal{C}(B)$ happens iff $B \in \mathcal{C}(A)$. This observation and the fact that $\mathcal{C}(A) = \mathcal{C}$ for all $A \in \mathcal{A}$ implies $\mathcal{A} \subset \mathcal{C}(B) \subset \mathcal{C}$ for all $B \in \mathcal{C}$. Again since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(B)$ is a monotone class we conclude that $\mathcal{C}(B) = \mathcal{C}$ for all $B \in \mathcal{C}$. That is to say, if $A, B \in \mathcal{C}$ then $A \in \mathcal{C} = \mathcal{C}(B)$ and hence $A \cap B$, $A \cap B^c$, $A^c \cap B \in \mathcal{C}$. So $\mathcal{C}$ is closed under complements (since $\Omega \in \mathcal{A} \subset \mathcal{C}$) and finite intersections and increasing unions from which it easily follows that $\mathcal{C}$ is a $\sigma -$ algebra. ■

## 6.3 * $\sigma$ - Algebra Regularity and Uniqueness Results

The goal of this appendix it to approximating measurable sets from inside and outside by classes of sets which are relatively easy to understand. Our first few results are already contained in Carathoédory's existence of measures proof. Nevertheless, we state these results again and give another somewhat independent proof.

**Theorem 6.24 (Finite Regularity Result).** *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} = \sigma(\mathcal{A})$ and $\mu : \mathcal{B} \to [0, \infty)$ is a finite measure, i.e. $\mu(\Omega) < \infty$. Then for every $\varepsilon > 0$ and $B \in \mathcal{B}$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$.*

**Proof.** Let $\mathcal{B}_0$ denote the collection of $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there here exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$. It is now clear that $\mathcal{A} \subset \mathcal{B}_0$ and that $\mathcal{B}_0$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}_0$ for $i = 1, 2, \ldots$ and $\varepsilon > 0$ is given. By assumption there exists $A_i \in \mathcal{A}_\delta$ and $C_i \in \mathcal{A}_\sigma$ such that $A_i \subset B_i \subset C_i$ and $\mu(C_i \setminus A_i) < 2^{-i}\varepsilon$.

Let $A := \cup_{i=1}^\infty A_i$, $A^N := \cup_{i=1}^N A_i \in \mathcal{A}_\delta$, $B := \cup_{i=1}^\infty B_i$, and $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$. Then $A^N \subset A \subset B \subset C$ and

$$C \setminus A = [\cup_{i=1}^\infty C_i] \setminus A = \cup_{i=1}^\infty [C_i \setminus A] \subset \cup_{i=1}^\infty [C_i \setminus A_i].$$

Therefore,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^\infty [C_i \setminus A]) \le \sum_{i=1}^\infty \mu(C_i \setminus A) \le \sum_{i=1}^\infty \mu(C_i \setminus A_i) < \varepsilon.$$

Since $C \setminus A^N \downarrow C \setminus A$, it also follows that $\mu\left(C \setminus A^N\right) < \varepsilon$ for sufficiently large $N$ and this shows $B = \cup_{i=1}^\infty B_i \in \mathcal{B}_0$. Hence $\mathcal{B}_0$ is a sub-$\sigma$-algebra of $\mathcal{B} = \sigma(\mathcal{A})$ which contains $\mathcal{A}$ which shows $\mathcal{B}_0 = \mathcal{B}$. ∎

Many theorems in the sequel will require some control on the size of a measure $\mu$. The relevant notion for our purposes (and most purposes) is that of a $\sigma$ – finite measure defined next.

**Definition 6.25.** *Suppose $\Omega$ is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^\Omega$ and $\mu : \mathcal{B} \to [0, \infty]$ is a function. The function $\mu$ is $\sigma$ – finite on $\mathcal{E}$ if there exists $E_n \in \mathcal{E}$ such that $\mu(E_n) < \infty$ and $\Omega = \cup_{n=1}^\infty E_n$. If $\mathcal{B}$ is a $\sigma$ – algebra and $\mu$ is a measure on $\mathcal{B}$ which is $\sigma$ – finite on $\mathcal{B}$ we will say $(\Omega, \mathcal{B}, \mu)$ is a $\sigma$ – finite measure space.*

The reader should check that if $\mu$ is a finitely additive measure on an algebra, $\mathcal{B}$, then $\mu$ is $\sigma$ – finite on $\mathcal{B}$ iff there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$.

**Corollary 6.26 ($\sigma$ – Finite Regularity Result).** *Theorem 6.24 continues to hold under the weaker assumption that $\mu : \mathcal{B} \to [0, \infty]$ is a measure which is $\sigma$ – finite on $\mathcal{A}$.*

**Proof.** Let $\Omega_n \in \mathcal{A}$ such that $\cup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all $n$.Since $A \in \mathcal{B} \to \mu_n(A) := \mu(\Omega_n \cap A)$ is a finite measure on $A \in \mathcal{B}$ for each $n$, by Theorem 6.24, for every $B \in \mathcal{B}$ there exists $C_n \in \mathcal{A}_\sigma$ such that $B \subset C_n$ and $\mu(\Omega_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n}\varepsilon$. Now let $C := \cup_{n=1}^\infty [\Omega_n \cap C_n] \in \mathcal{A}_\sigma$ and observe that $B \subset C$ and

$$\begin{aligned}\mu(C \setminus B) &= \mu(\cup_{n=1}^\infty ([\Omega_n \cap C_n] \setminus B)) \\ &\le \sum_{n=1}^\infty \mu([\Omega_n \cap C_n] \setminus B) = \sum_{n=1}^\infty \mu(\Omega_n \cap [C_n \setminus B]) < \varepsilon.\end{aligned}$$

Applying this result to $B^c$ shows there exists $D \in \mathcal{A}_\sigma$ such that $B^c \subset D$ and

$$\mu(B \setminus D^c) = \mu(D \setminus B^c) < \varepsilon.$$

So if we let $A := D^c \in \mathcal{A}_\delta$, then $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu([B \setminus A] \cup [(C \setminus B) \setminus A]) \le \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon$$

and the result is proved. ∎

**Exercise 6.3.** Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu$ and $\nu$ are two measures on $\mathcal{B} = \sigma(\mathcal{A})$.

a. Suppose that $\mu$ and $\nu$ are finite measures such that $\mu = \nu$ on $\mathcal{A}$. Show $\mu = \nu$.
b. Generalize the previous assertion to the case where you only assume that $\mu$ and $\nu$ are $\sigma$ – finite on $\mathcal{A}$.

**Corollary 6.27.** *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu : \mathcal{B} = \sigma(\mathcal{A}) \to [0, \infty]$ is a measure which is $\sigma$ – finite on $\mathcal{A}$. Then for all $B \in \mathcal{B}$, there exists $A \in \mathcal{A}_{\delta\sigma}$ and $C \in \mathcal{A}_{\sigma\delta}$ such that $A \subset B \subset C$ and $\mu(C \setminus A) = 0$.*

**Proof.** By Theorem 6.24, given $B \in \mathcal{B}$, we may choose $A_n \in \mathcal{A}_\delta$ and $C_n \in \mathcal{A}_\sigma$ such that $A_n \subset B \subset C_n$ and $\mu(C_n \setminus B) \le 1/n$ and $\mu(B \setminus A_n) \le 1/n$. By replacing $A_N$ by $\cup_{n=1}^N A_n$ and $C_N$ by $\cap_{n=1}^N C_n$, we may assume that $A_n \uparrow$ and $C_n \downarrow$ as $n$ increases. Let $A = \cup A_n \in \mathcal{A}_{\delta\sigma}$ and $C = \cap C_n \in \mathcal{A}_{\sigma\delta}$, then $A \subset B \subset C$ and

$$\begin{aligned}\mu(C \setminus A) &= \mu(C \setminus B) + \mu(B \setminus A) \le \mu(C_n \setminus B) + \mu(B \setminus A_n) \\ &\le 2/n \to 0 \text{ as } n \to \infty.\end{aligned}$$

∎

## 6.4 Outer Measures

**Definition 6.28 (Outer measures).** *Let $\Omega$ be a non-empty set. A function $\nu : 2^\Omega \to [0, \infty]$ is an **outer measure** (on $\Omega$) if;*

1. *$\nu(\emptyset) = 0$,*
2. *$\nu$ is **monotonic** (i.e. $\nu(A) \le \nu(B)$ whenever $A \subset B$) and*
3. *$\nu$ is **countably sub-additive** (i.e. $\nu(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \nu(A_n)$ for all $\{A_n\}_{n=1}^\infty \subset 2^\Omega$).*

*If item 3. is replaced by finite sub-additivity (i.e. $\nu(A \cup B) \le \nu(A) + \nu(B)$ for all $A, B \in 2^\Omega$) then we say $\nu$ is a **weak outer measure** on $\Omega$.*

**Proposition 6.29 (Example of an outer measure.).** *Let $\mathcal{E} \subset 2^\Omega$ be arbitrary collection of subsets of $\Omega$ such that $\emptyset, \Omega \in \mathcal{E}$. Let $\rho : \mathcal{E} \to [0,\infty]$ be a function such that $\rho(\emptyset) = 0$. For any $A \subset \Omega$, define*

$$\rho^*(A) = \inf\left\{\sum_{k=1}^\infty \rho(E_k) : A \subset \bigcup_{k=1}^\infty E_k \text{ with } E_k \in \mathcal{E}\right\}. \tag{6.2}$$

*Then $\rho^*$ is an outer measure.*

**Proof.** It is clear that $\rho^*$ is monotonic and $\rho^*(\emptyset) = 0$. Suppose $\{A_n\}_{n=1}^\infty \subset 2^\Omega$ and $\varepsilon > 0$ is given. By definition of $\rho^*(A_n)$, there exists $\{E_{nk}\}_{k=1}^\infty \subset \mathcal{E}$ such that $A_n \subset \bigcup_{k=1}^\infty E_{nk}$ and $\sum_{k=1}^\infty \rho(E_{nk}) \le \rho^*(A_n) + 2^{-n}\varepsilon$. Using these inequalities and the fact that $\bigcup_{n=1}^\infty A_n \subset \bigcup_{n,k=1}^\infty E_{nk}$, it follows that

$$\rho^*\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty\sum_{k=1}^\infty \rho(E_{nk}) \le \sum_{n=1}^\infty \left[\rho^*(A_n) + 2^{-n}\varepsilon\right] = \sum_{n=1}^\infty \rho^*(A_n) + \varepsilon.$$

The sub-additivity of $\rho^*$ then follows from this inequality as $\varepsilon > 0$ was arbitrary. ■

**Exercise 6.4 (*).** If $\mathcal{C}$ is a collection of subsets of $\Omega$ such that $\emptyset, \Omega \in \mathcal{C}$ and $\mathcal{C}$ is closed under finite unions and $\mu : \mathcal{C} \to [0,\infty]$ is a monotone and finitely subadditive on $\mathcal{C}$, then

$$\mu^\#(B) := \inf\{\mu(C) : B \subset C \in \mathcal{C}\}$$

is a weak outer measure on $\Omega$.

*Example 6.30 (*).* If $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure on a sub-algebra, $\mathcal{A} \subset 2^\Omega$, then

$$\mu^\#(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}\}$$

is a weak-outer measure on $\Omega$. [Weak outer measures are typically not going to be sufficient for constructing measures, see Exercise 6.6 below.]

**Lemma 6.31.** *If $(\Omega, \mathcal{A}, \mu)$ is a finitely additive measure space $\mu^*$ is the outer measure associated to $\mu$ as in Proposition 6.29, then for all $B \in 2^\Omega$ we have*

$$\mu^*(B) = \inf\left\{\lim_{n\to\infty}\mu(C_n) : \{C_n\}_{n=1}^\infty \subset \mathcal{A} \ni C_n \uparrow C_\infty \supset B\right\}. \tag{6.3}$$

**Proof.** Let $M \in [0,\infty]$ be the value of the right side of Eq. (6.3). If $\{C_n\}_{n=1}^\infty \subset \mathcal{A} \ni C_n \uparrow C_\infty \supset B$, then $B \subset C_\infty = \sum_{n=1}^\infty [C_n \setminus C_{n-1}]$ where $C_0 := \emptyset$. By definition of $\mu^*$ and the finite additivity of $\mu$ we conclude,

$$\begin{aligned}\mu^*(B) &\le \sum_{n=1}^\infty \mu(C_n \setminus C_{n-1}) = \lim_{N\to\infty}\sum_{n=1}^N \mu(C_n \setminus C_{n-1}) \\ &= \lim_{N\to\infty}\mu\left(\sum_{n=1}^N [C_n \setminus C_{n-1}]\right) = \lim_{N\to\infty}\mu(C_N)\end{aligned} \tag{6.4}$$

and this shows $\mu^*(B) \le M$. So we now have to show $\mu^*(B) \ge M$.

Suppose that $\{B_n\}_{n=1}^\infty \subset \mathcal{A}$ with $B \subset \cup_{n=1}^\infty B_n$. Let $B_0 = \tilde{B}_0 = \emptyset$ and $\tilde{B}_n := B_n \setminus [B_1 \cup \dots \cup B_{n-1}]$ for $n \in \mathbb{N}$. Then $\left\{\tilde{B}_n\right\}_{n=1}^\infty$ are pairwise disjoint, $\tilde{B}_n \subset B_n$ for all $n$, and $C_N := \cup_{n=1}^N B_n = \cup_{n=1}^N \tilde{B}_n$ for all $N \in \mathbb{N}\cup\{\infty\}$. Using the monotonicity of $\mu$ and the finite additivity of $\mu$ as in the previous paragraph we conclude that

$$\sum_{n=1}^\infty \mu(B_n) \ge \sum_{n=1}^\infty \mu\left(\tilde{B}_n\right) = \lim_{N\to\infty}\mu(C_N) \ge M.$$

As the sequence $\{B_n\}_{n=1}^\infty \subset \mathcal{A}$ with $B \subset \cup_{n=1}^\infty B_n$ was arbitrary we conclude from this previous equation that $\mu^*(B) \ge M$. ■

*Remark 6.32.* Let us recall from Proposition 6.4 and Remark 6.5 that a finitely additive measure $\mu : \mathcal{A} \to [0,\infty]$ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$.

**Corollary 6.33.** *If $(\Omega, \mathcal{A}, \mu)$ is a finitely additive measure space then $\mu^* \le \mu$ on $\mathcal{A}$ with equality on $\mathcal{A}$ iff $\mu$ is a premeasure.*

**Proof.** It should be clear that $\mu^* \le \mu$ on $\mathcal{A}$. If $\mu$ is not a premeasure, then there exists $A \in \mathcal{A}$ and $\{A_n\}_{n=1}^\infty$ such that $A_n \uparrow A$ yet $\lim_{n\to\infty}\mu(A_n) < \mu(A)$. By Lemma 6.31 we know that $\mu^*(A) \le \lim_{n\to\infty}\mu(A_n)$ and therefore $\mu^*(A) < \mu(A)$. ■

**Definition 6.34.** *Given a collection of subsets, $\mathcal{E}$, of $\Omega$, let $\mathcal{E}_\sigma$ denote the collection of subsets of $\Omega$ which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_\delta$ denote the collection of subsets of $\Omega$ which are finite or countable intersections of sets from $\mathcal{E}$. We also write $\mathcal{E}_{\sigma\delta} = (\mathcal{E}_\sigma)_\delta$ and $\mathcal{E}_{\delta\sigma} = (\mathcal{E}_\delta)_\sigma$, etc.*

**Lemma 6.35.** *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra. Then:*

1. *$\mathcal{A}_\sigma$ is closed under taking countable unions and finite intersections.*

2. *The map,* $\mathcal{A}_\sigma \ni A \to A^c \in \mathcal{A}_\delta$ *is a bijection.*
3. $\mathcal{A}_\delta$ *is closed under taking countable intersections and finite unions.*

**Proof. 1.** By construction $\mathcal{A}_\sigma$ is closed under countable unions. Moreover if $A = \cup_{i=1}^\infty A_i$ and $B = \cup_{j=1}^\infty B_j$ with $A_i, B_j \in \mathcal{A}$, then

$$A \cap B = \cup_{i,j=1}^\infty A_i \cap B_j \in \mathcal{A}_\sigma,$$

which shows that $\mathcal{A}_\sigma$ is also closed under finite intersections.

**2.** If $A_i \in \mathcal{A}$ and $A = \cup_{i=1}^\infty A_i \in \mathcal{A}_\sigma$, then $A^c = \cap_{i=1}^\infty A_i^c \in \mathcal{A}_\delta$ and visa versa.

**3.** This item follows directly from items 1. and 2. or may be proved directly. ■

**Proposition 6.36.** *Let* $(\Omega, \mathcal{A}, \mu)$ *be a premeasure space.*

1. *If* $A \in \mathcal{A}_\sigma$ *and* $\{A_n\} \subset \mathcal{A}$ *are such that* $A_n \uparrow A$, *then* $\mu(A_n) \uparrow \mu^*(A)$ *as* $n \to \infty$.
2. *If* $B \in \mathcal{A}_\sigma$, *then*

$$\mu^*(B) := \sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \text{ for all } B \in \mathcal{A}_\sigma. \tag{6.5}$$

3. *If* $B \in 2^\Omega$, *then*

$$\mu^*(B) = \inf\{\mu^*(C) : B \subset C \in \mathcal{A}_\sigma\} \text{ for all } B \in 2^\Omega.$$

**Proof.** 1. Suppose that $\{C_n\}_{n=1}^\infty \subset \mathcal{A} \ni C_n \uparrow C_\infty \supset A$. Then for each $n \in \mathbb{N}$, $A_n \cap C_m \uparrow A_n \cap C_\infty = A_n$ as $m \uparrow \infty$ and as $\mu$ is a premeasure we may conclude that

$$\mu(A_n) = \lim_{m\to\infty} \mu(A_n \cap C_m) \le \lim_{m\to\infty} \mu(C_m).$$

Passing to the limit as $n \to \infty$ in this equation shows $\lim_{n\to\infty} \mu(A_n) \le \lim_{m\to\infty} \mu(C_m)$. Taking the infimum over all $\{C_n\}_{n=1}^\infty \subset \mathcal{A} \ni C_n \uparrow C_\infty \supset A$ implies $\lim_{n\to\infty} \mu(A_n) \le \mu^*(A)$. However as we saw in Eq. (6.4), we always have $\mu^*(A) \le \lim_{n\to\infty} \mu(A_n)$.

2. Since if $A \in \mathcal{A}$ and $A \subset B \in \mathcal{A}_\sigma$, we have $\mu(A) = \mu^*(A) \le \mu^*(B)$ and so

$$\sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \le \mu^*(B).$$

On the other hand, since $B \in \mathcal{A}_\sigma$ there exists $A_n \in \mathcal{A}$ such that $A_n \uparrow B$ and therefore

$$\sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \ge \sup_n \mu(A_n) = \lim_{n\to\infty} \mu(A_n) = \mu^*(B)$$

where the last equality follows by item 1.

3. From Lemma 6.31 and item 1.,

$$\begin{aligned}\mu^*(B) &= \inf\left\{\lim_{n\to\infty} \mu(C_n) : \{C_n\}_{n=1}^\infty \subset \mathcal{A} \ni C_n \uparrow C_\infty \supset B\right\} \\ &= \inf\{\mu^*(C_\infty) : B \subset C_\infty\}.\end{aligned}$$

■

**Notation 6.37** *When* $(\Omega, \mathcal{A}, \mu)$ *is a premeasure space we let* $\mu_\sigma := \mu^*|_{\mathcal{A}_\sigma}$ *where* $\mu^*$ *is the outer measure associated to* $\mu$.

Although the notation $\mu_\sigma$ is not strictly needed, we will sometimes write $\mu_\sigma(A)$ rather than $\mu^*(A)$ to emphasize $A \in \mathcal{A}_\sigma$ and is not a general element of $2^\Omega$. As we have seen in Proposition 6.36, $\mu_\sigma = \mu^*|_{\mathcal{A}_\sigma}$ is relatively easy to compute and understand and $\mu^*$ may be computed (according to item 3. of Proposition 6.36) using

$$\mu^*(B) = \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\} \text{ for all } B \in 2^\Omega. \tag{6.6}$$

**Proposition 6.38.** *Given a premeasure,* $\mu : \mathcal{A} \to [0, \infty]$, *satisfies;*

1. *(**Monotonicity**) If* $A, B \in \mathcal{A}_\sigma$ *with* $A \subset B$ *then* $\mu_\sigma(A) \le \mu_\sigma(B)$.
2. *(**Continuity**) If* $A_n \in \mathcal{A}$ *and* $A_n \uparrow B \in \mathcal{A}_\sigma$, *then* $\mu(A_n) \uparrow \mu_\sigma(B)$ *as* $n \to \infty$.
3. *(**Strong Additivity on*** $\mathcal{A}_\sigma$*) If* $A, B \in \mathcal{A}_\sigma$, *then*

$$\mu_\sigma(A \cup B) + \mu_\sigma(A \cap B) = \mu_\sigma(A) + \mu_\sigma(B). \tag{6.7}$$

4. *(**Sub-Additivity on*** $\mathcal{A}_\sigma$*) The function* $\mu_\sigma$ *is sub-additive on* $\mathcal{A}_\sigma$, *i.e. if* $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$, *then*

$$\mu_\sigma(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu_\sigma(A_n). \tag{6.8}$$

5. *(σ - Additivity **on*** $\mathcal{A}_\sigma$*) The function* $\mu_\sigma$ *is countably additive on* $\mathcal{A}_\sigma$. *[This item is not necessary for the proofs to follow!]*

**Proof. 1.** Monotonicity holds on all subset of $\Omega$ as $\mu^*$ is an outer measure.

**2.** Continuity was proved in Proposition 6.36.

**3.** Suppose that $A, B \in \mathcal{A}_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,

$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$

proves Eq. (6.7). In particular, it follows that $\mu_\sigma$ is finitely additive on $\mathcal{A}_\sigma$.

**4.** Let $\{A_n\}_{n=1}^\infty$ be any sequence in $\mathcal{A}_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset \mathcal{A}$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,

$$\mu\left(\cup_{n=1}^{N} A_{n,N}\right) \leq \sum_{n=1}^{N} \mu\left(A_{n,N}\right) \leq \sum_{n=1}^{N} \mu_\sigma\left(A_n\right) \leq \sum_{n=1}^{\infty} \mu_\sigma\left(A_n\right). \tag{6.9}$$

Since $\mathcal{A} \ni \cup_{n=1}^{N} A_{n,N} \uparrow \cup_{n=1}^{\infty} A_n \in \mathcal{A}_\sigma$, we may let $N \to \infty$ in Eq. (6.9) to conclude Eq. (6.8) holds.

**5.** Let us now further assume that $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}_\sigma$ are pairwise disjoint. By the finite additivity and monotonicity of $\mu_\sigma$ on $\mathcal{A}_\sigma$, we have

$$\sum_{n=1}^{\infty} \mu_\sigma\left(A_n\right) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu_\sigma\left(A_n\right) = \lim_{N\to\infty} \mu_\sigma\left(\cup_{n=1}^{N} A_n\right) \leq \mu_\sigma\left(\cup_{n=1}^{\infty} A_n\right).$$

This inequality along with Eq. (6.8) shows that $\mu_\sigma$ is $\sigma$ – additive on $\mathcal{A}_\sigma$. ∎

**Corollary 6.39.** *Suppose that $(\Omega, \mathcal{A}, \mu)$ is a premeasure space. If $B \in \mathcal{A}_\sigma$ and $A \in \mathcal{A}$ with $A \subset B$, then $B \setminus A \in \mathcal{A}_\sigma$ and*

$$\mu_\sigma(B) = \mu(A) + \mu_\sigma(B \setminus A).$$

**Proof.** Choose $B_n \in \mathcal{A}$ so that $B_n \uparrow B$ as $n \to \infty$, then $\mathcal{A} \ni B_n \setminus A \uparrow B \setminus A$ as $n \to \infty$ which shows $B \setminus A \in \mathcal{A}_\sigma$. Moreover using the fact that $\mu$ is a premeasure along with item 2. of Proposition 6.38, we find

$$\mu_\sigma(B) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \left[\mu(B_n \cap A) + \mu(B_n \setminus A)\right] = \mu(A) + \mu_\sigma(B \setminus A).$$

∎

**Definition 6.40 (The outer metric).** *If $\nu$ is an (weak( outer measure on $\Omega$, the associate **(weak) outer pseudo-metric** $(d_\nu)$ on $2^\Omega$ is defined by*

$$d_\nu(A, B) := \nu(A \triangle B) \ \forall\ A, B \in 2^\Omega.$$

The following important proposition is outer measure version of Example 6.11.

**Proposition 6.41 (Properties of the outer metric).** *If (as in Definition 6.40) $d = d_\nu$ is the outer pseudo-metric associate to an outer measure $\nu : 2^\Omega \to [0, \infty]$, then;*

1. *$d$ is a pseudo metric.*
2. *$d(A^c, C^c) = d(A, C)$ for all $A, B \in 2^\Omega$.*
3. *If $\{A_n\}_{n=1}^{\infty}, \{B_n\}_{n=1}^{\infty} \subset 2^\Omega$, then*

$$d\left(\cup_{n=1}^{\infty} A_n, \cup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} d(A_n, B_n) \ \text{and} \tag{6.10}$$

$$d\left(\cap_{n=1}^{\infty} A_n, \cap_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} d(A_n, B_n). \tag{6.11}$$

*[* These results hold when $\nu$ is a weak outer measure except that in item 3. we need to replace $\infty$ by $N < \infty$ everywhere.]*

**Proof.** We take each item in turn.

1. Since $A \triangle B = B \triangle A$ it follows that $d(A, B) = d(B, A)$ and hence we need only shows that $d$ satisfies the triangle inequality. If $A, B, C \in 2^\Omega$, then
$$1_{A\triangle C} = |1_A - 1_C| \leq |1_A - 1_B| + |1_B - 1_C| = 1_{A\triangle B} + 1_{B\triangle C}.$$
From this equation (or directly from Exercise 4.4) it easily follows that
$$A \triangle C \subset [A \triangle B] \cup [B \triangle C]$$
and therefore,
$$\begin{aligned} d(A, C) &= \nu(A \triangle C) \leq \nu([A \triangle B] \cup [B \triangle C]) \\ &\leq \nu(A \triangle B) + \nu(B \triangle C) = d(A, B) + d(B, C) \end{aligned}$$
which completes the proof of the first item.
2. As
$$A^c \triangle C^c = [A^c \cap C] \cup [C^c \cap A] = [C \setminus A] \cup [A \setminus C] = A \triangle C$$
it follows that $d(A^c, C^c) = d(A, C)$. This also may be verifies alternatively as follows;
$$1_{A^c\triangle C^c} = |1_{A^c} - 1_{C^c}| = |[1 - 1_A] - [1 - 1_C]| = |1_A - 1_C| = 1_{A\triangle C}.$$
3. By Exercise 4.3,
$$\left[\cup_{n=1}^{\infty} A_n\right] \triangle \left[\cup_{n=1}^{\infty} B_n\right] \subset \cup_{n=1}^{\infty} \left[A_n \triangle B_n\right]. \tag{6.12}$$
Hence using the monotonicity and finite subadditivity of $\nu$ then gives;
$$\begin{aligned} d\left(\cup_{n=1}^{\infty} A_n, \cup_{n=1}^{\infty} B_n\right) &= \nu\left(\left[\cup_{n=1}^{\infty} A_n\right] \triangle \left[\cup_{n=1}^{\infty} B_n\right]\right) \leq \nu\left(\cup_{n=1}^{\infty} \left[A_n \triangle B_n\right]\right) \\ &\leq \sum_{n=1}^{\infty} \nu(A_n \triangle B_n) = \sum_{n=1}^{\infty} d(A_n, B_n), \end{aligned}$$
which proves Eq. (6.10). Equation (6.11) may be proved similarly or by combining item 2. with Eq. (6.10) as follows;
$$\begin{aligned} &d\left(\cap_{n=1}^{\infty} A_n, \cap_{n=1}^{\infty} B_n\right) \\ &\quad = d\left(\left[\cap_{n=1}^{\infty} A_n\right]^c, \left[\cap_{n=1}^{\infty} B_n\right]^c\right) \\ &\quad = d\left(\cup_{n=1}^{\infty} A_n^c, \cup_{n=1}^{\infty} B_n^c\right) \leq \sum_{n=1}^{\infty} d(A_n^c, B_n^c) = \sum_{n=1}^{\infty} d(A_n, B_n). \end{aligned}$$

■

**Exercise 6.5 (* ).** If $d = d_\nu$ is the weak pseudo-metric associate to a weak outer measure $\nu$ and $\mathcal{A} \subset 2^\Omega$ is a sub-algebra, then $\mathcal{B} := \bar{\mathcal{A}}^d$ is also a sub-algebra of $2^\Omega$.

**Theorem 6.42 (* A finitely additive extension theorem).** *Let $(\Omega, \mathcal{A}, \mu)$ be a finite and finitely additive measure space and let*

$$\mu^{\#}(B) := \inf\{\mu(A) : B \subset A\}$$

*be the associated weak outer measure. Further let $\bar{\mu}$ denote the unique $d_{\mu^{\#}}$ – continuous extension of $\mu$ to $\mathcal{B} := \bar{\mathcal{A}}^{d_{\mu^{\#}}}$. Then $\mathcal{B}$ is again an algebra and $\bar{\mu}$ is a finitely additive measure on $\mathcal{B}$.*

**Proof.** By Lemma 6.12 we know that $\mu$ is Lip-1 on $\mathcal{A}$ relative to $d$. It is now a simple matter to show that $\mu$ extends by uniform continuity to a function $\bar{\mu}$ on $\mathcal{B} := \bar{\mathcal{A}}^{d_\nu}$ with the same estimate holding. The fact that $\mathcal{B}$ is an algebra follows by item 4. of Proposition **??**. So it only remains to show $\bar{\mu}$ is finitely additive.

Let $B_i \in \mathcal{B}$ for $i = 1, 2$ and $\{A_n^i\} \subset \mathcal{A}$ be chosen so that $\lim_{n\to\infty} d(A_n^i, B_i) = 0$ for $i = 1, 2$. Because of Eqs. (**??**) and (**??**) and the continuity of $\bar{\mu}$ it follows that

$$\begin{aligned}\bar{\mu}(B_1 \cup B_2) + \bar{\mu}(B_1 \cap B_2) &= \lim_{n\to\infty}\left[\bar{\mu}\left(A_n^1 \cup A_n^2\right) + \bar{\mu}\left(A_n^1 \cap A_n^2\right)\right] \\ &= \lim_{n\to\infty}\left[\bar{\mu}\left(A_n^1\right) + \bar{\mu}\left(A_n^2\right)\right] = \bar{\mu}(B_1) + \bar{\mu}(B_2),\end{aligned}$$

which shows $\bar{\mu}$ is additive on $\mathcal{B}$. ■

**Exercise 6.6 (*).** Let $\Omega = (0, 1]$ and $\mathcal{A}$ be the algebra generated by sets of the form $(a, b]$ with $0 \le a \le b \le 1$ and let $\mu$ be the finitely additive length measure on $\mathcal{A}$, i.e. $\mu$ is determined uniquely by requiring $\mu((a, b]) = b - a$ for all $0 \le a \le b \le 1$. Further let $\mu^{\#}$ be the weak outer measure defined in Example 6.30 and let $B = \mathbb{Q} \cap (0, 1]$. Show $d_{\mu^{\#}}(A, B) = 1$ for all $A \in \mathcal{A}$. This shows $B \notin \bar{\mathcal{A}}^d$ and hence $\bar{\mathcal{A}}^d$ is **not** a $\sigma$ – algebra.

Exercise 6.6 indicates (by example) that weak outer measures are typically not going to be sufficient for our purposes. So in order to get the desired extension of $\mu$ on an algebra $\mathcal{A}$ it is necessary to use the smaller outer measure $\mu^* \le \mu^{\#}$ in order to get a bigger closure. This is the topic of the next section..

## 6.5 Construction of $\sigma$–Additive Finite Measures

For the rest of this section we suppose that $(\Omega, \mathcal{A}, \mu)$ is a premeasure space, $\mu^*$ is the associated outer measure on $\Omega$ as in Proposition 6.29 and Proposition 6.36. We further let

$$d(A, B) = d_{\mu^*}(A, B) = \mu^*(A \triangle B) \text{ for all } A, B \in 2^\Omega$$

be the associated pseudo-metric as in Definition 6.40, and $\mathcal{B} := \bar{\mathcal{A}}^d$ denote the closure of $\mathcal{A}$ relative to $d$.

**Proposition 6.43.** *If $(\Omega, \mathcal{A}, \mu)$ is a finite premeasure space, then (using the notation above), $\mathcal{B} = \bar{\mathcal{A}}^d = \bar{\mathcal{A}}_\sigma^d$ and $\mathcal{B} \subset 2^\Omega$ is a $\sigma$ – algebra.*

**Proof.**

1. $\left(\bar{\mathcal{A}}_\sigma^d = \bar{\mathcal{A}}^d\right)$ As $\mathcal{A} \subset \mathcal{A}_\sigma$ it suffices to show $\mathcal{A}_\sigma \subset \bar{\mathcal{A}}^d$. However, if $B \in \mathcal{A}_\sigma$ and $B_n \in \mathcal{A}$ are such that $B_n \uparrow B$, then by Corollary 6.39 and Proposition 6.38 it follows that;
$$d(B_n, B) = \mu^*(B \setminus B_n) = \mu_\sigma(B \setminus B_n) = \mu_\sigma(B) - \mu(B_n) \to 0.$$
2. (**$\mathcal{B}$ is a $\sigma$ – algebra.**) From Theorem 6.42 we know already know that $\mathcal{B}$ is a subalgebra of $2^\Omega$ so it only remains to show $\mathcal{B}$ is closed under countable unions. So let $B := \cup_{n=1}^\infty B_n$ with $\{B_n\}_{n=1}^\infty \subset \mathcal{B}$ and $\varepsilon > 0$ be given. By the definition of $\mathcal{B}$ being the closure of $\mathcal{A}$, we may choose $A_n \in \mathcal{A}$ so that $d(A_n, B_n) \le \varepsilon 2^{-n}$ for all $n$. With $A := \cup_{n=1}^\infty A_n \in \mathcal{A}_\sigma$, it follows that
$$d(A, B) = d\left(\cup_{n=1}^\infty A_n, \cup_{n=1}^\infty B_n\right) \le \sum_{n=1}^\infty d(A_n, B_n) \le \varepsilon. \tag{6.13}$$
As $\varepsilon > 0$, this shows $B \in \bar{\mathcal{A}}_\sigma = \bar{\mathcal{A}}^d = \mathcal{B}$.

■

**Exercise 6.7.** Let $\Omega = \mathbb{R}$, $\mathcal{A} \subset 2^{\mathbb{R}}$ be the algebra generated by half open intervals, $\mu$ be the length measure on $\mathcal{A}$, and set $B = \cup_{n=0}^\infty (2n, 2n+1] \in \mathcal{A}_\sigma$. Show $d(B, A) = \infty$ for all $A \in \mathcal{A}$ and use this to conclude that both assertions in Proposition 6.43 fail in this case.

Recall by Lemma 6.12 if $\mu(\Omega) < \infty$, then $\mu$ is Lip-1 on $\mathcal{A}$ relative to $d$ and hence by Lemma 2.28 there exist a unique Lip-1 extension of $(\bar{\mu})$ of $\mu$ to $\mathcal{B} = \bar{\mathcal{A}}^d$. The next theorem is a restatement of Theorem 6.13.

**Theorem 6.44 (Finite premeasure extension theorem).** *If $(\Omega, \mathcal{A}, \mu)$ is a finite premeasure space, $\mathcal{B} = \bar{\mathcal{A}}^d \subset 2^\Omega$, and $\bar{\mu}$ is the unique $d$ – continuous extension of $\mu$ from $\mathcal{A}$ to $\mathcal{B}$,then $(\Omega, \mathcal{B}, \bar{\mu})$ as $\sigma$ – additive measure space. [Clearly $\mathcal{A} \subset \mathcal{B}$ and hence $\sigma(\mathcal{A}) \subset \mathcal{B}$. It also worth spelling out that $A \in \mathcal{B}$ iff there exists $\{A_n\} \subset \mathcal{A}$ such that $\lim_{n\to\infty} \mu^*(A_n \triangle A) = 0$ and if which case $\bar{\mu}(A) := \lim_{n\to\infty} \mu(A_n)$.]*

**Proof.** From Proposition 6.43) we already know $\mathcal{B}$ is a $\sigma$ – algebra and by Theorem 6.42 we know that $\bar{\mu}$ is a finitely additive measure on $\mathcal{B}$. To see that $\bar{\mu}$ is $\sigma$ – additive, it suffices (see Proposition 5.2) to show $\bar{\mu}$ is countably sub-additive on $\mathcal{B}$ which we now do.

Let $\varepsilon > 0$ be given, $B := \cup_{n=1}^\infty B_n$ with $B_n \in \mathcal{B}$, and $A := \cup_{n=1}^\infty A_n \in \mathcal{A}_\sigma$ with $A_n \in \mathcal{A}$ so that $d(A_n, B_n) \le \varepsilon 2^{-n}$ for all $n$. Using Eq. (6.13), the fact that $\bar{\mu}$ is Lip $-1$, and $\mu_\sigma$ is sub-additivity on $\mathcal{A}_\sigma$, it follows that

$$\begin{aligned}\bar{\mu}(B) &\le \bar{\mu}(A) + \varepsilon = \mu_\sigma(A) + \varepsilon \le \sum_{n=1}^\infty \mu(A_n) + \varepsilon \\ &\le \sum_{n=1}^\infty \left[\bar{\mu}(B_n) + \varepsilon 2^{-n}\right] + \varepsilon = \sum_{n=1}^\infty \bar{\mu}(B_n) + 2\varepsilon.\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, we have proved the desired sub-additivity of $\bar{\mu}$ on $\mathcal{B}$. ■

**Corollary 6.45.** *If $(\Omega, \mathcal{A}, \mathcal{B}, \mu, \bar{\mu})$ be as in Theorem 6.44, then $B \subset \Omega$ is in $\mathcal{B}$ iff for every $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu_\sigma(C \setminus A) < \varepsilon$. Moreover, we may compute $\bar{\mu}$ using*

$$\bar{\mu}(B) = \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\}. \tag{6.14}$$

**Proof.** Let $B \in \mathcal{B}$ and $\varepsilon > 0$. As $\mathcal{B} = \bar{\mathcal{A}}^d$, there exists $D \in \mathcal{A}$ so that $d(B, D) = \mu^*(D \triangle B) < \varepsilon/4$ and so by the definition of $\mu^*$ there exists $F \in \mathcal{A}_\sigma$ so that $D \triangle B \subset F$ and $\mu_\sigma(F) < \varepsilon/4$. Letting $C := D \cup F \in \mathcal{A}_\sigma$, we have $B \subset D \cup [B \setminus D] \subset C$ and

$$d(B, C) = d(B \cup \emptyset, D \cup F) \le d(B, D) + d(\emptyset, F) < \varepsilon/4 + \varepsilon/4 = \varepsilon/2.$$

Applying this same argument to $B^c \in \mathcal{B}$, we may find $\tilde{C} \in \mathcal{A}_\sigma$ so that $B^c \subset \tilde{C}$ and $d\left(B^c, \tilde{C}\right) < \varepsilon/2$. Letting $A := \tilde{C}^c \in \mathcal{A}_\delta$ it the follows that $A \subset B \subset C$ and

$$\begin{aligned}\mu_\sigma(C \setminus A) &= \mu^*(C \setminus A) = d(A, C) \\ &\le d(A, B) + d(B, C) < \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

It now also follows that

$$\bar{\mu}(C) - \bar{\mu}(B) = \bar{\mu}(C \setminus B) \le \bar{\mu}(C \setminus A) < \varepsilon$$

which implies,

$$\bar{\mu}(B) \le \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\} < \bar{\mu}(B) + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this proves Eq. (6.14).

Conversely if $B \subset 2^\Omega$ and there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu_\sigma(C \setminus A) < \varepsilon$, then

$$d(B, C) = \mu^*(C \setminus B) \le \mu^*(C \setminus A) = \mu_\sigma(C \setminus A) < \varepsilon.$$

As $\varepsilon > 0$ is assumed to be arbitrary, it follows that $B \in \bar{\mathcal{A}}_\sigma = \mathcal{B}$. ■

**Exercise 6.8.** Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu$ and $\nu$ are two finite measures on $\mathcal{B} = \sigma(\mathcal{A})$ such that $\mu = \nu$ on $\mathcal{A}$. Show $\mu = \nu$ on $\mathcal{B}$.

**Exercise 6.9 (Look at but do not hand in).** Let $\mu$, $\bar{\mu}$, $\mathcal{A}$, and $\mathcal{B} := \mathcal{B}(\mu)$ be as in Theorem 6.44. Further suppose that $\mathcal{B}_0 \subset 2^\Omega$ is a $\sigma$ – algebra such that $\mathcal{A} \subset \mathcal{B}_0 \subset \mathcal{B}$ and $\nu : \mathcal{B}_0 \to [0, \mu(\Omega)]$ is a $\sigma$ – additive measure on $\mathcal{B}_0$ such that $\nu = \mu$ on $\mathcal{A}$. Show that $\nu = \bar{\mu}$ on $\mathcal{B}_0$ as well. (When $\mathcal{B}_0 = \sigma(\mathcal{A})$ this exercise is of course a consequence of Proposition 6.18. It is not necessary to use this information to complete the exercise.)

### 6.5.1 *Other characterizations of $\mathcal{B}$

This subsection may be skipped on first reading. Its purpose is to serve as motivation for Carathéodory's general construction theorem appearing in the optional Chapter 7. Lemma 6.46 and Notation 6.47 are redundant here.

**Lemma 6.46.** *If $\mathcal{C}$ is a collection of subsets of $\Omega$ such that $\emptyset, \Omega \in \mathcal{C}$ and $\mathcal{C}$ is closed under countable unions and $\mu : \mathcal{C} \to [0, \infty]$ is a monotone and countably subadditive on $\mathcal{C}$ then $\mu^* : 2^\Omega \to [0, \infty]$ defined by,*

$$\mu^*(B) := \inf\{\mu(C) : B \subset C \in \mathcal{C}\},$$

*is an outer measure on $\Omega$.*

**Proof.** It is clear that $\mu^*(\emptyset) = 0$ and that $\mu^*$ is monotone. If $B_n \subset \Omega$ and $C_n \in \mathcal{C}$ so that $B_n \subset C_n$, then we have $\cup_{n=1}^\infty B_n \subset \cup_{n=1}^\infty C_n \in \mathcal{C}$ and hence

$$\mu^*(\cup_{n=1}^\infty B_n) \le \mu(\cup_{n=1}^\infty C_n) \le \sum_{n=1}^\infty \mu(C_n).$$

By taking the infimum over all such $C_n$ it follows that

$$\mu^*(\cup_{n=1}^\infty B_n) \le \inf \sum_{n=1}^\infty \mu(C_n) = \sum_{n=1}^\infty \mu^*(B_n).$$

To see this is the case, if $\sum_{n=1}^{\infty} \mu^*(B_n) = \infty$ there is nothing to prove. If $\sum_{n=1}^{\infty} \mu^*(B_n) < \infty$ and $\varepsilon > 0$, we may choose $C_n \in \mathcal{C}$ so that $B_n \subset C_n$ and $\mu(C_n) \le \mu^*(B_n) + \varepsilon 2^{-n}$ and hence

$$\mu^*(\cup_{n=1}^{\infty} B_n) \le \sum_{n=1}^{\infty} \mu(C_n) \le \sum_{n=1}^{\infty} \mu^*(B_n) + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary we again conclude that $\mu^*(\cup_{n=1}^{\infty} B_n) \le \sum_{n=1}^{\infty} \mu^*(B_n)$. ■

Combining Propositions 6.38 and Lemma 6.46 gives us our most important example of an outer measure which we now introduce.

**Notation 6.47 (Outer measures)** *Let $\mu : \mathcal{A} \to [0,\infty)$ be a finite premeasure, $\mu_\sigma : \mathcal{A}_\sigma \to [0,\infty)$ be its extension to $\mathcal{A}_\sigma$ as in Propositions 6.38 and now (by abuse of notation) let $\mu^* := \mu_\sigma^*$ (as in Lemma 6.46), i.e.*

$$\mu^*(B) := \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\} \ \forall\ B \in 2^\Omega.$$

*We refer to $\mu^*(B)$ as the $\mu$ -* ***outer measure (or $\mu$ - outer content) of*** *$B$.*

**Proposition 6.48 (*Other characterizations of $\mathcal{B}$).** *Let*

$$\mathcal{B}' := \left\{B \subset \Omega : \mu^*(E \cap B) + \mu^*(E \setminus B) = \mu^*(E)\ \forall\ E \in 2^\Omega\right\}$$

*and*

$$\mathcal{B}'' := \{B \subset \Omega : \mu^*(B) + \mu^*(B^c) = \mu(\Omega)\}$$

*Then we have $\mathcal{B} = \mathcal{B}' = \mathcal{B}''$ and moreover $\bar{\mu} = \mu^*|_{\mathcal{B}}$.*

**Proof.** Let us first observe that $\mathcal{B}'$ may be expressed alternatively as;

$$\mathcal{B}' := \left\{B \subset \Omega : \mu^*(E \cap B) + \mu^*(E \setminus B) \le \mu^*(E)\ \forall\ E \in 2^\Omega\right\}.$$

This is because the subadditivty of $\mu^*$ automatically implies

$$\mu^*(E) \le \mu^*(E \cap B) + \mu^*(E \setminus B) \forall\ B, E \in 2^\Omega.$$

As the test for as set to be in $\mathcal{B}''$ is the same as one of the tests (namely $E = \Omega$) for being in $\mathcal{B}'$, we have $\mathcal{B}' \subset \mathcal{B}''$. We will now complete the proof that $\mathcal{B} = \mathcal{B}' = \mathcal{B}''$ by showing $\mathcal{B} \subset \mathcal{B}'$ and $\mathcal{B}'' \subset \mathcal{B}$.

$(\mathcal{B} \subset \mathcal{B}')$. If $B \in \mathcal{B}$, $E \in 2^\Omega$, and $C \in \mathcal{A}_\sigma$ such that $E \subset C$, then

$$\begin{aligned}\mu^*(E \cap B) + \mu^*(E \setminus B) &\le \mu^*(C \cap B) + \mu^*(C \setminus B) \\ &= \bar{\mu}(C \cap B) + \bar{\mu}(C \setminus B) = \bar{\mu}(C) = \mu_\sigma(C).\end{aligned}$$

Taking the infimum over all $C \in \mathcal{A}_\sigma$ such that $E \subset C$ shows

$$\mu^*(E \cap B) + \mu^*(E \setminus B) \le \mu^*(E)$$

and hence $B \in \mathcal{B}'$.

$(\mathcal{B}'' \subset \mathcal{B})$. If $B \in \mathcal{B}''$ and $\varepsilon > 0$ is given, there exists $C, D \in \mathcal{A}_\sigma$ so that $B \subset C$, $B^c \subset D$,

$$\bar{\mu}(C) = \mu_\sigma(C) \le \mu^*(B) + \varepsilon, \text{ and } \bar{\mu}(D) = \mu_\sigma(D) \le \mu^*(B^c) + \varepsilon.$$

Summing these inequalities while using $B \in \mathcal{B}''$ implies,

$$\bar{\mu}(C) + \bar{\mu}(D) \le \mu(\Omega) + 2\varepsilon \implies \bar{\mu}(C) \le \bar{\mu}(D^c) + 2\varepsilon.$$

As $D^c \subset B \subset C$ and $C \setminus D^c = C \cap D \in \mathcal{A}_\sigma$, it follows that

$$\mu^*(C \setminus D^c) = \mu_\sigma(C \setminus D^c) = \bar{\mu}(C \setminus D^c) \le 2\varepsilon.$$

Consequently we conclude that,

$$d(B, C) = \mu^*(C \setminus B) \le \mu^*(C \setminus D^c) \le 2\varepsilon \text{ and} \tag{6.15}$$
$$d(B, D^c) = \mu^*(B \setminus D^c) \le \mu^*(C \setminus D^c) \le 2\varepsilon. \tag{6.16}$$

Since $\varepsilon > 0$ is arbitrary these equations show $B \in \bar{\mathcal{A}}_\sigma^d = \mathcal{B}$ and also that $B \in \bar{\mathcal{A}}_\delta^d$. Moreover, since $\bar{\mu} = \mu_\sigma$ on $\mathcal{A}_\sigma$ and $\bar{\mu}$ is Lip $-1$ relative to $d$, we also find,

$$\begin{aligned}|\bar{\mu}(B) - \mu^*(B)| &\le |\bar{\mu}(B) - \bar{\mu}(C)| + |\mu_\sigma(C) - \mu^*(B)| \\ &\le 2\varepsilon + \varepsilon = 3\varepsilon.\end{aligned}$$

Again as $\varepsilon > 0$ is arbitrary we have also shown that $\bar{\mu} = \mu^*$ on $\mathcal{B}$. ■

**Notation 6.49 (Inner measures)** *If $\mu : \mathcal{A} \to [0,\infty)$ is a finite premeasure and $B \subset \Omega$ let*

$$\mu_*(B) := \sup\left\{\mu^\delta(A) : \mathcal{A}_\delta \ni A \subset B\right\},$$

*where $\mu^\delta := \bar{\mu}|_{\mathcal{A}_\delta}$. We refer to $\mu_*(B)$ as the* ***inner measure (or inner content)*** *of $B$.*

*Remark 6.50.* If $A \in \mathcal{A}_\delta$ and $A_n \in \mathcal{A}$ are such that $A_n \downarrow A$, then $\mu^\delta(A) = \lim_{n\to\infty} \mu(A_n)$. This shows that $\mu^\delta$ may be computed directly from $\mu$ without referring to the extension $\bar{\mu}$ as in Notation 6.49.

**Corollary 6.51.** *Let $(\Omega, \mathcal{A}, \mu)$ be a finite premeasure space and $\mu_*$ and $\mu^*$ be the inner and outer measures associated to $\mu$ as in Notations 6.49 and Proposition 6.29 and Proposition 6.36 respectively. Then (continuing to use the notation in Theorem 6.44)*

$$\mathcal{B} := \bar{\mathcal{A}}^d = \{B \subset \Omega : \mu_*(B) = \mu^*(B)\}. \tag{6.17}$$

**Proof.** From Proposition 6.48 we know that $B \in \mathcal{B}$ iff $\mu^*(B) + \mu^*(B^c) = \mu(\Omega)$. Now choose $C_n \in \mathcal{A}_\sigma$ such that $B^c \subset C_n$ and $\bar{\mu}(C_n) \downarrow \mu^*(B^c)$ and let $A_n := C_n^c \in \mathcal{A}_\delta$ with $A_n \subset B$, then

$$\mu^*(B) = \mu(\Omega) - \mu^*(B^c) = \lim_{n\to\infty} [\mu(\Omega) - \bar{\mu}(C_n)]$$
$$= \lim_{n\to\infty} \bar{\mu}(A_n) \le \mu_*(B).$$

As $\mu_*(B) \le \mu^*(B)$ holds for all $B \in 2^\Omega$ it follows that $B \in \mathcal{B}$ implies $\mu^*(B) = \mu_*(B)$.

Conversely if $B \in 2^\Omega$ satisfies $\mu^*(B) = \mu_*(B)$, there exists $\mathcal{A}_\delta \ni A_n \subset B \subset C_n \in \mathcal{A}_\sigma$ such that $\lim_{n\to\infty} \bar{\mu}(C_n) = \lim_{n\to\infty} \bar{\mu}(A_n)$ which implies

$$\lim_{n\to\infty} \mu^*(C_n \setminus A_n) = \lim_{n\to\infty} \bar{\mu}(C_n \setminus A_n) = 0$$

and hence $d(B, C_n) = \mu^*(C_n \setminus B) \le \mu^*(C_n \setminus A_n) \to 0$ as $n \to \infty$ which shows $B \in \bar{\mathcal{A}}_\sigma^d = \mathcal{B}$.

■

## 6.6 Construction of $\sigma$ - Finite Measures

The goal of this section is to generalize Theorem 6.44 and Corollary 6.45 to $\sigma$ – finite measures.

**Theorem 6.52 ($\sigma$ - Finite Premeasure Extension Theorem).** *If $\mu$ is a $\sigma$ – finite premeasure on an algebra $\mathcal{A}$, then there exists a unique measure, $\bar{\mu}$, on $\sigma(\mathcal{A})$ such that $\bar{\mu} = \mu$ on $\mathcal{A}$.*

**Proof. Existence of an extension.** Let $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{A}$ be a partition of $\Omega$ such that $\mu(\Omega_n) < \infty$ for all $n$. To each $n \in \mathbb{N}$ let $\mu_n$ be the finite premeasure on $\mathcal{A}$ defined by

$$\mu_n(A) := \mu_n(A \cap \Omega_n) \text{ for all } A \in \mathcal{A}$$

and let $\bar{\mu}_n$ be an extension of $\mu_n$ to $\sigma(\mathcal{A})$ guaranteed by Theorem 6.44. Then the measure, $\bar{\mu} := \sum_{n=1}^\infty \bar{\mu}_n$, is then an extension of $\mu$ to $\sigma(\mathcal{A})$.

**Uniqueness of the extension.** Suppose that $\nu$ is another measure on $\sigma(\mathcal{A})$ such that $\nu = \mu$ on $\mathcal{A}$. Given $n \in \mathbb{N}$, let $\nu_n$ be the measure on $\sigma(\mathcal{A})$ defined by $\nu_n(B) := \nu(B \cap \Omega_n)$ for all $B \in \sigma(\mathcal{A})$. As $\nu_n = \mu_n = \bar{\mu}_n$ on $\mathcal{A}$, it follows by Exercise 6.8 that $\nu_n = \bar{\mu}_n$ on $\sigma(\mathcal{A})$ and therefore

$$\nu = \sum_{n=1}^\infty \nu_n = \sum_{n=1}^\infty \bar{\mu}_n = \mu.$$

■

**Corollary 6.53.** *Suppose that $\mu$ is a $\sigma$ – finite premeasure on an algebra $\mathcal{A}$. If $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$ is given, there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu_\sigma(C \setminus A) < \varepsilon$. Moreover if $\bar{\mu}(B) < \infty$ and $\varepsilon > 0$ is given, there exists $A \in \mathcal{A}$ so that $\bar{\mu}(B \triangle A) < \varepsilon$.*

**Proof.** Let $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{A}$ be a partition of $\Omega$ such that $\mu(\Omega_n) < \infty$ for all $n$ and let $\bar{\mu}_n$ be the measure on $\sigma(\mathcal{A})$ defined by $\bar{\mu}_n(B) := \bar{\mu}(B \cap \Omega_n)$ for all $B \in \sigma(\mathcal{A})$. According to Corollary 6.45, to each $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$, there exists $A_n \in \mathcal{A}_\delta$ and $C_n \in \mathcal{A}_\sigma$ such that $A_n \subset B \subset C_n$ and

$$\bar{\mu}(C_n \cap \Omega_n \setminus [A_n \cap \Omega_n]) = \bar{\mu}_n(C_n \setminus A_n) < \varepsilon 2^{-n} \ \forall\ n \in \mathbb{N}.$$

Taking $C := \cup_n [C_n \cap \Omega_n] \in \mathcal{A}_\sigma$ it follows that

$$\bar{\mu}(C \setminus B) = \sum_{n=1}^\infty \bar{\mu}(C_n \cap \Omega_n \setminus [B \cap \Omega_n]) \le \sum_{n=1}^\infty \bar{\mu}(C_n \cap \Omega_n \setminus [A_n \cap \Omega_n]) < \varepsilon.$$

Applying this same result to $B^c \in \sigma(\mathcal{A})$, there exists $A \in \mathcal{A}_\delta$ so that $B^c \subset A^c \in \mathcal{A}_\sigma$ and

$$\bar{\mu}(B \setminus A) = \bar{\mu}(A^c \setminus B^c) < \varepsilon.$$

From this we conclude that $\mu_\sigma(C \setminus A) = \bar{\mu}(C \setminus A) < 2\varepsilon$ which suffices as $\varepsilon > 0$ was arbitrary.

Let us now further assume that $\bar{\mu}(B) < \infty$ and $\varepsilon > 0$ be given. By what we have just proved we may find $A_n \in \mathcal{A}$ so that $A_n \uparrow C \in \mathcal{A}_\sigma$ with $B \subset C$ and $\bar{\mu}(C \setminus B) < \varepsilon$. Hence it follows that,

$$\bar{\mu}(A_n \triangle B) = \bar{\mu}(A_n \setminus B) + \bar{\mu}(B \setminus A_n)$$
$$\le \bar{\mu}(C \setminus B) + \bar{\mu}(B \setminus A_n) < \varepsilon + \bar{\mu}(B \setminus A_n).$$

As $\bar{\mu}(B) < \infty$ and $B \setminus A_n \downarrow \emptyset$, we know that $\bar{\mu}(B \setminus A_n) \downarrow 0$ and therefore for $n$ sufficiently large we see that $\bar{\mu}(A_n \triangle B) < \varepsilon$. ■

Recall that

$$\mu_\sigma(C) = \sup\{\mu(A) : \mathcal{A} \ni A \subset C\} = \bar{\mu}(A) \text{ for all } C \in \mathcal{A}_\sigma.$$

**Corollary 6.54.** *Suppose that $\mu$ is a $\sigma$ – finite premeasure on an algebra $\mathcal{A}$ and $\bar{\mu}$ is its unique extension to a measure on $\sigma(\mathcal{A})$. Then $\bar{\mu}$ may be computed using,*

$$\bar{\mu}(B) = \mu^*(B) := \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\} \ \forall\ B \in \sigma(\mathcal{A}). \tag{6.18}$$

**Proof.** For $B \in \sigma(\mathcal{A})$, let

$$\nu(B) = \inf\{\mu_\sigma(C) : B \subset C \in \mathcal{A}_\sigma\}$$
$$= \inf\{\bar{\mu}(C) : B \subset C \in \mathcal{A}_\sigma\}.$$

Clearly we have $\nu(B) \le \bar{\mu}(B)$. Moreover if $\varepsilon > 0$ is given, by Corollary 6.53 there exists $C \in \mathcal{A}_\sigma$ such that $B \subset C$ and $\bar{\mu}(C \setminus B) < \varepsilon$. hence it follows that

$$\nu(B) \le \mu_\sigma(C) = \bar{\mu}(C) \le \bar{\mu}(B) + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, we have shown $\nu(B) \le \bar{\mu}(B)$ for all $B \in \sigma(\mathcal{A})$. ■

The following slight reformulation of Theorem 6.52 can be useful.

**Corollary 6.55 (*).** *Let $\mathcal{A}$ be an algebra of sets, $\{\Omega_m\}_{m=1}^\infty \subset \mathcal{A}$ is a given sequence of sets such that $\Omega_m \uparrow \Omega$ as $m \to \infty$. Let*

$$\mathcal{A}_f := \{A \in \mathcal{A} : A \subset \Omega_m \text{ for some } m \in \mathbb{N}\}.$$

*Notice that $\mathcal{A}_f$ is a ring, i.e. closed under differences, intersections and unions and contains the empty set. Further suppose that $\mu : \mathcal{A}_f \to [0,\infty)$ is an additive set function such that $\mu(A_n) \downarrow 0$ for any sequence, $\{A_n\} \subset \mathcal{A}_f$ such that $A_n \downarrow \emptyset$ as $n \to \infty$. Then $\mu$ extends uniquely to a $\sigma$ – finite measure on $\mathcal{A}$.*

**Proof. Existence.** By assumption, $\mu_m := \mu|_{\mathcal{A}_{\Omega_m}} : \mathcal{A}_{\Omega_m} \to [0,\infty)$ is a premeasure on $(\Omega_m, \mathcal{A}_{\Omega_m})$ and hence by Theorem 6.52 extends to a measure $\mu_m'$ on $(\Omega_m, \sigma(\mathcal{A}_{\Omega_m}) = \mathcal{B}_{\Omega_m})$. Let $\bar{\mu}_m(B) := \mu_m'(B \cap \Omega_m)$ for all $B \in \mathcal{B}$. Then $\{\bar{\mu}_m\}_{m=1}^\infty$ is an increasing sequence of measure on $(\Omega, \mathcal{B})$ and hence $\bar{\mu} := \lim_{m\to\infty} \bar{\mu}_m$ defines a measure on $(\Omega, \mathcal{B})$ such that $\bar{\mu}|_{\mathcal{A}_f} = \mu$.

Uniqueness. If $\mu_1$ and $\mu_2$ are two such extensions, then $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{A}$ and therefore by Proposition 6.18 or Exercise 6.3 we know that $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{B}$. We may now let $m \to \infty$ to see that in fact $\mu_1(B) = \mu_2(B)$ for all $B \in \mathcal{B}$, i.e. $\mu_1 = \mu_2$. ■

## 6.7 Radon Measures on $\mathbb{R}$

We say that a measure, $\mu$, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is a **Radon measure** if $\mu([a,b]) < \infty$ for all $-\infty < a < b < \infty$. In this section we will give a characterization of all Radon measures on $\mathbb{R}$. We first need the following general result characterizing premeasures on an algebra generated by a semi-algebra.

**Proposition 6.56.** *Suppose that $\mathcal{S} \subset 2^\Omega$ is a semi-algebra, $\mathcal{A} = \mathcal{A}(\mathcal{S})$ and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Then $\mu$ is a premeasure on $\mathcal{A}$ iff $\mu$ is countably sub-additive on $\mathcal{S}$.*

**Proof.** Clearly if $\mu$ is a premeasure on $\mathcal{A}$ then $\mu$ is $\sigma$ - additive and hence sub-additive on $\mathcal{S}$. Because of Proposition 5.2, to prove the converse it suffices to show that the sub-additivity of $\mu$ on $\mathcal{S}$ implies the sub-additivity of $\mu$ on $\mathcal{A}$.

So suppose $A = \sum_{n=1}^\infty A_n \in \mathcal{A}$ with each $A_n \in \mathcal{A}$. By Proposition 4.27 we may write $A = \sum_{j=1}^k E_j$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_j, E_{n,i} \in \mathcal{S}$. Intersecting the identity, $A = \sum_{n=1}^\infty A_n$, with $E_j$ implies

$$E_j = A \cap E_j = \sum_{n=1}^\infty A_n \cap E_j = \sum_{n=1}^\infty \sum_{i=1}^{N_n} E_{n,i} \cap E_j.$$

By the assumed sub-additivity of $\mu$ on $\mathcal{S}$,

$$\mu(E_j) \le \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j).$$

Summing this equation on $j$ and using the finite additivity of $\mu$ shows

$$\mu(A) = \sum_{j=1}^k \mu(E_j) \le \sum_{j=1}^k \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j)$$
$$= \sum_{n=1}^\infty \sum_{i=1}^{N_n} \sum_{j=1}^k \mu(E_{n,i} \cap E_j) = \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i}) = \sum_{n=1}^\infty \mu(A_n).$$

■

If $\mu$ is a Radon measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, then we can always find a function, $F : \mathbb{R} \to \mathbb{R}$, such that

$$\mu((a,b]) = F(b) - F(a) \text{ for all } -\infty < a \le b < \infty. \tag{6.19}$$

For example if $\mu(\mathbb{R}) < \infty$ we can take $F(x) = \mu((-\infty, x])$ while if $\mu(\mathbb{R}) = \infty$ we might take

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x \ge 0 \\ -\mu((x,0]) & \text{if } x \le 0 \end{cases}.$$

It is a simple exercise to show that Eq. (6.19) uniquely determines $F$ modulo an additive constant.

**Lemma 6.57.** *If $\mu$ is a Radon measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that $\mu((a,b]) = F(b) - F(a)$, then $F$ is increasing and right continuous.*

**Proof.** The function $F$ is increasing by the monotonicity of $\mu$. To see that $F$ is right continuous, let $b \in \mathbb{R}$ and choose $a \in (-\infty, b)$ and any sequence $\{b_n\}_{n=1}^\infty \subset (b, \infty)$ such that $b_n \downarrow b$ as $n \to \infty$. Since $\mu((a, b_1]) < \infty$ and $(a, b_n] \downarrow (a, b]$ as $n \to \infty$, it follows that

$$F(b_n) - F(a) = \mu((a, b_n]) \downarrow \mu((a,b]) = F(b) - F(a).$$

Since $\{b_n\}_{n=1}^\infty$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $\lim_{y \downarrow b} F(y) = F(b)$. ■

The key result of this section is the converse to this lemma.

**Theorem 6.58.** *Suppose $F:\mathbb{R}\to\mathbb{R}$ is a right continuous increasing function. Then there exists a unique Radon measure, $\mu=\mu_F$, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that Eq. (6.19) holds.*

**Proof. First Proof.** Let $\mathcal{S}:=\{(a,b]\cap\mathbb{R}:-\infty\le a\le b\le\infty\}$, and $\mathcal{A}=\mathcal{A}(\mathcal{S})$ consists of those sets, $A\subset\mathbb{R}$ which may be written as finite disjoint unions of sets from $\mathcal{S}$ as in Example 4.28. Recall that $\mathcal{B}_{\mathbb{R}}=\sigma(\mathcal{A})=\sigma(\mathcal{S})$. Further define $F(\pm\infty):=\lim_{x\to\pm\infty}F(x)$ and let $\mu=\mu_F$ be the finitely additive measure on $(\mathbb{R},\mathcal{A})$ described in Proposition 5.8 and Remark 5.9. To finish the proof it suffices by Theorem 6.52 to show that $\mu$ is a premeasure on $\mathcal{A}=\mathcal{A}(\mathcal{S})$ where $\mathcal{S}:=\{(a,b]\cap\mathbb{R}:-\infty\le a\le b\le\infty\}$. So in light of Proposition 6.56, to finish the proof it suffices to show $\mu$ is sub-additive on $\mathcal{S}$, i.e. we must show

$$\mu(J)\le\sum_{n=1}^{\infty}\mu(J_n). \tag{6.20}$$

where $J=\sum_{n=1}^{\infty}J_n$ with $J=(a,b]\cap\mathbb{R}$ and $J_n=(a_n,b_n]\cap\mathbb{R}$. Recall from Proposition 5.2 that the finite additivity of $\mu$ implies

$$\sum_{n=1}^{\infty}\mu(J_n)\le\mu(J). \tag{6.21}$$

We begin with the special case where $-\infty<a<b<\infty$. Our proof will be by "continuous induction." The strategy is to show $a\in\Lambda$ where

$$\Lambda:=\left\{\alpha\in[a,b]:\mu(J\cap(\alpha,b])\le\sum_{n=1}^{\infty}\mu(J_n\cap(\alpha,b])\right\}. \tag{6.22}$$

As $b\in J$, there exists an $k$ such that $b\in J_k$ and hence $(a_k,b_k]=(a_k,b]$ for this $k$. It now easily follows that $J_k\subset\Lambda$ so that $\Lambda$ is not empty. To finish the proof we are going to show $\bar a:=\inf\Lambda\in\Lambda$ and that $\bar a=a$.

- Let $\alpha_m\in\Lambda$ such that $\alpha_m\downarrow\bar a$, i.e.

$$\mu(J\cap(\alpha_m,b])\le\sum_{n=1}^{\infty}\mu(J_n\cap(\alpha_m,b]). \tag{6.23}$$

  The right continuity of $F$ implies $\alpha\to\mu(J_n\cap(\alpha,b])$ is right continuous. So by the dominated convergence theorem[2] for sums,

$$\begin{aligned}\mu(J\cap(\bar a,b])&=\lim_{m\to\infty}\mu(J\cap(\alpha_m,b])\\&\le\lim_{m\to\infty}\sum_{n=1}^{\infty}\mu(J_n\cap(\alpha_m,b])\\&=\sum_{n=1}^{\infty}\lim_{m\to\infty}\mu(J_n\cap(\alpha_m,b])=\sum_{n=1}^{\infty}\mu(J_n\cap(\bar a,b]),\end{aligned}$$

  i.e. $\bar a\in\Lambda$.
- If $\bar a>a$, then $\bar a\in J_l=(a_l,b_l]$ for some $l$. Letting $\alpha=a_l<\bar a$, we have,

$$\begin{aligned}\mu(J\cap(\alpha,b])&=\mu(J\cap(\alpha,\bar a])+\mu(J\cap(\bar a,b])\\&\le\mu(J_l\cap(\alpha,\bar a])+\sum_{n=1}^{\infty}\mu(J_n\cap(\bar a,b])\\&=\mu(J_l\cap(\alpha,\bar a])+\mu(J_l\cap(\bar a,b])+\sum_{n\neq l}\mu(J_n\cap(\bar a,b])\\&=\mu(J_l\cap(\alpha,b])+\sum_{n\neq l}\mu(J_n\cap(\bar a,b])\\&\le\sum_{n=1}^{\infty}\mu(J_n\cap(\alpha,b]).\end{aligned}$$

  This shows $\alpha\in\Lambda$ and $\alpha<\bar a$ which violates the definition of $\bar a$. Thus we must conclude that $\bar a=a$.

The hard work is now done but we still have to check the cases where $a=-\infty$ or $b=\infty$. For example, suppose that $b=\infty$ so that

$$J=(a,\infty)=\sum_{n=1}^{\infty}J_n$$

with $J_n=(a_n,b_n]\cap\mathbb{R}$. Then

$$I_M:=(a,M]=J\cap I_M=\sum_{n=1}^{\infty}J_n\cap I_M$$

and so by what we have already proved,

$$F(M)-F(a)=\mu(I_M)\le\sum_{n=1}^{\infty}\mu(J_n\cap I_M)\le\sum_{n=1}^{\infty}\mu(J_n).$$

Now let $M\to\infty$ in this last inequality to find that

[2] DCT applies as $\mu(J_n\cap(\alpha_m,b])\le\mu(J_n)$ and $\sum_{n=1}^{\infty}\mu(J_n)\le\mu(J)<\infty$ by Eq. (6.23).

$$\mu((a,\infty)) = F(\infty) - F(a) \leq \sum_{n=1}^{\infty} \mu(J_n).$$

The other cases where $a = -\infty$ and $b \in \mathbb{R}$ and $a = -\infty$ and $b = \infty$ are handled similarly. ■

**Proof. Second proof.** Case 1. First suppose that $-\infty < a < b < \infty$. Choose numbers $\tilde{a} > a$, $\tilde{b}_n > b_n$ in which case $I := (\tilde{a}, b] \subset J$,

$$\tilde{J}_n := (a_n, \tilde{b}_n] \supset \tilde{J}_n^o := (a_n, \tilde{b}_n) \supset J_n.$$

Since $\bar{I} = [\tilde{a}, b]$ is compact and $\bar{I} \subset J \subset \bigcup_{n=1}^{\infty} \tilde{J}_n^o$ there exists[3] $N < \infty$ such that

$$I \subset \bar{I} \subset \bigcup_{n=1}^{N} \tilde{J}_n^o \subset \bigcup_{n=1}^{N} \tilde{J}_n.$$

Hence by **finite** sub-additivity of $\mu$,

$$F(b) - F(\tilde{a}) = \mu(I) \leq \sum_{n=1}^{N} \mu(\tilde{J}_n) \leq \sum_{n=1}^{\infty} \mu(\tilde{J}_n).$$

Using the right continuity of $F$ and letting $\tilde{a} \downarrow a$ in the above inequality,

$$\begin{aligned} \mu(J) = \mu((a,b]) = F(b) - F(a) &\leq \sum_{n=1}^{\infty} \mu\left(\tilde{J}_n\right) \\ &= \sum_{n=1}^{\infty} \mu(J_n) + \sum_{n=1}^{\infty} \mu(\tilde{J}_n \setminus J_n). \end{aligned} \tag{6.24}$$

Given $\varepsilon > 0$, we may use the right continuity of $F$ to choose $\tilde{b}_n$ so that

$$\mu(\tilde{J}_n \setminus J_n) = F(\tilde{b}_n) - F(b_n) \leq \varepsilon 2^{-n} \ \forall\, n \in \mathbb{N}.$$

Using this in Eq. (6.24) shows

$$\mu(J) = \mu((a,b]) \leq \sum_{n=1}^{\infty} \mu(J_n) + \varepsilon$$

which verifies Eq. (6.20) since $\varepsilon > 0$ was arbitrary. ■

[3] To see this, let $c := \sup\left\{x \leq b : [\tilde{a}, x] \text{ is finitely covered by } \left\{\tilde{J}_n^o\right\}_{n=1}^{\infty}\right\}$. If $c < b$, then $c \in \tilde{J}_m^o$ for some $m$ and there exists $x \in \tilde{J}_m^o$ such that $[\tilde{a}, x]$ is finitely covered by $\left\{\tilde{J}_n^o\right\}_{n=1}^{\infty}$, say by $\left\{\tilde{J}_n^o\right\}_{n=1}^{N}$. We would then have that $\left\{\tilde{J}_n^o\right\}_{n=1}^{\max(m,N)}$ finitely covers $[a, c']$ for all $c' \in \tilde{J}_m^o$. But this contradicts the definition of $c$.

**Corollary 6.59.** *Suppose $-\infty < \alpha < \beta < \infty$ and $F : [\alpha, \beta] \to \mathbb{R}$ is a right continuous increasing function. Then there exists a unique measure, $\mu = \mu_F$, on $\left([\alpha, \beta], \mathcal{B}_{[\alpha,\beta]}\right)$ such that*

$$\mu([\alpha, b]) = F(b) \text{ for all } \alpha \leq a \leq b \leq \beta.$$

**Proof.** Extend $F$ to $\mathbb{R}$ by setting $F(x) = F(\alpha)$ for $x \geq \alpha$ and $F(x) = F(\beta)$ for $x \leq \beta$. The extension now satisfies the hypothesis of Theorem 6.58 and hence there exists a measure $\hat{\mu}$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that Eq. (6.19) holds. The measure $\hat{\mu}$ satisfies,

$$\begin{aligned} \hat{\mu}([\alpha, b]) = \hat{\mu}([\alpha, b]) &= \lim_{n\to\infty} \hat{\mu}\left((\alpha - \frac{1}{n}, b]\right) \\ &= \lim_{n\to\infty}\left[F(b) - F\left(\alpha - \frac{1}{n}\right)\right] = F(b) - F(\alpha). \end{aligned}$$

Therefore the desired measure is given by $\mu(A) := F(\alpha)\delta_\alpha(A) + \hat{\mu}(A)$ for all $A \in \mathcal{B}_{[\alpha,\beta]}$. ■

## 6.8 Metric-Measure Space Regularity Exercises

See Section **??** of the analysis notes for more general regularity results along these lines. In particular we have.

**Exercise 6.10.** If $(X, \rho)$ is a metric space and $\mu$ is a finite measure on $(X, \mathcal{B}_X)$, then for all $A \in \mathcal{B}_X$ and $\varepsilon > 0$ there exists a closed set $F$ and open set $V$ such that $F \subset A \subset V$ and $d_\mu(F, V) = \mu(V \setminus F) < \varepsilon$.

**Exercise 6.11.** If $(X, \rho)$ is a metric space and $\mu$ is a measure on $(X, \mathcal{B}_X)$ such that there exists open sets, $\{V_n\}_{n=1}^{\infty}$, of $X$ such that $V_n \uparrow X$ and $\mu(V_n) < \infty$ for all $n$. The for all $A \in \mathcal{B}_X$ and $\varepsilon > 0$ there exists a closed set $F$ and open set $V$ such that $F \subset A \subset V$ and $d_\mu(F, V) = \mu(V \setminus F) < \varepsilon$.

**Exercise 6.12.** Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma(\{\text{open subsets of } \mathbb{R}^n\})$ be the Borel $\sigma-$ algebra on $\mathbb{R}^n$ and $\mu$ be a probability measure on $\mathcal{B}$. Further, let $\mathcal{B}_0$ denote those sets $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there exists $F \subset B \subset V$ such that $F$ is closed, $V$ is open, and $\mu(V \setminus F) < \varepsilon$. Show:

1. $\mathcal{B}_0$ contains all closed subsets of $\mathcal{B}$. **Hint:** given a closed subset, $F \subset \mathbb{R}^n$ and $k \in \mathbb{N}$, let $V_k := \cup_{x \in F} B(x, 1/k)$, where $B(x, \delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}$. Show, $V_k \downarrow F$ as $k \to \infty$.
2. Show $\mathcal{B}_0$ is a $\sigma-$ algebra and use this along with the first part of this exercise to conclude $\mathcal{B} = \mathcal{B}_0$. **Hint:** follow closely the method used in the first step of the proof of Theorem 6.24.
3. Show for every $\varepsilon > 0$ and $B \in \mathcal{B}$, there exist a compact subset, $K \subset \mathbb{R}^n$, such that $K \subset B$ and $\mu(B \setminus K) < \varepsilon$. **Hint:** take $K := F \cap \{x \in \mathbb{R}^n : |x| \leq n\}$ for some sufficiently large $n$.

## 6.9 Lebesgue Measure

**Definition 6.60 (Lebesgue Measure).** *If $F(x) = x$ for all $x \in \mathbb{R}$, we denote $\mu_F$ by $m$ and call $m$ **Lebesgue measure** on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.*

**Exercise 6.13.** Given $x \in \mathbb{R} \setminus \{0\}$ let

$$x + B := \{x + y : y \in B\} \text{ and } x \cdot B := \{xy : y \in B\}. \tag{6.25}$$

Use the $\pi - \lambda$ Theorem 6.17 to show $x + B$ and $x \cdot B$ are in $\mathcal{B}_{\mathbb{R}}$ for all $B \in \mathcal{B}_{\mathbb{R}}$.

**Theorem 6.61.** *Lebesgue measure $m$ is invariant under translations, i.e. for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$,*

$$m(x + B) = m(B). \tag{6.26}$$

*Lebesgue measure, $m$, is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0, 1]) = 1$ and Eq. (6.26) holds for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$. Moreover, $m$ has the scaling property*

$$m(\lambda B) = |\lambda| \, m(B) \tag{6.27}$$

*where $\lambda \in \mathbb{R}$, $B \in \mathcal{B}_{\mathbb{R}}$ and $\lambda B := \{\lambda x : x \in B\}$.*

**Proof.** Let $m_x(B) := m(x+B)$, then one easily shows that $m_x$ is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a, b]) = b - a$ for all $a < b$. Therefore, $m_x = m$ by the uniqueness assertion in Exercise 6.3. For the converse, suppose that $m$ is translation invariant and $m((0, 1]) = 1$. Given $n \in \mathbb{N}$, we have

$$(0, 1] = \cup_{k=1}^{n} (\frac{k-1}{n}, \frac{k}{n}] = \cup_{k=1}^{n} \left( \frac{k-1}{n} + (0, \frac{1}{n}] \right).$$

Therefore,

$$\begin{aligned} 1 = m((0, 1]) &= \sum_{k=1}^{n} m \left( \frac{k-1}{n} + (0, \frac{1}{n}] \right) \\ &= \sum_{k=1}^{n} m((0, \frac{1}{n}]) = n \cdot m((0, \frac{1}{n}]). \end{aligned}$$

That is to say

$$m((0, \frac{1}{n}]) = 1/n.$$

Similarly, $m((0, \frac{l}{n}]) = l/n$ for all $l, n \in \mathbb{N}$ and therefore by the translation invariance of $m$,

$$m((a, b]) = b - a \text{ for all } a, b \in \mathbb{Q} \text{ with } a < b.$$

Finally for $a, b \in \mathbb{R}$ such that $a < b$, choose $a_n, b_n \in \mathbb{Q}$ such that $b_n \downarrow b$ and $a_n \uparrow a$, then $(a_n, b_n] \downarrow (a, b]$ and thus

$$m((a, b]) = \lim_{n \to \infty} m((a_n, b_n]) = \lim_{n \to \infty} (b_n - a_n) = b - a,$$

i.e. $m$ is Lebesgue measure. To prove Eq. (6.27) we may assume that $\lambda \neq 0$ since this case is trivial to prove. Now let $m_\lambda(B) := |\lambda|^{-1} m(\lambda B)$. It is easily checked that $m_\lambda$ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

$$m_\lambda((a, b]) = \lambda^{-1} m((\lambda a, \lambda b]) = \lambda^{-1}(\lambda b - \lambda a) = b - a$$

if $\lambda > 0$ and

$$m_\lambda((a, b]) = |\lambda|^{-1} m([\lambda b, \lambda a)) = -|\lambda|^{-1} (\lambda b - \lambda a) = b - a$$

if $\lambda < 0$. Hence $m_\lambda = m$. ■

## 6.10 A Discrete Kolmogorov's Extension Theorem

For this section, let $S$ be a finite or countable set (we refer to $S$ as **state space)**, $\Omega := S^\infty := S^{\mathbb{N}}$ (think of $\mathbb{N}$ as time and $\Omega$ as **path space)**

$$\mathcal{A}_n := \{B \times \Omega : B \subset S^n\} \text{ for all } n \in \mathbb{N},$$

$\mathcal{A} := \cup_{n=1}^{\infty} \mathcal{A}_n$, and $\mathcal{B} := \sigma(\mathcal{A})$. We call the elements, $A \in \mathcal{A}$, the **cylinder subsets of** $\Omega$. Notice that $A \subset \Omega$ is a cylinder set iff there exists $n \in \mathbb{N}$ and $B \subset S^n$ such that

$$A = B \times \Omega := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Also observe that we may write $A$ as $A = B' \times \Omega$ where $B' = B \times S^k \subset S^{n+k}$ for any $k \geq 0$.

**Exercise 6.14.** Show;

1. $\mathcal{A}_n$ is a $\sigma -$ algebra for each $n \in \mathbb{N}$,
2. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all $n$, and
3. $\mathcal{A} \subset 2^\Omega$ is an algebra of subsets of $\Omega$. (In fact, you might show that $\mathcal{A} = \cup_{n=1}^{\infty} \mathcal{A}_n$ is an algebra whenever $\{\mathcal{A}_n\}_{n=1}^{\infty}$ is an increasing sequence of algebras.)

**Lemma 6.62 (Baby Tychonov Theorem 0.).** *Let $S$ be a finite set and $\Omega = S^{\mathbb{N}}$. If $\{A_n\}_{n=1}^{\infty} \subset \Omega$ is a collection of decreasing non-empty sets which are closed under sequential limits then $\cap_{n=1}^{\infty} A_n \neq \emptyset$.*

**Proof.** By assumption for each $n \in \mathbb{N}$ there exists $\omega(n) \in A_n$. We now choose $\omega_1 \in S$ so that $\Gamma_1 := \{n \in \mathbb{N} : \omega_1(n) = \omega_1\}$ is an infinite set. Then choose $\omega_2 \in S$ such that

$$\Gamma_2 := \{n \in \Gamma_1 : \omega_2(n) = \omega_2\} \text{ is an infinite set.}$$

We continue this way inductively so as to choose $\omega = \{\omega_i\}_{i=1}^\infty \subset S$ and infinite subsets $\{\Gamma_i\}_{i=1}^\infty$ of $\mathbb{N}$ such that $\Gamma_{i+1} := \{n \in \Gamma_i : \omega_i(n) = \omega_i\}$. We now let (a Cantor's diagonalization argument)

$$n_1 = \min \Gamma_1, \dots, n_{k+1} := \min\{n > n_k : n \in \Gamma_{k+1}\}.$$

Then $\omega_i(n_k) = \omega_i$ for all large $k$ and $\omega(n_k) \in A_n$ for all $k \geq n$ and therefore $\omega := \lim_{k\to\infty} \omega(n_k) \in A_n$ for all $n$, i.e. $\omega \in \cap_{n=1}^\infty A_n$. ■

**Theorem 6.63 (Kolmogorov's Extension Theorem I.).** *Suppose that $S$ is a finite set. Then **every** finitely additive probability measure, $P : \mathcal{A} \to [0,1]$, has a unique extension to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

**Proof.** If $A \in \mathcal{A}$ then there exists $m \in \mathbb{N}$ and $B \subset S^m$ such that $A = B \times \Omega$. Therefore $\omega(n) \in A$ iff $(\omega_1(n), \dots, \omega_m(n)) \in B$ for all $m$. If we assume $\omega := \lim_{n\to\infty} \omega(n)$ exists then $(\omega_1(n), \dots, \omega_m(n)) = (\omega_1, \dots, \omega_m)$ for a.a. $n$ and this shows $\omega \in A$, i.e. $A$ is necessarily sequentially closed.

From Theorem 6.44, it suffices to show $\lim_{n\to\infty} P(A_n) = 0$ whenever $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ with $A_n \downarrow \emptyset$. However from Lemma 6.62 above or Corollary 6.70 below the only way $A_n \downarrow \emptyset$ is if $A_n = \emptyset$ for a.a. $n$ and in particular $P(A_n) = 0$ for a.a. $n$. This certainly implies $\lim_{n\to\infty} P(A_n) = 0$. ■

**Exercise 6.15 (Look at but do not hand in).** Suppose $S$ is a finite set, $\Omega = S^{\mathbb{N}}$, and $\mathcal{B} := \sigma(\mathcal{A})$ as above. Show every sequence of probability measures $\{P_n\}_{n=1}^\infty$ possesses a subsequence $\{P'_k = P_{n_k}\}_{k=1}^\infty$ such that $\lim_{k\to\infty} P'_k(A) = P(A)$ for all $A \in \mathcal{A}$ where $P$ is a probability measure on $(\Omega, \mathcal{B})$. **Hint:** use Cantor's diagonalization argument.

**Definition 6.64 (A one point compactifiction).** *If $S$ is a countably infinite set let $\bar{S} = S \cup \{\infty\}$ where "$\infty$" (read infinity) is simply another point not in $S$. Let $\{x_n\}_{n=1}^\infty \subset \bar{S}$ be a sequence, then we way $\lim_{n\to\infty} x_n = \infty$ if for every $A \subset_f S$,[4] $x_n \notin A$ for almost all $n$ and we say that $\lim_{n\to\infty} x_n = s \in S$ iff $x_n = s$ for almost all $n$.*

For example this is the usual notion of convergence for $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ and $\bar{S} = S \cup \{0\} \subset [0,1]$, where 0 is playing the role of infinity here.

[4] Recall that $A \subset_f S$ means $A$ is a finite subset of $S$.

**Lemma 6.65 ($\bar{S}$ is sequentially compact).** *Let $\bar{S} := S$ is $S$ is a finite set and $\bar{S} = S \cup \{\infty\}$ if $S$ is an infinite set. Then every sequence $\{x_n\}_{n=1}^\infty \subset \bar{S}$ has a convergent subsequence.*

**Proof.** If $S$ is a finite set so that $\bar{S} = S$. In this case there exists $x \in S$ such that $x_n = x$ for infinitely many $n$. Hence we may find $n_1 < n_2 < \dots < n_k \uparrow \infty$ such that $x_{n_k} = x$ for all $k$ and therefore $\lim_{k\to\infty} x_{n_k} = x$.

Now suppose $S$ is an infinite set so that $\bar{S} = S \cup \{\infty\}$. Observe that either $\lim_{n\to\infty} x_n = \infty$ or there exists a finite subset $F \subset S$ such that $x_n \in F$ infinitely often. In the latter case we may follow the proof above to $n_1 < n_2 < \dots < n_k \uparrow \infty$ such that $x_{n_k} = x$ for all $k$ and therefore $\lim_{k\to\infty} x_{n_k} = x$. ■

**Definition 6.66.** *Let $\bar{\Omega} := \bar{S}^{\mathbb{N}}$ and $\{\omega(n)\}_{n=1}^\infty$ be a sequence in $\bar{\Omega}$. We say $\lim_{n\to\infty} \omega(n) = \omega \in \bar{\Omega}$ iff $\lim_{n\to\infty} \omega_i(n) = \omega_i \in S$ for all $i \in \mathbb{N}$. We say $F \subset \bar{\Omega}$ is **sequential closed** iff for every sequence $\{\omega(n)\}_{n=1}^\infty \subset F$ such that $\omega = \lim_{n\to\infty} \omega(n)$ exists in $\bar{\Omega}$ we have $\omega \in F$.*

*Example 6.67.* Suppose that $F = K \times \bar{\Omega}$ where $K \subset_f S^N$ for some $N \in \mathbb{N}$.[5] Indeed if $\{\omega(n)\}_{n=1}^\infty \subset F$ then $(\omega_1(n), \dots, \omega_N(n)) \in K$ for all $n$. Moreover since $\lim_{n\to\infty} \omega(n) = \omega$ exits we must in fact have $\omega_i(n) = \omega_i$ for a.a. $n$ for each $1 \leq i \leq n$, i.e. $(\omega_1, \dots, \omega_N) = (\omega_1(n), \dots, \omega_N(n))$ for a.a. $n$ and this shows $(\omega_1, \dots, \omega_N) \in K$, i.e. $\omega \in F$.

**Lemma 6.68 (Baby Tychonov Theorem I.).** *The set $\bar{\Omega} := \bar{S}^{\mathbb{N}}$ is sequentially compact. In detail, to every sequence $\{\omega(n)\}_{n=1}^\infty \subset \bar{\Omega}$, there exits a subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$ such that $\lim_{k\to\infty} \omega(n_k)$ exists in $\bar{\Omega}$.*

**Proof.** This follows by the usual Cantor's diagonalization argument. Indeed, let $\{n_k^1\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ be chosen so that $\lim_{k\to\infty} \omega_1(n_k^1) = s_1 \in \bar{S}$ exists. Then choose $\{n_k^2\}_{k=1}^\infty \subset \{n_k^1\}_{k=1}^\infty$ so that $\lim_{k\to\infty} \omega_2(n_k^2) = s_2 \in \bar{S}$ exists. Continue on this way to inductively choose

$$\{n_k^1\}_{k=1}^\infty \supset \{n_k^2\}_{k=1}^\infty \supset \dots \supset \{n_k^l\}_{k=1}^\infty \supset \dots$$

such that $\lim_{k\to\infty} \omega_l(n_k^l) = s_l \in \bar{S}$. The subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$, may now be defined by, $n_k = n_k^k$. ■

**Corollary 6.69 (Baby Tychonov Theorem II.).** *Suppose that $\{F_n\}_{n=1}^\infty \subset \bar{\Omega}$ is decreasing sequence of non-empty sets which are closed under taking sequential limits, then $\cap_{n=1}^\infty F_n \neq \emptyset$.*

**Proof.** Since $F_n \neq \emptyset$ there exists $\omega(n) \in F_n$ for all $n$. Using Lemma 6.68, there exists $\{n_k\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ such that $\omega := \lim_{k\to\infty} \omega(n_k)$ exits in $\bar{\Omega}$. Since $\omega(n_k) \in F_n$ for all $k \geq n$, it follows that $\omega \in F_n$ for all $n$, i.e. $\omega \in \cap_{n=1}^\infty F_n$ and hence $\cap_{n=1}^\infty F_n \neq \emptyset$. ■

[5] Recall that $K \subset_f S^N$ means $K$ is a finite subset of $S^N$.

**Corollary 6.70.** *If $S$ is a finite set and $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

**Proof.** Since $S$ is a finite set $\bar{\Omega} = \Omega$ and $A_n = B_n \times \Omega$ with $B_n \subset S^{N_n}$ is a sequentially closed set by Example 6.67. Therefore the result follows from Corollary 6.69 with $F_n = A_n$. ∎

**Corollary 6.71 (of Corollary 6.69).** *Let $S$ be a countably infinite set. Suppose that $1 \leq N_1 < N_2 < N_3 < \dots$, $F_n = K_n \times \Omega$ with $K_n \subset_f S^{N_n}$ such that $\{F_n\}_{n=1}^\infty \subset \Omega$ is a decreasing sequence of non-empty sets. Then $\cap_{n=1}^\infty F_n \neq \emptyset$.*

**Proof.** Let $\bar{F}_n := K_n \times \bar{\Omega}$ in which case $\bar{F}_n$ are non − empty sequential closed set by Example 6.67. Therefore by Corollary 6.69, $\cap_n \bar{F}_n \neq \emptyset$. This completes the proof since it is easy to check that $\cap_{n=1}^\infty F_n = \cap_n \bar{F}_n \neq \emptyset$. [Here we use that $N_k$ is strictly increasing in $k$.] ∎

**Exercise 6.16 ("Finite" approximation theorem).** Suppose that $S$ is a countable set and the hypothesis of Theorem 6.72 are in force. Given $n \ni \mathbb{N}$, $\varepsilon > 0$, and $A \in \mathcal{A}_n$, show there exists $F \in \mathcal{A}_n$ such that $F \subset A$, $P(A \setminus F) < \varepsilon$, and $F = K \times \Omega$ with $K \subset_f S^n$.

**Exercise 6.17 (Look at but do not hand in).** Suppose that $S$ is a countable set and the hypothesis of Theorem 6.72 are in force. Let $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ be a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$. [As you should check (but do not hand in), we may choose $N_m \in \mathbb{N}$ strictly increasing so that $A_m = B_m \times \Omega$ for some $B_m \subset S^{N_m}$.] Using Exercise 6.16, we may choose and $K_m \subset_f S^{N_m}$ such that $F_m = K_m \times \Omega \subset A_m$ with $P(A_m \setminus F_m) \leq \varepsilon/2^{m+1}$. Further define $C_m := F_1 \cap \dots \cap F_m$ for each $m$. Show;

1. Show $A_m \setminus C_m \subset (A_1 \setminus F_1) \cup (A_2 \setminus F_2) \cup \dots \cup (A_m \setminus F_m)$ and use this to conclude that $P(A_m \setminus C_m) \leq \varepsilon/2$.
2. Conclude $P(C_m) \geq \varepsilon/2$ and in particular that $C_m$ is not empty for all $m$.
3. Use Corollary 6.71 to conclude that $\emptyset \neq \cap_{m=1}^\infty C_m \subset \cap_{m=1}^\infty A_m$.

**Theorem 6.72 (Kolmogorov's Extension Theorem II).** *Suppose now that $S$ is countably infinite set and $P : \mathcal{A} \to [0,1]$ is a finitely additive measure such that $P|_{\mathcal{A}_n}$ is a $\sigma$ – additive measure for each $n \in \mathbb{N}$. Then $P$ extends uniquely to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

**Proof.** From Theorem 6.44 it suffice to show; if $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$, then $\cap_{m=1}^\infty A_m \neq \emptyset$. But this last statement is a consequence of item 3. of Exercise 6.17. ∎

*Example 6.73 (Markov Chain Probabilities).* Let $S$ be a finite or at most countable state space and $p : S \times S \to [0,1]$ be a **Markov kernel,** i.e.

$$\sum_{y \in S} p(x,y) = 1 \text{ for all } x \in S. \tag{6.28}$$

Also let $\pi : S \to [0,1]$ be a probability function, i.e. $\sum_{x \in S} \pi(x) = 1$. We now take

$$\Omega := S^{\mathbb{N}_0} = \{\omega = (s_0, s_1, \dots) : s_j \in S\}$$

and let $X_n : \Omega \to S$ be given by

$$X_n(s_0, s_1, \dots) = s_n \text{ for all } n \in \mathbb{N}_0.$$

Then there exists a unique probability measure, $P_\pi$, on $\sigma(\mathcal{A})$ such that

$$P_\pi(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

for all $n \in \mathbb{N}_0$ and $x_0, x_1, \dots, x_n \in S$. To see such a measure exists, we need only verify that

$$p_n(x_0, \dots, x_n) := \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

verifies the hypothesis of Exercise 6.18 taking into account a shift of the $n$ – index.

For the next three exercises, suppose that $S$ is a finite set and continue the notation from above. Further suppose that $P : \sigma(\mathcal{A}) \to [0,1]$ is a probability measure and for $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, let

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \tag{6.29}$$

**Exercise 6.18 (Consistency Conditions).** If $p_n$ is defined as above, show:

1. $\sum_{s \in S} p_1(s) = 1$ and
2. for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$,

$$p_n(s_1, \dots, s_n) = \sum_{s \in S} p_{n+1}(s_1, \dots, s_n, s).$$

**Exercise 6.19 (Converse to 6.18).** Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : S^n \to [0,1]$ such that the consistency conditions in Exercise 6.18 hold. Then there exists a unique probability measure, $P$ on $\sigma(\mathcal{A})$ such that Eq. (6.29) holds for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$.

*Example 6.74 (Existence of iid simple R.V.s).* Suppose now that $q : S \to [0,1]$ is a function such that $\sum_{s \in S} q(s) = 1$. Then there exists a unique probability measure $P$ on $\sigma(\mathcal{A})$ such that, for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, we have

$$P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}) = q(s_1) \dots q(s_n).$$

This is a special case of Exercise 6.19 with $p_n(s_1, \dots, s_n) := q(s_1) \dots q(s_n)$.

**Exercise 6.20.** Convince yourself that the results of Exercise 6.18 and 6.19 are valid when $S$ is a countable set. (See Example 5.6.)

**In summary**, the main result of this section states, to any sequence of functions, $p_n : S^n \to [0,1]$, such that $\sum_{\lambda \in S^n} p_n(\lambda) = 1$ and $\sum_{s \in S} p_{n+1}(\lambda, s) = p_n(\lambda)$ for all $n$ and $\lambda \in S^n$, there exists a unique probability measure, $P$, on $\mathcal{B} := \sigma(\mathcal{A})$ such that

$$P(B \times \Omega) = \sum_{\lambda \in B} p_n(\lambda) \ \forall \ B \subset S^n \text{ and } n \in \mathbb{N}.$$

## 6.11 Appendix: Completions of Measure Spaces*

**Definition 6.75.** *A set $E \subset \Omega$ is a* ***null*** *set if $E \in \mathcal{B}$ and $\mu(E) = 0$. If $P$ is some "property" which is either true or false for each $x \in \Omega$, we will use the terminology $P$ a.e. (to be read $P$ almost everywhere) to mean*

$$E := \{x \in \Omega : P \text{ is false for } x\}$$

*is a null set. For example if $f$ and $g$ are two measurable functions on $(\Omega, \mathcal{B}, \mu)$, $f = g$ a.e. means that $\mu(f \neq g) = 0$.*

**Definition 6.76.** *A measure space $(\Omega, \mathcal{B}, \mu)$ is* ***complete*** *if every subset of a null set is in $\mathcal{B}$, i.e. for all $F \subset \Omega$ such that $F \subset E \in \mathcal{B}$ with $\mu(E) = 0$ implies that $F \in \mathcal{B}$.*

**Proposition 6.77 (Completion of a Measure).** *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Set*

$$\mathcal{N} = \mathcal{N}^{\mu} := \{N \subset \Omega : \exists\, F \in \mathcal{B} \text{ such that } N \subset F \text{ and } \mu(F) = 0\},$$
$$\mathcal{B} = \bar{\mathcal{B}}^{\mu} := \{A \cup N : A \in \mathcal{B} \text{ and } N \in \mathcal{N}\} \text{ and}$$
$$\bar{\mu}(A \cup N) := \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N},$$

*see Fig. 6.2. Then $\bar{\mathcal{B}}$ is a $\sigma$ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{B}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{B}}$ which extends $\mu$ on $\mathcal{B}$, and $(\Omega, \bar{\mathcal{B}}, \bar{\mu})$ is complete measure space. The $\sigma$-algebra, $\bar{\mathcal{B}}$, is called the* ***completion*** *of $\mathcal{B}$ relative to $\mu$ and $\bar{\mu}$, is called the* ***completion of*** *$\mu$.*

**Proof.** Clearly $\Omega, \emptyset \in \bar{\mathcal{B}}$. Let $A \in \mathcal{B}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{B}$ such that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$\begin{aligned}(A \cup N)^c &= A^c \cap N^c = A^c \cap (F \setminus N \cup F^c) \\ &= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]\end{aligned}$$

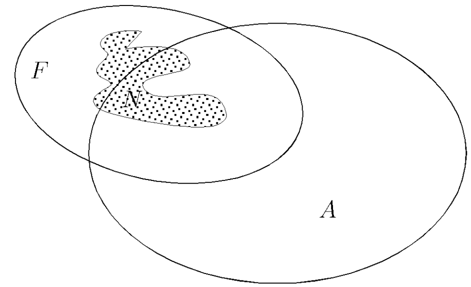



**Fig. 6.2.** Completing a $\sigma$ – algebra.

where $[A^c \cap (F \setminus N)] \in \mathcal{N}$ and $[A^c \cap F^c] \in \mathcal{B}$. Thus $\bar{\mathcal{B}}$ is closed under complements. If $A_i \in \mathcal{B}$ and $N_i \subset F_i \in \mathcal{B}$ such that $\mu(F_i) = 0$ then $\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \bar{\mathcal{B}}$ since $\cup A_i \in \mathcal{B}$ and $\cup N_i \subset \cup F_i$ and $\mu(\cup F_i) \leq \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{B}}$ is a $\sigma$ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{B}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \leq \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \leq \mu(A)$ so that $\mu(A) = \mu(B)$ and hence $\bar{\mu}(A \cup N) := \mu(A)$ is well defined. It is left as an exercise to show $\bar{\mu}$ is a measure, i.e. that it is countable additive. ■

## 6.12 Supplement: Generalizations of Theorem 6.58 to $\mathbb{R}^n$

**Notation 6.78** *Let $a, b \in \bar{\mathbb{R}}^n$ we say $a < b$ if $a_i < b_i$ for all $i$ and $a \leq b$ if $a_i \leq b_i$ for all $i$. For $a < b$ we let*

$$(a, b] := (a_1, b_1] \cap \mathbb{R} \times \cdots \times (a_n, b_n] \cap \mathbb{R}.$$

*Also define $x \wedge y := (x_1 \wedge y_1, \ldots, x_n \wedge y_n)$ for all $x, y \in \mathbb{R}^n$.*

**Definition 6.79.** *A function $F : \mathbb{R}^n \to \mathbb{C}$ (or any normed space) is* ***right continuous*** *at $a \in \mathbb{R}^n$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|F(b) - F(a)| < \varepsilon$ for all $b \in [a, \infty)$ such that $|b - a| \leq \delta$. Alternative put, $F$ is right continuous at $a$ iff for every $\varepsilon > 0$ there exists $b \in \mathbb{R}^n$ such that $b > a$ and $|F(y) - F(a)| < \varepsilon$ for all $a \leq y < b$.*

**Definition 6.80.** *Let $\mathcal{A} = \mathcal{A}(\mathbb{R}^n)$ be the sub-algebra of $\mathcal{B}$ generated by sets of the form $(a, b] \cap \mathbb{R}^n$ with $a, b \in \bar{\mathbb{R}}^n$. [Note that $(a, b] = \emptyset$ if $a_i \geq b_i$ for some $i$.]*

**Theorem 6.81.** *Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is a right continuous function and $\mu_F : \mathcal{A}(\mathbb{R}^n) \to \mathbb{R}$ is the associated finitely additive measure as given in Eq. (5.40)* *of Proposition 5.41. If $\mu_F$ is positive*[6] *in the sense that $\mu_F((a,b]) \geq 0$ for all $a \leq b$, then there exists a unique measure $\mu = \bar{\mu}_F$ on $\mathcal{B}_{\mathbb{R}^n}$ such that*

$$\mu((a,b]) = \mu_F((a,b]) = \sum_{\Lambda \subset \{1,2,\dots,n\}} (-1)^{|\Lambda|} F(a_\Lambda \times b_{\Lambda^c})$$

*for all $a \leq b$ with $a, b \in \mathbb{R}^n$.*

**Proof.** We let $\mu_F$ be the finitely additive measure on $\mathcal{A} = \mathcal{A}(\mathbb{R}^n)$ given in Proposition 5.41, which (by assumption) is non-negative on $\mathcal{A}(\mathbb{R}^n)$. So to finish the proof we need only show $\mu := \mu_F$ is a premeasure on $\mathcal{A}$ which we will do by showing $\mu$ is subadditive on $\mathcal{E} := \{(a,b] : a, b \in \mathbb{R}^n \text{ and } a \leq b\}$, i.e. if $(a,b] = \sum_{n=1}^\infty (a_n, b_n]$ then we must show

$$\mu((a,b]) \leq \sum_{n=1}^\infty \mu((a_n, b_n]).$$

We may suppose that $a < b$ (i.e. $a_i < b_i$ for all $i$) for otherwise $(a,b] = \emptyset$ and $\mu((a,b]) = 0$ and there will be nothing to prove. For any choice of $a < \tilde{a} < b$ and $\tilde{b}_n > b_n$ for all $n \in \mathbb{N}$, we have

$$[\tilde{a}, b] \subset (a,b] \subset \cup_n \left(a_n, \tilde{b}_n\right).$$

So by compactness there exists $N < \infty$ such that

$$(\tilde{a}, b] \subset [\tilde{a}, b] \subset \cup_{n=1}^N \left(a_n, \tilde{b}_n\right) \subset \cup_{n=1}^N (a_n, \tilde{b}_n].$$

By finite sub-additivity and monotinicity of $\mu$, we have

$$\mu((\tilde{a}, b]) \leq \sum_{n=1}^N \mu\left((a_n, \tilde{b}_n]\right) \leq \sum_{n=1}^\infty \mu\left((a_n, \tilde{b}_n]\right).$$

Using the right continuity of $F$ it follows that

$$\mu((a,b]) = \lim_{\tilde{a} \downarrow a} \mu((\tilde{a}, b]) \leq \sum_{n=1}^\infty \mu\left((a_n, \tilde{b}_n]\right). \tag{6.30}$$

Also using the right continuity of $F$, for every $\varepsilon > 0$ we may choose $\tilde{b}_n > b_n$ such that $\mu\left((a_n, \tilde{b}_n]\right) \leq \mu((a_n, b_n]) + \varepsilon 2^{-n}$ which combined with Eq. (6.30) implies,

$$\mu((a,b]) \leq \sum_{n=1}^\infty \left[\mu((a_n, b_n]) + \varepsilon 2^{-n}\right] = \sum_{n=1}^\infty \mu((a_n, b_n]) + \varepsilon.$$

As $\varepsilon > 0$ is arbitrary the proof is complete in the case $a < b$.

The proof for the cases where some of the components of $a$ are $-\infty$ and or some the components of $b$ are $+\infty$ follows as in the proof to Theorem 6.58. ∎

**Lemma 6.82 (Right Continuous Versions).** *Suppose $G : \mathbb{R}^n \to \mathbb{R}$ is increasing in each of its variables.*[7] *For $x \in \mathbb{R}^n$ let $F(x) := \inf_{y > x} G(y)$. Then $F$ is increasing in each of its variables and $F$ is right continuous.*

**Proof.** If $a \leq b$ and $y > b$ then $y > a$ so that $F(a) \leq G(y)$. Therefore $F(a) \leq \inf_{y > b} G(y) = F(b)$ and so $F$ is increasing in each of its variables. Now suppose that $\varepsilon > 0$ there exists $\beta > a$ such that $0 \leq G(\beta) - F(a) < \varepsilon$. Then for any $a < b < \beta$, we have

$$0 \leq G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon \text{ for all } b < y < \beta.$$

From this it follows that

$$0 \leq F(b) - F(a) \leq \inf_{b < y < \beta} G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon$$

which proves the right continuity of $F$. ∎

## 6.13 Appendix: Alternate measure extension construction

(The reader wanting for more motivation of the construction measure in this section may wish to read Section 6.3 below first.)

Suppose $\mu$ is a **finite** premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and $A \in \mathcal{A}_\delta \cap \mathcal{A}_\sigma$. Since $A, A^c \in \mathcal{A}_\sigma$ and $\Omega = A \cup A^c$, it follows that $\mu(\Omega) = \mu(A) + \mu(A^c)$. From this observation we may extend $\mu$ to a function on $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ by defining

$$\mu(A) := \mu(\Omega) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{6.31}$$

[6] It is not sufficient to assume that $F$ is non-decreasing in each of it variables. For example $F(x,y) := x + y - xy$ on $[0,1]^2$ satisfies, $F_x = 1 - y \geq 0$ and $F_y = 1 - x \geq 0$ while $F_{x,y} = -1 < 0$. In this case $\mu_F\left([0,1]^2\right) = -1 < 0$ and more generally $\mu_F((\mathbf{a}, \mathbf{b}]) = -(b_1 - a_1)(b_2 - a_2) < 0$ for all $\mathbf{a} < \mathbf{b}$. Moreover if $F$ is sufficiently smooth, then non-decreasing in each of its variables means $\partial_i F \geq 0$ for all $i$ whereas $d\mu_F = \partial_1 \dots \partial_n F dm$ will be a positive measure iff $\partial_1 \dots \partial_n F \geq 0$.

[7] The increasing in each of its variables assumption is a bit of a red herring. For example, if $F(x,y) = xy - x - y$, then $\mu_F =$ Lebesgue measure on the plane since $F_{x,y} = 1$. However $F_x = y - 1$ and $F_y = x - 1$ has variable signs. Actually we could use the simpler example of $F(x,y) = xy$ just as well.

**Lemma 6.83.** *Suppose $\mu$ is a finite premeasure on an algebra, $\mathcal{A} \subset 2^{\Omega}$, and $\mu$ has been extended to $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ as described in Proposition 6.38 and Eq. (6.31) above.*

1. *If $A \in \mathcal{A}_\delta$ then $\mu(A) = \inf\{\mu(B) : A \subset B \in \mathcal{A}\}$.*
2. *If $A \in \mathcal{A}_\delta$ and $A_n \in \mathcal{A}$ such that $A_n \downarrow A$, then $\mu(A) = \downarrow \lim_{n\to\infty} \mu(A_n)$.*
3. *$\mu$ is strongly additive when restricted to $\mathcal{A}_\delta$.*
4. *If $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset C$, then $\mu(C \setminus A) = \mu(C) - \mu(A)$.*

**Proof.**

1. Since $\mu(B) = \mu(\Omega) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that
$$\begin{aligned}\inf\{\mu(B) : A \subset B \in \mathcal{A}\} &= \inf\{\mu(\Omega) - \mu(B^c) : \mathcal{A} \ni B^c \subset A^c\} \\ &= \mu(\Omega) - \sup\{\mu(B) : \mathcal{A} \ni B \subset A^c\} \\ &= \mu(\Omega) - \mu(A^c) = \mu(A).\end{aligned}$$
2. Similarly, since $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 6.38 it follows that
$$\begin{aligned}\mu(A) = \mu(\Omega) - \mu(A^c) &= \mu(\Omega) - \uparrow \lim_{n\to\infty} \mu(A_n^c) \\ &= \downarrow \lim_{n\to\infty} [\mu(\Omega) - \mu(A_n^c)] = \downarrow \lim_{n\to\infty} \mu(A_n).\end{aligned}$$
3. Suppose $A, B \in \mathcal{A}_\delta$ and $A_n, B_n \in \mathcal{A}$ such that $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \cup B_n \downarrow A \cup B$ and $A_n \cap B_n \downarrow A \cap B$ and therefore,
$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= \lim_{n\to\infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] \\ &= \lim_{n\to\infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).\end{aligned}$$
All we really need is the finite additivity of $\mu$ which can be proved as follows. Suppose that $A, B \in \mathcal{A}_\delta$ are disjoint, then $A \cap B = \emptyset$ implies $A^c \cup B^c = \Omega$. So by the strong additivity of $\mu$ on $\mathcal{A}_\sigma$ it follows that
$$\mu(\Omega) + \mu(A^c \cap B^c) = \mu(A^c) + \mu(B^c)$$
from which it follows that
$$\begin{aligned}\mu(A \cup B) &= \mu(\Omega) - \mu(A^c \cap B^c) \\ &= \mu(\Omega) - [\mu(A^c) + \mu(B^c) - \mu(\Omega)] \\ &= \mu(A) + \mu(B).\end{aligned}$$
4. Since $A^c, C \in \mathcal{A}_\sigma$ we may use the strong additivity of $\mu$ on $\mathcal{A}_\sigma$ to conclude,
$$\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).$$
Because $\Omega = A^c \cup C$, and $\mu(A^c) = \mu(\Omega) - \mu(A)$, the above equation may be written as
$$\mu(\Omega) + \mu(C \setminus A) = \mu(\Omega) - \mu(A) + \mu(C)$$
which finishes the proof.

■

If $B \subset \Omega$ has the same inner and outer content (see Notations 6.49 and 6.47 respectively) it is reasonable to define the measure of $B$ as this common value. As we will see in Theorem 6.86 below, this extension becomes a $\sigma$ – additive measure on a $\sigma$ – algebra of subsets of $\Omega$.

**Definition 6.84 (Measurable Sets).** *Suppose $\mu$ is a finite premeasure on an algebra $\mathcal{A} \subset 2^{\Omega}$. We say that $B \subset \Omega$ is **measurable** if $\mu_*(B) = \mu^*(B)$ where $\mu_*(B)$ and $\mu^*(B)$ as in Notations 6.49 and 6.47 respectively. We will denote the collection of measurable subsets of $\Omega$ by $\mathcal{B} = \mathcal{B}(\mu)$ and define $\bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)]$ by*
$$\bar{\mu}(B) := \mu_*(B) = \mu^*(B) \text{ for all } B \in \mathcal{B}. \tag{6.32}$$

*Remark 6.85.* Observe that $\mu_*(B) = \mu^*(B)$ iff for all $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and
$$\mu(C \setminus A) = \mu(C) - \mu(A) < \varepsilon, \tag{6.33}$$
wherein we have used Lemma 6.83 for the first equality. Moreover we will use below that if $B \in \mathcal{B}$ and $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$, then
$$\mu(A) \le \mu_*(B) = \bar{\mu}(B) = \mu^*(B) \le \mu(C). \tag{6.34}$$

**Theorem 6.86 (Finite Premeasure Extension Theorem (revisited)).** *Suppose $\mu$ is a finite premeasure on an algebra $\mathcal{A} \subset 2^{\Omega}$ and $\bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(\Omega)]$ be as in Definition 6.84. Then $\mathcal{B}$ is a $\sigma$ – algebra on $\Omega$ which contains $\mathcal{A}$ and $\bar{\mu}$ is a $\sigma$ – additive measure on $\mathcal{B}$. Moreover, $\bar{\mu}$ is the unique measure on $\mathcal{B}$ such that $\bar{\mu}|_{\mathcal{A}} = \mu$.*

**Proof.** 1. **$\mathcal{B}$ is an algebra.** It is clear that $\mathcal{A} \subset \mathcal{B}$ and that $\mathcal{B}$ is closed under complementation – see Eq. (6.33) and use the fact that $A^c \setminus C^c = C \setminus A$. Now suppose that $B_i \in \mathcal{B}$ for $i = 1, 2$ and $\varepsilon > 0$ is given. We may then choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon$ for $i = 1, 2$. Then with $A = A_1 \cup A_2$, $B = B_1 \cup B_2$ and $C = C_1 \cup C_2$, we have $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$. Since
$$C \setminus A = (C_1 \setminus A) \cup (C_2 \setminus A) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2),$$
it follows from the sub-additivity of $\mu$ that

$$\mu(C\setminus A)\le\mu(C_1\setminus A_1)+\mu(C_2\setminus A_2)<2\varepsilon.$$

Since $\varepsilon>0$ was arbitrary, we have shown that $B\in\mathcal{B}$ which completes the proof that $\mathcal{B}$ is an algebra.

2. $\mathcal{B}$ **is a** $\sigma-$**algebra.** As we know $\mathcal{B}$ is an algebra, to show $\mathcal{B}$ is a $\sigma-$ algebra it suffices to show that $B=\sum_{n=1}^\infty B_n\in\mathcal{B}$ whenever $\{B_n\}_{n=1}^\infty$ is a disjoint sequence in $\mathcal{B}$. To this end, let $\varepsilon>0$ be given and choose $A_i\subset B_i\subset C_i$ such that $A_i\in\mathcal{A}_\delta$, $C_i\in\mathcal{A}_\sigma$, and $\mu(C_i\setminus A_i)<\varepsilon 2^{-i}$ for all $i$. Let $C:=\cup_{i=1}^\infty C_i\in\mathcal{A}_\sigma$ and for $n\in\mathbb{N}$ let $A^n:=\sum_{i=1}^n A_i\in\mathcal{A}_\delta$. Since the $\{A_i\}_{i=1}^\infty$ are pairwise disjoint we may use Lemma 6.83 to show,

$$\sum_{i=1}^n\mu(C_i)=\sum_{i=1}^n(\mu(A_i)+\mu(C_i\setminus A_i))$$
$$=\mu(A^n)+\sum_{i=1}^n\mu(C_i\setminus A_i)\le\mu(\Omega)+\sum_{i=1}^n\varepsilon 2^{-i}$$

which on letting $n\to\infty$ implies

$$\sum_{i=1}^\infty\mu(C_i)\le\mu(\Omega)+\varepsilon<\infty. \tag{6.35}$$

Using

$$C\setminus A^n=\cup_{i=1}^\infty(C_i\setminus A^n)\subset\left[\cup_{i=1}^n(C_i\setminus A_i)\right]\cup\left[\cup_{i=n+1}^\infty C_i\right]\in\mathcal{A}_\sigma,$$

and the sub-additivity of $\mu$ on $\mathcal{A}_\sigma$ it follows that

$$\mu(C\setminus A^n)\le\sum_{i=1}^n\mu(C_i\setminus A_i)+\sum_{i=n+1}^\infty\mu(C_i)\le\varepsilon\sum_{i=1}^n 2^{-i}+\sum_{i=n+1}^\infty\mu(C_i)$$
$$\le\varepsilon+\sum_{i=n+1}^\infty\mu(C_i)\to\varepsilon\text{ as }n\to\infty,$$

wherein we have used Eq. (6.35) in computing the limit. In summary, $B=\cup_{i=1}^\infty B_i$, $\mathcal{A}_\delta\ni A^n\subset B\subset C\in\mathcal{A}_\sigma$, $C\setminus A^n\in\mathcal{A}_\sigma$ with $\mu(C\setminus A^n)\le 2\varepsilon$ for all $n$ sufficiently large. Since $\varepsilon>0$ is arbitrary, it follows that $B\in\mathcal{B}$.

**3.** $\bar\mu$ **is a measure.** Continuing the notation in step 2, we have

$$\sum_{i=1}^\infty\mu(A_i)\overset{n\to\infty}{\longleftarrow}\sum_{i=1}^n\mu(A_i)=\mu(A^n)\le\bar\mu(B)\le\mu(C)\le\sum_{i=1}^\infty\mu(C_i). \tag{6.36}$$

On the other hand, since $A_i\subset B_i\subset C_i$, it follows (see Eq. (6.34)) that $\mu(A_i)\le\bar\mu(B_i)\le\mu(C_i)$ and therefore that



$$\sum_{i=1}^\infty\mu(A_i)\le\sum_{i=1}^\infty\bar\mu(B_i)\le\sum_{i=1}^\infty\mu(C_i). \tag{6.37}$$

Equations (6.36) and (6.37) show that $\bar\mu(B)$ and $\sum_{i=1}^\infty\bar\mu(B_i)$ are both between $\sum_{i=1}^\infty\mu(A_i)$ and $\sum_{i=1}^\infty\mu(C_i)$ and so

$$\left|\bar\mu(B)-\sum_{i=1}^\infty\bar\mu(B_i)\right|\le\sum_{i=1}^\infty\mu(C_i)-\sum_{i=1}^\infty\mu(A_i)=\sum_{i=1}^\infty\mu(C_i\setminus A_i)\le\sum_{i=1}^\infty\varepsilon 2^{-i}=\varepsilon.$$

Since $\varepsilon>0$ is arbitrary, we have shown $\bar\mu(B)=\sum_{i=1}^\infty\bar\mu(B_i)$, i.e. $\bar\mu$ is a measure on $\mathcal{B}$.

Since we really had no choice as to how to extend $\mu$, it is to be expected that the extension is unique. You are asked to supply the details in Exercise 6.9 below. ■

**Corollary 6.87.** *Suppose that* $\mathcal{A}\subset 2^\Omega$ *is an algebra and* $\mu:\mathcal{B}_0:=\sigma(\mathcal{A})\to[0,\mu(\Omega)]$ *is a* $\sigma$ – *additive finite measure. Then for every* $B\in\sigma(\mathcal{A})$ *and* $\varepsilon>0$*;*

1. *there exists* $\mathcal{A}_\delta\ni A\subset B\subset C\in\mathcal{A}_\sigma$ *and* $\varepsilon>0$ *such that* $\mu(C\setminus A)<\varepsilon$ *and*
2. *there exists* $A\in\mathcal{A}$ *such that* $\mu(A\bigtriangleup B)<\varepsilon$.

**Exercise 6.21 (Look at but do not hand in).** Prove corollary 6.87 by considering $\bar\nu$ where $\nu:=\mu|_{\mathcal{A}}$. **Hint:** you may find Exercise 5.6 useful here.

# 7

# Carathéodory's Construction of Measures*

This chapter deals with Carathéodory's very general measure construction theorem. This chapter may be safely skipped as we will not make direct use of the resutls here in the remainder of this book.

## 7.1 General Extension and Construction Theorem

Proposition 6.48 motivates the following definition.

**Definition 7.1.** *Let $\mu^* : 2^\Omega \to [0,\infty]$ be an outer measure, see Definition 6.28. Define the $\mu^*$-**measurable sets** to be*

$$\mathcal{M}(\mu^*) := \{B \subset \Omega : \mu^*(E) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \ \forall \ E \subset \Omega\}.$$

*Because of the sub-additivity of $\mu^*$, we may equivalently define $\mathcal{M}(\mu^*)$ by*

$$\mathcal{M}(\mu^*) = \{B \subset \Omega : \mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c) \ \forall \ E \subset \Omega\}. \tag{7.1}$$

**Theorem 7.2 (Carathéodory's Construction Theorem).** *Let $\mu^*$ be an outer measure on $\Omega$ and $\mathcal{M} := \mathcal{M}(\mu^*)$. Then $\mathcal{M}$ is a $\sigma$-algebra and $\mu := \mu^*|_{\mathcal{M}}$ is a complete measure.*

**Proof.** Clearly $\emptyset, \Omega \in \mathcal{M}$ and if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$. So to show that $\mathcal{M}$ is an algebra we must show that $\mathcal{M}$ is closed under finite unions, i.e. if $A, B \in \mathcal{M}$ and $E \in 2^\Omega$ then

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu(E \setminus (A \cup B)).$$

Using the definition of $\mathcal{M}$ three times, we have

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \setminus A) &(7.2)\\ &= \mu^*(E \cap A \cap B) + \mu^*((E \cap A) \setminus B)\\ &\quad + \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B). &(7.3)\end{aligned}$$

By the sub-additivity of $\mu^*$ and the set identity,

$$\begin{aligned}E \cap (A \cup B) &= (E \cap A) \cup (E \cap B)\\ &= [((E \cap A) \setminus B) \cup (E \cap A \cap B)] \cup [((E \cap B) \setminus A) \cup (E \cap A \cap B)]\\ &= [E \cap A \cap B] \cup [(E \cap A) \setminus B] \cup [(E \setminus A) \cap B],\end{aligned}$$

we have

$$\mu^*(E \cap A \cap B) + \mu^*((E \cap A) \setminus B) + \mu^*((E \setminus A) \cap B) \geq \mu^*(E \cap (A \cup B)).$$

Using this inequality in Eq. (7.3) shows

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B)) \tag{7.4}$$

which implies $A \cup B \in \mathcal{M}$. So $\mathcal{M}$ is an algebra. Now suppose $A, B \in \mathcal{M}$ are disjoint, then taking $E = A \cup B$ in Eq. (7.2) implies

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$$

and $\mu = \mu^*|_{\mathcal{M}}$ is finitely additive on $\mathcal{M}$.

We now must show that $\mathcal{M}$ is a $\sigma -$ algebra and the $\mu$ is $\sigma -$ additive. Let $A_i \in \mathcal{M}$ (without loss of generality assume $A_i \cap A_j = \emptyset$ if $i \neq j$) $B_n = \bigcup_{i=1}^n A_i$, and $B = \bigcup\limits_{j=1}^{\infty} A_j$, then for $E \subset \Omega$ we have

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c)\\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}).\end{aligned}$$

and so by induction,

$$\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k). \tag{7.5}$$

Therefore we find that

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c)\\ &= \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap B_n^c)\\ &\geq \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap B^c)\end{aligned}$$

where the last inequality is a consequence of the monotonicity of $\mu^*$ and the fact that $B^c \subset B_n^c$. Letting $n \to \infty$ in this equation shows that

$$\begin{aligned}\mu^*(E) &\geq \sum_{k=1}^{\infty}\mu^*(E\cap A_k)+\mu^*(E\cap B^c)\\ &\geq \mu^*(\cup_k(E\cap A_k))+\mu^*(E\setminus B)\\ &= \mu^*(E\cap B)+\mu^*(E\setminus B)\geq \mu^*(E),\end{aligned}$$

wherein we have used the sub-additivity $\mu^*$ twice. Hence $B\in\mathcal{M}$ and we have shown $\mathcal{M}$ is a $\sigma-$ algebra. Since $\mu^*(E)\geq\mu^*(E\cap B_n)$ we may let $n\to\infty$ in Eq. (7.5) to find

$$\mu^*(E)\geq\sum_{k=1}^{\infty}\mu^*(E\cap A_k).$$

Letting $E=B=\cup A_k$ in this inequality then implies $\mu^*(B)\geq\sum_{k=1}^{\infty}\mu^*(A_k)$ and hence, by the sub-additivity of $\mu^*$, $\mu^*(B)=\sum_{k=1}^{\infty}\mu^*(A_k)$. Therefore, $\mu=\mu^*|_{\mathcal{M}}$ is countably additive on $\mathcal{M}$.

Finally we show $\mu$ is complete. If $N\subset F\in\mathcal{M}$ and $\mu(F)=0=\mu^*(F)$, then $\mu^*(N)=0$ and

$$\mu^*(E)\leq\mu^*(E\cap N)+\mu^*(E\cap N^c)=\mu^*(E\cap N^c)\leq\mu^*(E).$$

which shows that $N\in\mathcal{M}$. ■

## 7.2 Extensions of General Premeasures

In this subsection let $\Omega$ be a set, $\mathcal{A}$ be a subalgebra of $2^\Omega$ and $\mu_0:\mathcal{A}\to[0,\infty]$ be a premeasure on $\mathcal{A}$.

**Theorem 7.3.** *Let $\mathcal{A}\subset 2^\Omega$ be an algebra, $\mu$ be a premeasure on $\mathcal{A}$ and $\mu^*$ be the associated outer measure as defined in Eq. (7.1) with $\rho=\mu$. Let $\mathcal{M}:=\mathcal{M}(\mu^*)\supset\sigma(\mathcal{A})$, then:*

1. *$\mathcal{A}\subset\mathcal{M}(\mu^*)$ and $\mu^*|_{\mathcal{A}}=\mu$.*
2. *$\bar{\mu}=\mu^*|_{\mathcal{M}}$ is a measure on $\mathcal{M}$ which extends $\mu$.*
3. *If $\nu:\mathcal{M}\to[0,\infty]$ is another measure such that $\nu=\mu$ on $\mathcal{A}$ and $B\in\mathcal{M}$, then $\nu(B)\leq\bar{\mu}(B)$ and $\nu(B)=\bar{\mu}(B)$ whenever $\bar{\mu}(B)<\infty$.*
4. *If $\mu$ is $\sigma$-finite on $\mathcal{A}$ then the extension, $\bar{\mu}$, of $\mu$ to $\mathcal{M}$ is unique and moreover $\mathcal{M}=\overline{\sigma(\mathcal{A})}^{\bar{\mu}|_{\sigma(\mathcal{A})}}$.*

**Proof.** Recall from Proposition 6.38 that $\mu$ extends to a countably additive function on $\mathcal{A}_\sigma$ and $\mu^*=\mu$ on $\mathcal{A}$.

1. Let $A\in\mathcal{A}$ and $E\subset\Omega$ such that $\mu^*(E)<\infty$. Given $\varepsilon>0$ choose pairwise disjoint sets, $B_j\in\mathcal{A}$, such that $E\subset B:=\sum_{j=1}^\infty B_j$ and
$$\mu^*(E)+\varepsilon\geq\mu(B)=\sum_{j=1}^{\infty}\mu(B_j).$$
Since $A\cap E\subset\sum_{j=1}^\infty(B_j\cap A^c)$ and $E\cap A^c\subset\sum_{j=1}^\infty(B_j\cap A^c)$, using the sub-additivity of $\mu^*$ and the additivity of $\mu$ on $\mathcal{A}$ we have,
$$\begin{aligned}\mu^*(E)+\varepsilon\geq\sum_{j=1}^{\infty}\mu(B_j)&=\sum_{j=1}^{\infty}[\mu(B_j\cap A)+\mu(B_j\cap A^c)]\\ &\geq\mu^*(E\cap A)+\mu^*(E\cap A^c).\end{aligned}$$
Since $\varepsilon>0$ is arbitrary this shows that
$$\mu^*(E)\geq\mu^*(E\cap A)+\mu^*(E\cap A^c)$$
and therefore that $A\in\mathcal{M}(\mu^*)$.
2. This is a direct consequence of item 1. and Theorem 7.2.
3. If $A:=\sum_{j=1}^\infty A_j$ with $\{A_j\}_{j=1}^\infty\subset\mathcal{A}$ being a collection of pairwise disjoint sets, then
$$\nu(A)=\sum_{j=1}^{\infty}\nu(A_j)=\sum_{j=1}^{\infty}\mu(A_j)=\mu(A).$$
This shows $\nu=\mu=\bar{\mu}$ on $\mathcal{A}_\sigma$. Consequently, if $B\in\mathcal{M}$, then
$$\begin{aligned}\nu(B)&\leq\inf\{\nu(A):B\subset A\in\mathcal{A}_\sigma\}\\ &=\inf\{\mu(A):B\subset A\in\mathcal{A}_\sigma\}=\mu^*(B)=\bar{\mu}(B).\end{aligned}\tag{7.6}$$
If $\bar{\mu}(B)<\infty$ and $\varepsilon>0$ is given, there exists $A\in\mathcal{A}_\sigma$ such that $B\subset A$ and $\bar{\mu}(A)=\mu(A)\leq\bar{\mu}(B)+\varepsilon$. From Eq. (7.6), this implies
$$\nu(A\setminus B)\leq\bar{\mu}(A\setminus B)\leq\varepsilon.$$
Therefore,
$$\nu(B)\leq\bar{\mu}(B)\leq\bar{\mu}(A)=\nu(A)=\nu(B)+\nu(A\setminus B)\leq\nu(B)+\varepsilon$$
which shows $\bar{\mu}(B)=\nu(B)$ because $\varepsilon>0$ was arbitrary.
4. For the $\sigma-$ finite case, choose $\Omega_j\in\mathcal{M}$ such that $\Omega_j\uparrow\Omega$ and $\bar{\mu}(\Omega_j)<\infty$ then
$$\bar{\mu}(B)=\lim_{j\to\infty}\bar{\mu}(B\cap\Omega_j)=\lim_{j\to\infty}\nu(B\cap\Omega_j)=\nu(B).$$

■

**Theorem 7.4 (Regularity Theorem).** *Suppose that $\mu$ is a $\sigma$ – finite premeasure on an algebra $\mathcal{A}$, $\bar{\mu}$ is the extension described in Theorem 7.3 and $B \in \mathcal{M} := \mathcal{M}(\mu^*)$. Then:*

1. $$\bar{\mu}(B) := \inf\{\bar{\mu}(C) : B \subset C \in \mathcal{A}_\sigma\}.$$
2. *For any $\varepsilon > 0$ there exists $A \subset B \subset C$ such that $A \in \mathcal{A}_\delta$, $C \in \mathcal{A}_\sigma$ and $\bar{\mu}(C \setminus A) < \varepsilon$.*
3. *There exists $A \subset B \subset C$ such that $A \in \mathcal{A}_{\delta\sigma}$, $C \in \mathcal{A}_{\sigma\delta}$ and $\bar{\mu}(C \setminus A) = 0$.*
4. *The $\sigma$-algebra, $\mathcal{M}$, is the completion of $\sigma(\mathcal{A})$ with respect to $\bar{\mu}|_{\sigma(\mathcal{A})}$.*

**Proof.** The proofs of items 1. – 3. are the same as the proofs of the corresponding results in Theorem 6.24, Corollary 6.26, and Theorem 6.52 and so will be omitted. Moreover, item 4. is a simple consequence of item 3. and Proposition 6.77. ∎

The following proposition shows that measures may be "restricted" to non-measurable sets.

**Proposition 7.5.** *Suppose that $(\Omega, \mathcal{M}, \mu)$ is a probability space and $\Omega_0 \subset \Omega$ is any set. Let $\mathcal{M}_{\Omega_0} := \{A \cap \Omega_0 : A \in \mathcal{M}\}$ and set $P(A \cap \Omega_0) := \mu^*(A \cap \Omega_0)$. Then $P$ is a measure on the $\sigma$ - algebra $\mathcal{M}_{\Omega_0}$. Moreover, if $P^*$ is the outer measure generated by $P$, then $P^*(A) = \mu^*(A)$ for all $A \subset \Omega_0$.*

**Proof.** Let $A, B \in \mathcal{M}$ such that $A \cap B = \emptyset$. Then since $A \in \mathcal{M} \subset \mathcal{M}(\mu^*)$ it follows from Eq. (7.1) with $E := (A \cup B) \cap \Omega_0$ that

$$\begin{aligned}\mu^*((A \cup B) \cap \Omega_0) &= \mu^*((A \cup B) \cap \Omega_0 \cap A) + \mu^*((A \cup B) \cap \Omega_0 \cap A^c) \\ &= \mu^*(\Omega_0 \cap A) + \mu^*(B \cap \Omega_0)\end{aligned}$$

which shows that $P$ is finitely additive. Now suppose $A = \sum_{j=1}^\infty A_j$ with $A_j \in \mathcal{M}$ and let $B_n := \sum_{j=n+1}^\infty A_j \in \mathcal{M}$. By what we have just proved,

$$\mu^*(A \cap \Omega_0) = \sum_{j=1}^n \mu^*(A_j \cap \Omega_0) + \mu^*(B_n \cap \Omega_0) \geq \sum_{j=1}^n \mu^*(A_j \cap \Omega_0).$$

Passing to the limit as $n \to \infty$ in this last expression and using the sub-additivity of $\mu^*$ we find

$$\sum_{j=1}^\infty \mu^*(A_j \cap \Omega_0) \geq \mu^*(A \cap \Omega_0) \geq \sum_{j=1}^\infty \mu^*(A_j \cap \Omega_0).$$

Thus

$$\mu^*(A \cap \Omega_0) = \sum_{j=1}^\infty \mu^*(A_j \cap \Omega_0)$$

and we have shown that $P = \mu^*|_{\mathcal{M}_{\Omega_0}}$ is a measure. Now let $P^*$ be the outer measure generated by $P$. For $A \subset \Omega_0$, we have

$$\begin{aligned}P^*(A) &= \inf\{P(B) : A \subset B \in \mathcal{M}_{\Omega_0}\} \\ &= \inf\{P(B \cap \Omega_0) : A \subset B \in \mathcal{M}\} \\ &= \inf\{\mu^*(B \cap \Omega_0) : A \subset B \in \mathcal{M}\}\end{aligned} \tag{7.7}$$

and since $\mu^*(B \cap \Omega_0) \leq \mu^*(B)$,

$$\begin{aligned}P^*(A) &\leq \inf\{\mu^*(B) : A \subset B \in \mathcal{M}\} \\ &= \inf\{\mu(B) : A \subset B \in \mathcal{M}\} = \mu^*(A).\end{aligned}$$

On the other hand, for $A \subset B \in \mathcal{M}$, we have $\mu^*(A) \leq \mu^*(B \cap \Omega_0)$ and therefore by Eq. (7.7)

$$\mu^*(A) \leq \inf\{\mu^*(B \cap \Omega_0) : A \subset B \in \mathcal{M}\} = P^*(A).$$

and we have shown

$$\mu^*(A) \leq P^*(A) \leq \mu^*(A).$$

∎

## 7.3 More Motivation of Carathéodory's Construction Theorem 7.2

The next Proposition helps to motivate this definition and the Carathéodory's construction Theorem 7.2.

**Proposition 7.6.** *Suppose $\mathcal{E} = \mathcal{M}$ is a $\sigma$ – algebra, $\rho = \mu : \mathcal{M} \to [0, \infty]$ is a measure and $\mu^*$ is defined as in Eq. (6.2). Then*

1. *For $A \subset X$*
$$\mu^*(A) = \inf\{\mu(B) : B \in \mathcal{M} \text{ and } A \subset B\}.$$
*In particular, $\mu^* = \mu$ on $\mathcal{M}$.*
2. *Then $\mathcal{M} \subset \mathcal{M}(\mu^*)$, i.e. if $A \in \mathcal{M}$ and $E \subset X$ then*
$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c). \tag{7.8}$$
3. *Assume further that $\mu$ is $\sigma$ – finite on $\mathcal{M}$, then $\mathcal{M}(\mu^*) = \bar{\mathcal{M}} = \bar{\mathcal{M}}^\mu$ and $\mu^*|_{\mathcal{M}(\mu^*)} = \bar{\mu}$ where $(\bar{\mathcal{M}} = \bar{\mathcal{M}}^\mu, \bar{\mu})$ is the completion of $(\mathcal{M}, \mu)$.*

**Proof.** Item 1. If $E_i \in \mathcal{M}$ such that $A \subset \cup E_i = B$ and $\widetilde{E}_i = E_i \setminus (E_1 \cup \cdots \cup E_{i-1})$ then

$$\sum \mu(E_i) \geq \sum \mu(\widetilde{E}_i) = \mu(B)$$

so

$$\mu^*(A) \leq \sum \mu(\widetilde{E}_i) = \mu(B) \leq \sum \mu(E_i).$$

Therefore, $\mu^*(A) = \inf\{\mu(B) : B \in \mathcal{M} \text{ and } A \subset B\}$.

Item 2. If $\mu^*(E) = \infty$ Eq. (7.8) holds trivially. So assume that $\mu^*(E) < \infty$. Let $\varepsilon > 0$ be given and choose, by Item 1., $B \in \mathcal{M}$ such that $E \subset B$ and $\mu(B) \leq \mu^*(E) + \varepsilon$. Then

$$\begin{aligned}\mu^*(E) + \varepsilon \geq \mu(B) &= \mu(B \cap A) + \mu(B \cap A^c) \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap A^c).\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary we are done.

Item 3. Let us begin by assuming the $\mu(X) < \infty$. We have already seen that $\mathcal{M} \subset \mathcal{M}(\mu^*)$. Suppose that $A \in 2^X$ satisfies,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \, \forall \, E \in 2^X. \tag{7.9}$$

By Item 1., there exists $B_n \in \mathcal{M}$ such that $A \subset B_n$ and $\mu^*(B_n) \leq \mu^*(A) + \frac{1}{n}$ for all $n \in \mathbb{N}$. Therefore $B = \cap B_n \supset A$ and $\mu(B) \leq \mu^*(A) + \frac{1}{n}$ for all $n$ which implies that $\mu(B) \leq \mu^*(A)$ which implies that $\mu(B) = \mu^*(A)$. Similarly there exists $C \in \mathcal{M}$ such that $A^c \subset C$ and $\mu^*(A^c) = \mu(C)$. Taking $E = X$ in Eq. (7.9) shows

$$\mu(X) = \mu^*(A) + \mu^*(A^c) = \mu(B) + \mu(C)$$

so

$$\mu(C^c) = \mu(X) - \mu(C) = \mu(B).$$

Thus letting $D = C^c$, we have

$$D \subset A \subset B \text{ and } \mu(D) = \mu^*(A) = \mu(B)$$

so $\mu(B \setminus D) = 0$ and hence

$$A = D \cup [(B \backslash D) \cap A]$$

where $D \in \mathcal{M}$ and $(B \backslash D) \cap A \in \mathcal{N}$ showing that $A \in \bar{\mathcal{M}}$ and $\mu^*(A) = \bar{\mu}(A)$.

Now if $\mu$ is $\sigma-$finite, choose $X_n \in \mathcal{M}$ such that $\mu(X_n) < \infty$ and $X_n \uparrow X$. Given $A \in \mathcal{M}(\mu^*)$ set $A_n = X_n \cap A$. Therefore

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \; \forall \; E \in 2^X.$$

Replace $E$ by $X_n$ to learn,

$$\mu^*(X_n) = \mu^*(A_n) + \mu^*(X_n \setminus A) = \mu^*(A_n) + \mu^*(X_n \setminus A_n).$$

The same argument as above produces sets $D_n \subset A_n \subset B_n$ such that $\mu(D_n) = \mu^*(A_n) = \mu(B_n)$. Hence $A_n = D_n \cup N_n$ and $N_n := (B_n \setminus D_n) \cap A_n \in \mathcal{N}$. So we learn that

$$A = D \cup N := (\cup D_n) \cup (\cup N_n) \in \mathcal{M} \cup \mathcal{N} = \bar{\mathcal{M}}.$$

We also see that $\mu^*(A) = \mu(D)$ since $D \subset A \subset D \cup F$ where $F \in \mathcal{M}$ such that $N \subset F$ and

$$\mu(D) = \mu^*(D) \leq \mu^*(A) \leq \mu(D \cup F) = \mu(D).$$

■

# Random Variables

**Notation 8.1** *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E) | E \in \mathcal{E}\}.$$

*If $\mathcal{G} \subset 2^X$, let*

$$f_*\mathcal{G} := \{A \in 2^Y | f^{-1}(A) \in \mathcal{G}\}.$$

**Definition 8.2.** *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

The following results will be used frequently (often without further reference) in the sequel.

**Lemma 8.3 (A key measurability lemma).** *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$, then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})). \tag{8.1}$$

*In particular, if $A \subset Y$ then*

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{8.2}$$

*(Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.)*

**Proof.** Since $\mathcal{E} \subset \sigma(\mathcal{E})$, it follows that $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. Moreover, by Exercise 8.1 below, $f^{-1}(\sigma(\mathcal{E}))$ is a $\sigma$ – algebra and therefore,

$$\sigma(f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma(\mathcal{E})).$$

To finish the proof we must show $f^{-1}(\sigma(\mathcal{E})) \subset \sigma(f^{-1}(\mathcal{E}))$, i.e. that $f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))$ for all $B \in \sigma(\mathcal{E})$. To do this we follow the usual measure theoretic mantra, namely let

$$\mathcal{M} := \left\{B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))\right\} = f_*\sigma(f^{-1}(\mathcal{E})).$$

We will now finish the proof by showing $\sigma(\mathcal{E}) \subset \mathcal{M}$. This is easily achieved by observing that $\mathcal{M}$ is a $\sigma$ – algebra (see Exercise 8.1) which contains $\mathcal{E}$ and therefore $\sigma(\mathcal{E}) \subset \mathcal{M}$.

Equation (8.2) is a special case of Eq. (8.1). Indeed, $f = i_A : A \to X$ we have

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

∎

**Exercise 8.1.** If $f : X \to Y$ is a function and $\mathcal{F} \subset 2^Y$ and $\mathcal{B} \subset 2^X$ are $\sigma$ – algebras (algebras), then $f^{-1}\mathcal{F}$ and $f_*\mathcal{B}$ are $\sigma$ – algebras (algebras).

*Example 8.4.* Let $\mathcal{E} = \{(a,b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel $\sigma$ – field on $\mathbb{R}$. Then

$$\mathcal{E}_{(0,1]} = \{(a,b] : 0 \le a < b \le 1\}$$

and we have

$$\mathcal{B}_{(0,1]} = \sigma\left(\mathcal{E}_{(0,1]}\right).$$

In particular, if $A \in \mathcal{B}$ such that $A \subset (0,1]$, then $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

## 8.1 Measurable Functions

**Definition 8.5.** *A* ***measurable space*** *is a pair $(X, \mathcal{M})$, where $X$ is a set and $\mathcal{M}$ is a $\sigma$ – algebra on $X$.*

To motivate the notion of a measurable function, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$ is a function. Roughly speaking, we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i \mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a,b]) \in \mathcal{M}$ for all $a < b$. Because of Corollary 8.11 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

**Definition 8.6.** *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *of more precisely, $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable, if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$, i.e. if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{F}$.*

*Remark 8.7.* Let $f : X \to Y$ be a function. Given a $\sigma$ – algebra $\mathcal{F} \subset 2^Y$, the $\sigma$ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest $\sigma$ – algebra on $X$ such that $f$ is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a $\sigma$ - algebra on $X$ then

$$\mathcal{F} = f_*\mathcal{M} = \{A \in 2^Y | f^{-1}(A) \in \mathcal{M}\}$$

is the largest $\sigma$ – algebra on $Y$ such that $f$ is $(\mathcal{M}, \mathcal{F})$ - measurable.

*Example 8.8 (Indicator Functions).* Let $(X,\mathcal{M})$ be a measurable space and $A \subset X$. Then $1_A$ is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{M}$. Indeed, $1_A^{-1}(W)$ is either $\emptyset$, $X$, $A$ or $A^c$ for any $W \subset \mathbb{R}$ with $1_A^{-1}(\{1\}) = A$.

*Example 8.9.* Suppose $f : X \to Y$ with $Y$ being a finite or countable set and $\mathcal{F} = 2^Y$. Then $f$ is measurable iff $f^{-1}(\{y\}) \in \mathcal{M}$ for all $y \in Y$.

**Proposition 8.10.** *Suppose that $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ are measurable spaces and further assume $\mathcal{E} \subset \mathcal{F}$ generates $\mathcal{F}$, i.e. $\mathcal{F} = \sigma(\mathcal{E})$. Then a map, $f : X \to Y$ is measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.*

**Proof.** If $f$ is $\mathcal{M}/\mathcal{F}$ measurable, then $f^{-1}(\mathcal{E}) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}$. Conversely if $f^{-1}(\mathcal{E}) \subset \mathcal{M}$ then $\sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}$ and so making use of Lemma 8.3,

$$f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}.$$

■

**Corollary 8.11.** *Suppose that $(X,\mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *$f$ is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable,*
2. *$f^{-1}((a,\infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a,\infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty,a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

**Exercise 8.2 (Look at but do not hand in).** Prove Corollary 8.11. **Hint:** See Exercise 4.9.

**Exercise 8.3.** If $\mathcal{M}$ is the $\sigma$ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the $\sigma$ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

**Exercise 8.4.** Let $(X,\mathcal{M})$ be a measure space and $f_n : X \to \mathbb{R}$ be a sequence of measurable functions on $X$. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{R}\} \in \mathcal{M}$. Similarly show the same holds if $\mathbb{R}$ is replaced by $\mathbb{C}$.

**Exercise 8.5.** Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_{\mathbb{R}},\mathcal{B}_{\mathbb{R}})$ – measurable.

**Definition 8.12.** *Given measurable spaces $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff $f$ is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

**Proposition 8.13 (Localizing Measurability).** *Let $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If $f$ is measurable and $A \subset X$ then $f|_A : A \to Y$ is $\mathcal{M}_A/\mathcal{F}$ – measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}/\mathcal{F}$ – measurable for all $n$, then $f$ is $\mathcal{M}$ – measurable.*

**Proof.** 1. If $f : X \to Y$ is measurable, $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{F}$ and therefore

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A \text{ for all } B \in \mathcal{F}.$$

2. If $B \in \mathcal{F}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty \left(f^{-1}(B) \cap A_n\right) = \cup_{n=1}^\infty f|_{A_n}^{-1}(B).$$

Since each $A_n \in \mathcal{M}$, $\mathcal{M}_{A_n} \subset \mathcal{M}$ and so the previous displayed equation shows $f^{-1}(B) \in \mathcal{M}$. ■

**Lemma 8.14 (Composing Measurable Functions).** *Suppose that $(X,\mathcal{M})$, $(Y,\mathcal{F})$ and $(Z,\mathcal{G})$ are measurable spaces. If $f : (X,\mathcal{M}) \to (Y,\mathcal{F})$ and $g : (Y,\mathcal{F}) \to (Z,\mathcal{G})$ are measurable functions then $g \circ f : (X,\mathcal{M}) \to (Z,\mathcal{G})$ is measurable as well.*

**Proof.** By assumption $g^{-1}(\mathcal{G}) \subset \mathcal{F}$ and $f^{-1}(\mathcal{F}) \subset \mathcal{M}$ so that

$$(g \circ f)^{-1}(\mathcal{G}) = f^{-1}\left(g^{-1}(\mathcal{G})\right) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}.$$

■

**Definition 8.15 ($\sigma$ – Algebras Generated by Functions).** *Let $X$ be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha,\mathcal{F}_\alpha) : \alpha \in I\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$. Let $\sigma(f_\alpha : \alpha \in I)$ denote the smallest $\sigma$ – algebra on $X$ such that each $f_\alpha$ is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

*Example 8.16.* Suppose that $Y$ is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1,\dots,y_N) = y_i$. Then, as the reader should check,

$$\sigma(\pi_1,\dots,\pi_n) = \left\{A \times \Lambda^{N-n} : A \subset \Lambda^n\right\}.$$

**Proposition 8.17.** *Assuming the notation in Definition 8.15 (so $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$) and additionally let $(Z,\mathcal{M})$ be a measurable space. Then $g : Z \to X$ is $(\mathcal{M},\sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ g$ $\left(Z \xrightarrow{g} X \xrightarrow{f_\alpha} Y_\alpha\right)$ is $(\mathcal{M},\mathcal{F}_\alpha)$–measurable for all $\alpha \in I$.*

**Proof.** ($\Rightarrow$) If $g$ is $(\mathcal{M},\sigma(f_\alpha : \alpha \in I))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M},\mathcal{F}_\alpha)$ – measurable by Lemma 8.14.

($\Leftarrow$) Since $\sigma(f_\alpha : \alpha \in I) = \sigma(\mathcal{E})$ where $\mathcal{E} := \cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)$, according to Proposition 8.10, it suffices to show $g^{-1}(A) \in \mathcal{M}$ for $A \in f_\alpha^{-1}(\mathcal{F}_\alpha)$. But this is true since if $A = f_\alpha^{-1}(B)$ for some $B \in \mathcal{F}_\alpha$, then $g^{-1}(A) = g^{-1}\left(f_\alpha^{-1}(B)\right) = (f_\alpha \circ g)^{-1}(B) \in \mathcal{M}$ because $f_\alpha \circ g : Z \to Y_\alpha$ is assumed to be measurable. ■

**Definition 8.18.** *If $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, then the product measure space, $(Y, \mathcal{F})$, is $Y := \prod_{\alpha \in I} Y_\alpha$, $\mathcal{F} := \sigma(\pi_\alpha : \alpha \in I)$ where $\pi_\alpha : Y \to Y_\alpha$ is the $\alpha$ – component projection. We call $\mathcal{F}$ the product $\sigma$ – algebra and denote it by, $\mathcal{F} = \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

Let us record an important special case of Proposition 8.17.

**Corollary 8.19.** *If $(Z, \mathcal{M})$ is a measure space, then $g : Z \to Y = \prod_{\alpha \in I} Y_\alpha$ is $(\mathcal{M}, \mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha)$ – measurable iff $\pi_\alpha \circ g : Z \to Y_\alpha$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$.*

As a special case of the above corollary, if $A = \{1, 2, \dots, n\}$, then $Y = Y_1 \times \cdots \times Y_n$ and $g = (g_1, \dots, g_n) : Z \to Y$ is measurable iff each component, $g_i : Z \to Y_i$, is measurable. Here is another closely related result.

**Proposition 8.20.** *Suppose $X$ is a set, $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, and we are given maps, $f_\alpha : X \to Y_\alpha$, for all $\alpha \in I$. If $f : X \to Y := \prod_{\alpha \in I} Y_\alpha$ is the unique map, such that $\pi_\alpha \circ f = f_\alpha$, then*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(f) = f^{-1}(\mathcal{F})$$

*where $\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

**Proof.** Since $\pi_\alpha \circ f = f_\alpha$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ it follows from Corollary 8.19 that $f : X \to Y$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}$ – measurable. Since $\sigma(f)$ is the smallest $\sigma$ – algebra on $X$ such that $f$ is measurable we may conclude that $\sigma(f) \subset \sigma(f_\alpha : \alpha \in I)$.

Conversely, for each $\alpha \in I$, $f_\alpha = \pi_\alpha \circ f$ is $\sigma(f) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ being the composition of two measurable functions. Since $\sigma(f_\alpha : \alpha \in I)$ is the smallest $\sigma$ – algebra on $X$ such that each $f_\alpha : X \to Y_\alpha$ is measurable, we learn that $\sigma(f_\alpha : \alpha \in I) \subset \sigma(f)$. ■

**Exercise 8.6.** Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\mathcal{E}_i$ is a subset of $\mathcal{F}_i$ such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for $i = 1$ and 2. Show $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{E})$ where $\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{E}_i \text{ for } i = 1, 2\}$. **Hints:**

1. First show that if $Y$ is a set and $\mathcal{S}_1$ and $\mathcal{S}_2$ are two non-empty subsets of $2^Y$, then $\sigma(\sigma(\mathcal{S}_1) \cup \sigma(\mathcal{S}_2)) = \sigma(\mathcal{S}_1 \cup \mathcal{S}_2)$. (In fact, one has that $\sigma(\cup_{\alpha \in I} \sigma(\mathcal{S}_\alpha)) = \sigma(\cup_{\alpha \in I} \mathcal{S}_\alpha)$ for any collection of non-empty subsets, $\{\mathcal{S}_\alpha\}_{\alpha \in I} \subset 2^Y$.)
2. After this you might start your proof as follows;

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\left(\pi_1^{-1}(\mathcal{F}_1) \cup \pi_2^{-1}(\mathcal{F}_2)\right) = \sigma\left(\pi_1^{-1}(\sigma(\mathcal{E}_2)) \cup \pi_2^{-1}(\sigma(\mathcal{E}_2))\right) = \dots.$$

*Remark 8.21.* The reader should convince herself that Exercise 8.6 admits the following extension. If $I$ is any finite or countable index set, $\{(Y_i, \mathcal{F}_i)\}_{i \in I}$ are measurable spaces and $\mathcal{E}_i \subset \mathcal{F}_i$ are such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for all $i \in I$, then

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j \text{ for all } j \in I\right\}\right)$$

and in particular,

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{F}_j \text{ for all } j \in I\right\}\right).$$

The last fact is easily verified directly without the aid of Exercise 8.6.

**Exercise 8.7 (Look at but do not hand in).** Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\emptyset \neq B_i \subset Y_i$ for $i = 1, 2$. Show

$$[\mathcal{F}_1 \otimes \mathcal{F}_2]_{B_1 \times B_2} = [\mathcal{F}_1]_{B_1} \otimes [\mathcal{F}_2]_{B_2}.$$

**Hint:** you may find it useful to use the result of Exercise 8.6 with

$$\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{F}_i \text{ for } i = 1, 2\}.$$

**Definition 8.22.** *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

**Proposition 8.23.** *Let $X$ and $Y$ be two topological spaces and $f : X \to Y$ be a continuous function. Then $f$ is Borel measurable.*

**Proof.** Using Lemma 8.3 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

■

*Example 8.24.* For $i = 1, 2, \dots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each $\pi_i$ is continuous and therefore $\mathcal{B}_{\mathbb{R}^n} / \mathcal{B}_{\mathbb{R}}$ – measurable.

**Lemma 8.25.** *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \dots, \pi_n) = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ and in particular, $A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \dots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the $\sigma$ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$. (Also see Remark 8.21.)*

**Proof. Assertion 1.** Since $\mathcal{E} \subset \mathcal{B}_{\mathbb{R}^n}$, it follows that $\sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}$. Let

$$\mathcal{E}_0 := \{(a,b) : a,b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a,b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1,2,\dots,n$ and let

$$(a,b) = (a_1,b_1) \times \dots \times (a_n,b_n). \quad (8.3)$$

Since every open set, $V \subset \mathbb{R}^n$, may be written as a (necessarily) countable union of elements from $\mathcal{E}_0$, we have

$$V \in \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}),$$

i.e. $\sigma(\mathcal{E}_0)$ and hence $\sigma(\mathcal{E})$ contains all open subsets of $\mathbb{R}^n$. Hence we may conclude that

$$\mathcal{B}_{\mathbb{R}^n} = \sigma(\text{open sets}) \subset \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}.$$

**Assertion 2.** Since each $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous, it is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore, $\sigma(\pi_1,\dots,\pi_n) \subset \mathcal{B}_{\mathbb{R}^n}$. Moreover, if $(a,b)$ is as in Eq. (8.3), then

$$(a,b) = \cap_{i=1}^n \pi_i^{-1}((a_i,b_i)) \in \sigma(\pi_1,\dots,\pi_n).$$

Therefore, $\mathcal{E} \subset \sigma(\pi_1,\dots,\pi_n)$ and $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E}) \subset \sigma(\pi_1,\dots,\pi_n)$.

**Assertion 3.** If $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1,2,\dots,n$, then

$$A_1 \times \dots \times A_n = \cap_{i=1}^n \pi_i^{-1}(A_i) \in \sigma(\pi_1,\dots,\pi_n) = \mathcal{B}_{\mathbb{R}^n}.$$

■

**Corollary 8.26.** *If $(X,\mathcal{M})$ is a measurable space, then*

$$f = (f_1,f_2,\dots,f_n) : X \to \mathbb{R}^n$$

*is $(\mathcal{M},\mathcal{B}_{\mathbb{R}^n})$ – measurable iff $f_i : X \to \mathbb{R}$ is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable for each $i$. In particular, a function $f : X \to \mathbb{C}$ is $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable.*

**Proof.** This is an application of Lemma 8.25 and Corollary 8.19 with $Y_i = \mathbb{R}$ for each $i$. ■

**Corollary 8.27.** *Let $(X,\mathcal{M})$ be a measurable space and $f,g : X \to \mathbb{C}$ be $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable.*

**Proof.** Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_{\pm} : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x),g(x))$, $A_{\pm}(w,z) = w \pm z$ and $M(w,z) = wz$. Then $A_{\pm}$ and $M$ are continuous and hence $(\mathcal{B}_{\mathbb{C}^2},\mathcal{B}_{\mathbb{C}})$ – measurable. Also $F$ is $(\mathcal{M},\mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_{\pm} \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ■

**Lemma 8.28.** *Let $\alpha \in \mathbb{C}$, $(X,\mathcal{M})$ be a measurable space and $f : X \to \mathbb{C}$ be a $(\mathcal{M},\mathcal{B}_{\mathbb{C}})$ – measurable function. Then*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if} \quad f(x) \neq 0 \\ \alpha & \text{if} \quad f(x) = 0 \end{cases}$$

*is measurable.*

**Proof.** Define $i : \mathbb{C} \to \mathbb{C}$ by

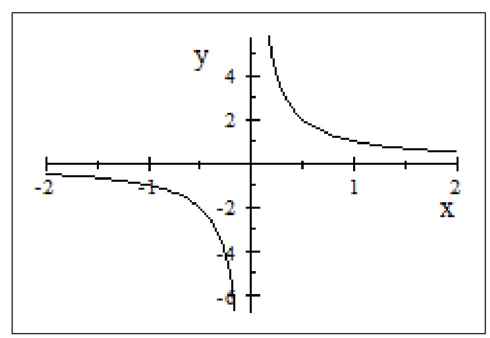
$$i(z) = \begin{cases} \frac{1}{z} & \text{if} \quad z \neq 0 \\ 0 & \text{if} \quad z = 0. \end{cases}$$

For any open set $V \subset \mathbb{C}$ we have

$$i^{-1}(V) = i^{-1}(V \setminus \{0\}) \cup i^{-1}(V \cap \{0\})$$

Because $i$ is continuous except at $z = 0$, $i^{-1}(V \setminus \{0\})$ is an open set and hence in $\mathcal{B}_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in \mathcal{B}_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset \mathcal{B}_{\mathbb{C}}$ and hence $i^{-1}(\mathcal{B}_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) = \sigma(i^{-1}(\tau_{\mathbb{C}})) \subset \mathcal{B}_{\mathbb{C}}$ which shows that $i$ is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, $F$ is also measurable. ■

*Remark 8.29.* For the real case of Lemma 8.28, define $i$ as above but now take $z$ to real. From the plot of $i$, Figure 8.29, the reader may easily verify that $i^{-1}((-\infty,a])$ is an infinite half interval for all $a$ and therefore $i$ is measurable. See Example 8.34 for another proof of this fact.



We will often deal with functions $f : X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. When talking about measurability in this context we will refer to the $\sigma$ – algebra on $\bar{\mathbb{R}}$ defined by

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma\left(\{[a, \infty] : a \in \mathbb{R}\}\right). \tag{8.4}$$

**Proposition 8.30 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$).** *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{8.5}$$

*In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.*

**Proof.** Let us first observe that

$$\{-\infty\} = \cap_{n=1}^{\infty}[-\infty, -n) = \cap_{n=1}^{\infty}[-n, \infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}},$$
$$\{\infty\} = \cap_{n=1}^{\infty}[n, \infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}.$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

$$\begin{aligned} i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) &= \sigma\left(i^{-1}\left(\left\{[a, \infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}\left([a, \infty]\right) : a \in \bar{\mathbb{R}}\right\}\right) \\ &= \sigma\left(\left\{[a, \infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a, \infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}. \end{aligned}$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \Longrightarrow A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \triangle B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (8.5). ■

The proofs of the next two corollaries are left to the reader, see Exercises 8.8 and 8.9.

**Corollary 8.31.** *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \bar{\mathbb{R}}$ be a function. Then the following are equivalent*

1. *$f$ is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ - measurable,*
2. *$f^{-1}((a, \infty]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
4. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^0 : X \to \mathbb{R}$ defined by*

$$f^0(x) := \begin{cases} f(x) & \text{if } \quad f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm\infty\} \end{cases}$$

*is measurable.*

**Corollary 8.32.** *Let $(X, \mathcal{M})$ be a measurable space, $f, g : X \to \bar{\mathbb{R}}$ be functions and define $f \cdot g : X \to \bar{\mathbb{R}}$ and $(f + g) : X \to \bar{\mathbb{R}}$ using the conventions, $0 \cdot \infty = 0$ and $(f + g)(x) = 0$ if $f(x) = \infty$ and $g(x) = -\infty$ or $f(x) = -\infty$ and $g(x) = \infty$. Then $f \cdot g$ and $f + g$ are measurable functions on $X$ if both $f$ and $g$ are measurable.*

**Exercise 8.8.** Prove Corollary 8.31 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 8.11. Use Proposition 8.30 to handle item 4.

**Exercise 8.9.** Prove Corollary 8.32.

**Proposition 8.33 (Closure under sups, infs and limits).** *Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*

$$\sup_j f_j, \quad \inf_j f_j, \quad \limsup_{j\to\infty} f_j \text{ and } \liminf_{j\to\infty} f_j$$

*are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X, \mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)*

**Proof.** Define $g_+(x) := \sup_j f_j(x)$, then

$$\begin{aligned} \{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \ \forall j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M} \end{aligned}$$

so that $g_+$ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then

$$\{x : g_-(x) \ge a\} = \cap_j \{x : f_j(x) \ge a\} \in \mathcal{M}.$$

Since

$$\begin{aligned} \limsup_{j\to\infty} f_j &= \inf_n \sup\{f_j : j \ge n\} \text{ and} \\ \liminf_{j\to\infty} f_j &= \sup_n \inf\{f_j : j \ge n\} \end{aligned}$$

we are done by what we have already proved. ■

*Example 8.34.* As we saw in Remark 8.29, $i : \mathbb{R} \to \mathbb{R}$ defined by

$$i(z) = \begin{cases} \frac{1}{z} & \text{if} \quad z \ne 0 \\ 0 & \text{if} \quad z = 0. \end{cases} \tag{8.6}$$

is measurable by a simple direct argument. For an alternative argument, let

$$i_n(z) := \frac{z}{z^2 + \frac{1}{n}} \text{ for all } n \in \mathbb{N}.$$

Then $i_n$ is continuous and $\lim_{n\to\infty} i_n(z) = i(z)$ for all $z \in \mathbb{R}$ from which it follows that $i$ is Borel measurable.

Similarly we may consider $i$ defined in Eq. (8.6) to be a function from $\mathbb{C}$ to $\mathbb{C}$. Again $i(\cdot)$ is Borel measurable (see also Lemma 8.28) since $i(z) = \lim_{n\to\infty} i_n(z)$ for all $z \in \mathbb{C}$ where,

$$i_n(z) := \frac{\bar{z}}{|z|^2 + \frac{1}{n}} \text{ for all } n \in \mathbb{N}.$$

*Example 8.35.* Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q} \cap [0,1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x - r_n|}} = 5 \text{ if } x = r_n.$$

Then $f : \mathbb{R} \to \bar{\mathbb{R}}$ is measurable. Indeed, if

$$g_n(x) = \begin{cases} \frac{1}{\sqrt{|x - r_n|}} & \text{if } x \neq r_n \\ 0 & \text{if } x = r_n \end{cases}$$

then $g_n(x) = \sqrt{|i(x - r_n)|}$ is measurable as the composition of measurable is measurable. Therefore $g_n + 5 \cdot 1_{\{r_n\}}$ is measurable as well. Finally,

$$f(x) = \lim_{N\to\infty} \sum_{n=1}^N 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

is measurable since sums of measurable functions are measurable and limits of measurable functions are measurable. **Moral:** if you can explicitly write a function $f : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ down then it is going to be measurable.

**Definition 8.36.** *Given a function* $f : X \to \bar{\mathbb{R}}$ *let* $f_+(x) := \max\{f(x), 0\}$ *and* $f_-(x) := \max(-f(x), 0) = -\min(f(x), 0)$. *Notice that* $f = f_+ - f_-$.

**Corollary 8.37.** *Suppose* $(X, \mathcal{M})$ *is a measurable space and* $f : X \to \bar{\mathbb{R}}$ *is a function. Then* $f$ *is measurable iff* $f_\pm$ *are measurable.*

**Proof.** If $f$ is measurable, then Proposition 8.33 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ■

**Definition 8.38.** *Let* $(X, \mathcal{M})$ *be a measurable space. A function* $\varphi : X \to \mathbb{F}$ *($\mathbb{F}$ denotes either $\mathbb{R}$, $\mathbb{C}$ or $[0,\infty] \subset \bar{\mathbb{R}}$) is a* ***simple function*** *if* $\varphi$ *is* $\mathcal{M}$ – $\mathcal{B}_\mathbb{F}$ *measurable and* $\varphi(X)$ *contains only finitely many elements.*

Any such simple functions can be written as

$$\varphi = \sum_{i=1}^n \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{8.7}$$

Indeed, take $\lambda_1, \lambda_2, \ldots, \lambda_n$ to be an enumeration of the range of $\varphi$ and $A_i = \varphi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\varphi = \sum_{y\in\mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{8.8}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

**Theorem 8.39 (Approximation Theorem).** *Let* $f : X \to [0,\infty]$ *be measurable and define, see Figure 8.1,*

$$\begin{aligned} \varphi_n(x) &:= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)}(x) + 2^n 1_{f^{-1}((2^n,\infty])}(x) \\ &= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}}(x) + 2^n 1_{\{f > 2^n\}}(x) \end{aligned}$$

*then* $\varphi_n \le f$ *for all* $n$, $\varphi_n(x) \uparrow f(x)$ *for all* $x \in X$ *and* $\varphi_n \uparrow f$ *uniformly on the sets* $X_M := \{x \in X : f(x) \le M\}$ *with* $M < \infty$.

*Moreover, if* $f : X \to \mathbb{C}$ *is a measurable function, then there exists simple functions* $\varphi_n$ *such that* $\lim_{n\to\infty} \varphi_n(x) = f(x)$ *for all* $x$ *and* $|\varphi_n| \uparrow |f|$ *as* $n \to \infty$.

**Proof.** Since $f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)$ and $f^{-1}((2^n,\infty])$ are in $\mathcal{M}$ as $f$ is measurable, $\varphi_n$ is a measurable simple function for each $n$. Because

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right],$$

if $x \in f^{-1}\left(\left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right]\right)$ then $\varphi_n(x) = \varphi_{n+1}(x) = \frac{2k}{2^{n+1}}$ and if $x \in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right]\right)$ then $\varphi_n(x) = \frac{2k}{2^{n+1}} < \frac{2k+1}{2^{n+1}} = \varphi_{n+1}(x)$. Similarly

$$(2^n, \infty] = (2^n, 2^{n+1}] \cup (2^{n+1}, \infty],$$

and so for $x \in f^{-1}((2^{n+1},\infty])$, $\varphi_n(x) = 2^n < 2^{n+1} = \varphi_{n+1}(x)$ and for $x \in f^{-1}((2^n, 2^{n+1}])$, $\varphi_{n+1}(x) \ge 2^n = \varphi_n(x)$. Therefore $\varphi_n \le \varphi_{n+1}$ for all $n$. It is

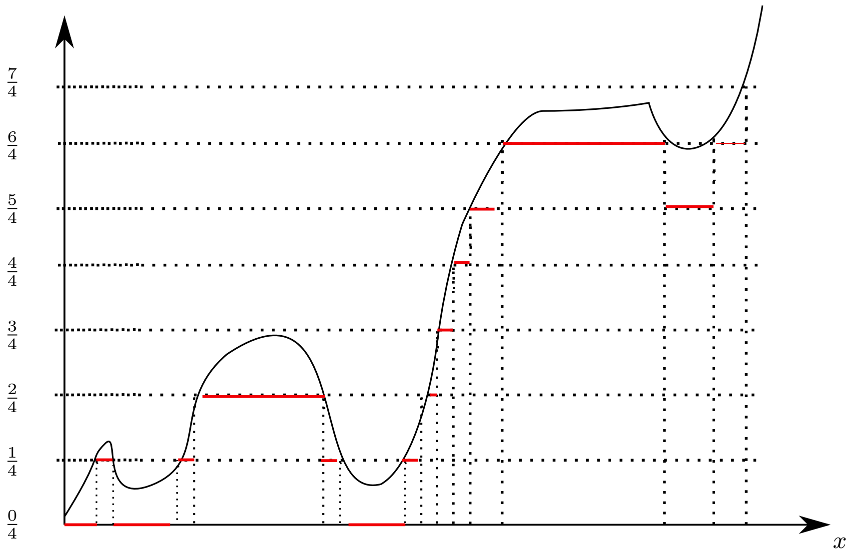



**Fig. 8.1.** Constructing the simple function, $\varphi_2$, approximating a function, $f : X \to [0,\infty]$. The graph of $\varphi_2$ is in red.

clear by construction that $0 \le \varphi_n(x) \le f(x)$ for all $x$ and that $0 \le f(x) - \varphi_n(x) \le 2^{-n}$ if $x \in X_{2^n} = \{f \le 2^n\}$. Hence we have shown that $\varphi_n(x) \uparrow f(x)$ for all $x \in X$ and $\varphi_n \uparrow f$ uniformly on bounded sets.

For the second assertion, first assume that $f : X \to \mathbb{R}$ is a measurable function and choose $\varphi_n^\pm$ to be non-negative simple functions such that $\varphi_n^\pm \uparrow f_\pm$ as $n \to \infty$ and define $\varphi_n = \varphi_n^+ - \varphi_n^-$. Then (using $\varphi_n^+ \cdot \varphi_n^- \le f_+ \cdot f_- = 0$)

$$|\varphi_n| = \varphi_n^+ + \varphi_n^- \le \varphi_{n+1}^+ + \varphi_{n+1}^- = |\varphi_{n+1}|$$

and clearly $|\varphi_n| = \varphi_n^+ + \varphi_n^- \uparrow f_+ + f_- = |f|$ and $\varphi_n = \varphi_n^+ - \varphi_n^- \to f_+ - f_- = f$ as $n \to \infty$. Now suppose that $f : X \to \mathbb{C}$ is measurable. We may now choose simple function $u_n$ and $v_n$ such that $|u_n| \uparrow |\mathrm{Re}\, f|$, $|v_n| \uparrow |\mathrm{Im}\, f|$, $u_n \to \mathrm{Re}\, f$ and $v_n \to \mathrm{Im}\, f$ as $n \to \infty$. Let $\varphi_n = u_n + iv_n$, then

$$|\varphi_n|^2 = u_n^2 + v_n^2 \uparrow |\mathrm{Re}\, f|^2 + |\mathrm{Im}\, f|^2 = |f|^2$$

and $\varphi_n = u_n + iv_n \to \mathrm{Re}\, f + i\,\mathrm{Im}\, f = f$ as $n \to \infty$. ■

## 8.2 Factoring Random Variables

**Lemma 8.40.** *Suppose that* $(\mathbb{Y},\mathcal{F})$ *is a measurable space and* $Y : \Omega \to \mathbb{Y}$ *is a map. Then to every* $(\sigma(Y),\mathcal{B}_{\bar{\mathbb{R}}})$ *– measurable function,* $h : \Omega \to \bar{\mathbb{R}}$*, there is a* $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ *– measurable function* $H : \mathbb{Y} \to \bar{\mathbb{R}}$ *such that* $h = H \circ Y$*. More generally,* $\bar{\mathbb{R}}$ *may be replaced by any "standard Borel space,"*[1] *i.e. a space,* $(S,\mathcal{B}_S)$ *which is measure theoretic isomorphic to a Borel subset of* $\mathbb{R}$*.*

$$\begin{array}{ccc} (\Omega,\sigma(Y)) & \xrightarrow{Y} & (\mathbb{Y},\mathcal{F}) \\ h\downarrow & \swarrow H & \\ (S,\mathcal{B}_S) & & \end{array}$$

**Proof.** First suppose that $h = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $H = 1_B$. More generally if $h = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $h = H \circ Y$ with $H := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $h$, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions $h_n$ converging to $h$. Let $H_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $h_n = H_n \circ Y$. Then it follows that

$$h = \lim_{n\to\infty} h_n = \limsup_{n\to\infty} h_n = \limsup_{n\to\infty} H_n \circ Y = H \circ Y$$

where $H := \limsup_{n\to\infty} H_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$.

For the last assertion we may assume that $S \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_S = (\mathcal{B}_{\mathbb{R}})_S = \{A \cap S : A \in \mathcal{B}_{\mathbb{R}}\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \bar{\mathbb{R}}$ which is $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable such that $h = i_S \circ h = H \circ Y$. The only problems with $H$ is that $H(\mathbb{Y})$ may not be contained in $S$. To fix this, let

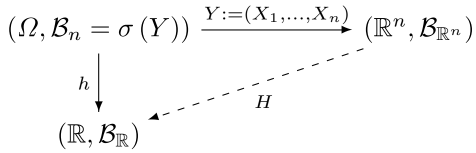
$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on } H^{-1}(S) \\ * & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

where $*$ is some fixed arbitrary point in $S$. It follows from Proposition 8.13 that $H_S : \mathbb{Y} \to S$ is $(\mathcal{F},\mathcal{B}_S)$ – measurable and we still have $h = H_S \circ Y$ as the range of $Y$ must necessarily be in $H^{-1}(S)$. ■

Here is how this lemma will often be used in these notes.

**Corollary 8.41.** *Suppose that* $(\Omega,\mathcal{B})$ *is a measurable space,* $X_n : \Omega \to \mathbb{R}$ *are* $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ *– measurable functions, and* $\mathcal{B}_n := \sigma(X_1,\dots,X_n) \subset \mathcal{B}$ *for each* $n \in \mathbb{N}$*. Then* $h : \Omega \to \mathbb{R}$ *is* $\mathcal{B}_n$ *– measurable iff there exists* $H : \mathbb{R}^n \to \mathbb{R}$ *which is* $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ *– measurable such that* $h = H(X_1,\dots,X_n)$*.*

[1] Standard Borel spaces include almost any measurable space that we will consider in these notes. For example they include all complete seperable metric spaces equipped with the Borel $\sigma$ – algebra, see Section 11.11.




**Proof.** By Lemma 8.25 and Corollary 8.19, the map, $Y := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ is $(\mathcal{B}, \mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}})$ – measurable and by Proposition 8.20, $\mathcal{B}_n = \sigma(X_1, \dots, X_n) = \sigma(Y)$. Thus we may apply Lemma 8.40 to see that there exists a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$, such that $h = H \circ Y = H(X_1, \dots, X_n)$. ■

## 8.3 Summary of Measurability Statements

It may be worthwhile to gather the statements of the main measurability results of Sections 8.1 and 8.2 in one place. To do this let $(\Omega, \mathcal{B})$, $(X, \mathcal{M})$, and $\{(Y_\alpha, \mathcal{F}_\alpha)\}_{\alpha \in I}$ be measurable spaces and $f_\alpha : \Omega \to Y_\alpha$ be given maps for all $\alpha \in I$. Also let $\pi_\alpha : Y \to Y_\alpha$ be the $\alpha$ – projection map,

$$\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha := \sigma(\pi_\alpha : \alpha \in I)$$

be the product $\sigma$ – algebra on $Y$, and $f : \Omega \to Y$ be the unique map determined by $\pi_\alpha \circ f = f_\alpha$ for all $\alpha \in I$. Then the following measurability results hold;

1. For $A \subset \Omega$, the indicator function, $1_A$, is $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{B}$. (Example 8.8).
2. If $\mathcal{E} \subset \mathcal{M}$ generates $\mathcal{M}$ (i.e. $\mathcal{M} = \sigma(\mathcal{E})$), then a map, $g : \Omega \to X$ is $(\mathcal{B}, \mathcal{M})$ – measurable iff $g^{-1}(\mathcal{E}) \subset \mathcal{B}$ (Lemma 8.3 and Proposition 8.10).
3. The notion of measurability may be localized (Proposition 8.13).
4. Composition of measurable functions are measurable (Lemma 8.14).
5. Continuous functions between two topological spaces are also Borel measurable (Proposition 8.23).
6. $\sigma(f) = \sigma(f_\alpha : \alpha \in I)$ (Proposition 8.20).
7. A map, $h : X \to \Omega$ is $(\mathcal{M}, \sigma(f) = \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Proposition 8.17).
8. A map, $h : X \to Y$ is $(\mathcal{M}, \mathcal{F})$ – measurable iff $\pi_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Corollary 8.19).
9. If $I = \{1, 2, \dots, n\}$, then
$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i \text{ for } i \in I\}),$$
this is a special case of Remark 8.21.
10. $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ ($n$ - times) for all $n \in \mathbb{N}$, i.e. the Borel $\sigma$ – algebra on $\mathbb{R}^n$ is the same as the product $\sigma$ – algebra. (Lemma 8.25).
11. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ is closed under the usual pointwise algebraic operations (Corollary 8.32). They are also closed under the countable supremums, infimums, and limits (Proposition 8.33).
12. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ is closed under the usual pointwise algebraic operations and countable limits. (Corollary 8.27 and Proposition 8.33). The limiting assertion follows by considering the real and imaginary parts of all functions involved.
13. The class of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ may be well approximated by measurable simple functions (Theorem 8.39).

14. If $X_i : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_\mathbb{R}$ – measurable maps and $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff $h = H(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_\mathbb{R}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$ (Corollary 8.41).
15. We also have the more general factorization Lemma 8.40.

For the most part most of our future measurability issues can be resolved by one or more of the items on this list.



## 8.4 Distributions / Laws of Random Vectors

The proof of the following proposition is routine and will be left to the reader.

**Proposition 8.42.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$. Then $\nu$ is a measure on $(Y, \mathcal{F})$. (In the future we will denote $\nu$ by $f_*\mu$ or $\mu \circ f^{-1}$ or $\mathrm{Law}_\mu(f)$ and call $f_*\mu$ the **push-forward of** $\mu$ **by** $f$ or the **law of** $f$ **under** $\mu$.*

**Definition 8.43.** *Suppose that $\{X_i\}_{i=1}^n$ is a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$. The probability measure,*

$$\mu = (X_1, \dots, X_n)_* P = P \circ (X_1, \dots, X_n)^{-1} \text{ on } \mathcal{B}_\mathbb{R}$$

*(see Proposition 8.42) is called the **joint distribution** (or **law**) of $(X_1, \dots, X_n)$. To be more explicit,*

$$\mu(B) := P((X_1, \dots, X_n) \in B) := P(\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\})$$

*for all $B \in \mathcal{B}_{\mathbb{R}^n}$.*

**Corollary 8.44.** *The joint distribution, $\mu$ is uniquely determined from the knowledge of*

$$P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) \text{ for all } A_i \in \mathcal{B}_\mathbb{R}$$

*or from the knowledge of*

$$P(X_1 \le x_1, \dots, X_n \le x_n) \text{ for all } A_i \in \mathcal{B}_\mathbb{R}$$

*for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.*

**Proof.** Apply Proposition 6.18 with $\mathcal{P}$ being the $\pi$ – systems defined by

$$\mathcal{P} := \{A_1 \times \dots \times A_n \in \mathcal{B}_{\mathbb{R}^n} : A_i \in \mathcal{B}_\mathbb{R}\}$$

for the first case and

$$\mathcal{P} := \{(-\infty, x_1] \times \dots \times (-\infty, x_n] \in \mathcal{B}_{\mathbb{R}^n} : x_i \in \mathbb{R}\}$$

for the second case. ∎

**Definition 8.45.** *Suppose that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two finite sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ respectively. We write $(X_1, \dots, X_n) \overset{d}{=} (Y_1, \dots, Y_n)$ if $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the **same distribution / law**, i.e. if*

$$P((X_1,\dots,X_n)\in B)=P'((Y_1,\dots,Y_n)\in B) \text{ for all } B\in\mathcal{B}_{\mathbb{R}^n}.$$

*More generally, if* $\{X_i\}_{i=1}^\infty$ *and* $\{Y_i\}_{i=1}^\infty$ *are two sequences of random variables on two probability spaces,* $(\Omega,\mathcal{B},P)$ *and* $(\Omega',\mathcal{B}',P')$ *we write* $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ *iff* $(X_1,\dots,X_n)\overset{d}{=}(Y_1,\dots,Y_n)$ *for all* $n\in\mathbb{N}$.

**Proposition 8.46.** *Let us continue using the notation in Definition 8.45. Further let*

$$X=(X_1,X_2,\dots):\Omega\to\mathbb{R}^{\mathbb{N}} \text{ and } Y:=(Y_1,Y_2,\dots):\Omega'\to\mathbb{R}^{\mathbb{N}}$$

*and let* $\mathcal{F}:=\otimes_{n\in\mathbb{N}}\mathcal{B}_{\mathbb{R}}$ *– be the product* $\sigma$ *– algebra on* $\mathbb{R}^{\mathbb{N}}$*. Then* $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ *iff* $X_*P=Y_*P'$ *as measures on* $(\mathbb{R}^{\mathbb{N}},\mathcal{F})$.

**Proof.** Let

$$\mathcal{P}:=\cup_{n=1}^\infty\left\{A_1\times A_2\times\dots\times A_n\times\mathbb{R}^{\mathbb{N}}:A_i\in\mathcal{B}_{\mathbb{R}} \text{ for } 1\le i\le n\right\}.$$

Notice that $\mathcal{P}$ is a $\pi$ – system and it is easy to show $\sigma(\mathcal{P})=\mathcal{F}$ (see Exercise 8.6). Therefore by Proposition 6.18, $X_*P=Y_*P'$ iff $X_*P=Y_*P'$ on $\mathcal{P}$. Now for $A_1\times A_2\times\dots\times A_n\times\mathbb{R}^{\mathbb{N}}\in\mathcal{P}$ we have,

$$X_*P\left(A_1\times A_2\times\dots\times A_n\times\mathbb{R}^{\mathbb{N}}\right)=P((X_1,\dots,X_n)\in A_1\times A_2\times\dots\times A_n)$$

and hence the condition becomes,

$$P((X_1,\dots,X_n)\in A_1\times A_2\times\dots\times A_n)=P'((Y_1,\dots,Y_n)\in A_1\times A_2\times\dots\times A_n)$$

for all $n\in\mathbb{N}$ and $A_i\in\mathcal{B}_{\mathbb{R}}$. Another application of Proposition 6.18 or using Corollary 8.44 allows us to conclude that shows that $X_*P=Y_*P'$ iff $(X_1,\dots,X_n)\overset{d}{=}(Y_1,\dots,Y_n)$ for all $n\in\mathbb{N}$. ■

**Corollary 8.47.** *Continue the notation above and assume that* $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$*. Further let*

$$X_\pm=\begin{cases}\limsup_{n\to\infty}X_n & if +\\ \liminf_{n\to\infty}X_n & if -\end{cases}$$

*and define* $Y_\pm$ *similarly. Then* $(X_-,X_+)\overset{d}{=}(Y_-,Y_+)$ *as random variables into* $\left(\bar{\mathbb{R}}^2,\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}\right)$*. In particular,*

$$P\left(\lim_{n\to\infty}X_n \text{ exists in } \mathbb{R}\right)=P'\left(\lim_{n\to\infty}Y_n \text{ exists in } \mathbb{R}\right). \tag{8.9}$$

**Proof.** First suppose that $(\Omega',\mathcal{B}',P')=\left(\mathbb{R}^{\mathbb{N}},\mathcal{F},P':=X_*P\right)$ where $Y_i(a_1,a_2,\dots):=a_i=\pi_i(a_1,a_2,\dots)$. Then for $C\in\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}$ we have,

$$X^{-1}(\{(Y_-,Y_+)\in C\})=\{(Y_-\circ X,Y_+\circ X)\in C\}=\{(X_-,X_+)\in C\},$$

since, for example,

$$Y_-\circ X=\liminf_{n\to\infty}Y_n\circ X=\liminf_{n\to\infty}X_n=X_-.$$

Therefore it follows that

$$P((X_-,X_+)\in C)=P\circ X^{-1}(\{(Y_-,Y_+)\in C\})=P'(\{(Y_-,Y_+)\in C\}). \tag{8.10}$$

The general result now follows by two applications of this special case.

For the last assertion, take

$$C=\{(x,x):x\in\mathbb{R}\}\in\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\subset\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}.$$

Then $(X_-,X_+)\in C$ iff $X_-=X_+\in\mathbb{R}$ which happens iff $\lim_{n\to\infty}X_n$ exists in $\mathbb{R}$. Similarly, $(Y_-,Y_+)\in C$ iff $\lim_{n\to\infty}Y_n$ exists in $\mathbb{R}$ and therefore Eq. (8.9) holds as a consequence of Eq. (8.10). ■

**Exercise 8.10.** Let $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ be two sequences of random variables such that $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n:=X_1+\dots+X_n$ and $T_n:=Y_1+\dots+Y_n$. Prove the following assertions.

1. Suppose that $f:\mathbb{R}^n\to\mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1,\dots,X_n)\overset{d}{=}f(Y_1,\dots,Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty \overset{d}{=} \{T_n\}_{n=1}^\infty$.
   **Hint:** Apply item 1. with $k=n$ after making a judicious choice for $f:\mathbb{R}^n\to\mathbb{R}^n$.

## 8.5 Generating All Distributions from the Uniform Distribution

If $U\in(0,1)$ is a random variable with the uniform distribution, $G:(0,1)\to\mathbb{R}$ is a non-decreasing function, and $F(x):=P(G(U)\le x)$ is the cumulative distribution function of $G(U)$, then

$$F(x)=m(\{y\in(0,1):G(y)\le x\})=\sup\{y\in(0,1):G(y)\le x\}. \tag{8.11}$$

Now suppose that $F:\mathbb{R}\to[0,1]$ is a cumulative distribution function of a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$, i.e. $F$ is non-decreasing, right continuous, and $\lim_{x\to\infty}F(x)=1$, and $\lim_{x\to-\infty}F(x)=0$. We would like to find a function $G$ as above such that Eq. (8.11) holds. If $F$ happened to be continuous and

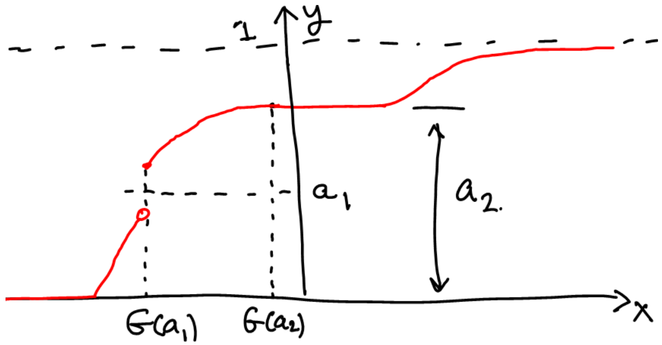



**Fig. 8.2.** A pictorial definition of $G$.

strictly increasing we should take $G(y) = F^{-1}(y)$. For general distributions functions $(F)$ we will show that

$$G(y) := \inf\{x \in \mathbb{R} : y \le F(x)\} \text{ for all } y \in (0,1), \tag{8.12}$$

(see see Figure 8.2) is the required function.

**Theorem 8.48.** *If $F : \mathbb{R} \to [0,1]$ is a cumulative distribution function of a probability measure $\mu = \mu_F$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $G$ is defined as in Eq. (8.12), then $G : (0,1) \to \mathbb{R}$ is a non-decreasing (hence Borel measurable) function such that* $\mathrm{Law}_P(G(U)) = G_* m = \mu_F$.

**Proof.** Since $y < 1$, $G(y) < \infty$ and since $y > 0$, $G(y) > -\infty$ wherein we have used $\lim_{x\to\infty} F(x) = 1$, and $\lim_{x\to-\infty} F(x) = 0$ respectively. Since $F$ is non-decreasing it is easily seen that $G$ is non-decreasing. To finish the proof it suffices to prove $G(y) \le x$ iff $y \le F(x)$. For once this is done it easily follows that Eq. (8.11) is valid.

By definition of $G(y)$, $G(y) \le x$ iff there exists a non-increasing sequence $\{x_n\} \subset \mathbb{R}$ such that $y \le F(x_n)$ for all $n$ and $\lim_{n\to\infty} x_n \le x$. Since $\lim_{\xi\downarrow-\infty} F(\xi) = 0$ and $y > 0$, it follows that $\lim_{n\to\infty} x_n = x_0 \in \mathbb{R}$ for some $x_0 \in (-\infty, x]$. Because $F$ is right continuous we may conclude that $F(x_0) \ge y$. Thus we have shown $G(y) \le x$ iff there exists $x_0 \le x$ such that $y \le F(x_0)$. As $F$ is non-decreasing this last equivalence is equivalent to $G(y) \le x$ iff $y \le F(x)$. ∎

**Theorem 8.49 (Durret's Version).** *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $Y : (0,1) \to \mathbb{R}$ be defined (see Figure 8.3) by,*

$$Y(x) := \sup\{y : F(y) < x\}.$$

*Then $Y : (0,1) \to \mathbb{R}$ is Borel measurable and $Y_* m = \mu_F$ where $\mu_F$ is the unique measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.*

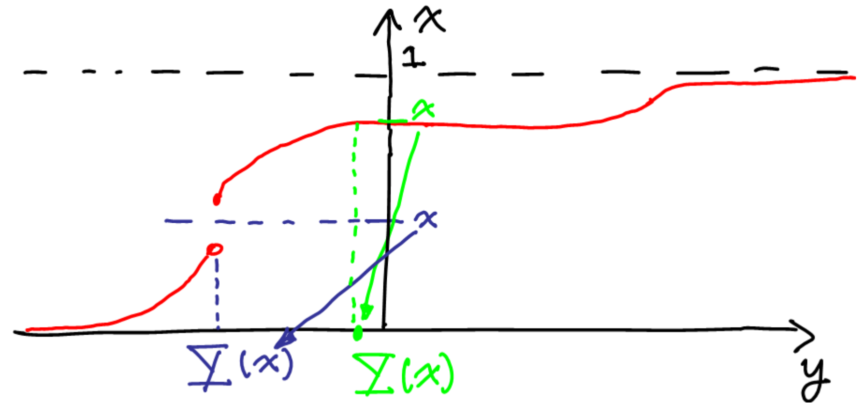



**Fig. 8.3.** A pictorial definition of $Y(x)$.

**Proof.** Since $Y : (0,1) \to \mathbb{R}$ is a non-decreasing function, $Y$ is measurable. Also observe, if $y < Y(x)$, then $F(y) < x$ and hence,

$$F(Y(x)-) = \lim_{y\uparrow Y(x)} F(y) \le x.$$

For $y > Y(x)$, we have $F(y) \ge x$ and therefore,

$$F(Y(x)) = F(Y(x)+) = \lim_{y\downarrow Y(x)} F(y) \ge x$$

and so we have shown

$$F(Y(x)-) \le x \le F(Y(x)).$$

We will now show

$$\{x \in (0,1) : Y(x) \le y_0\} = (0, F(y_0)] \cap (0,1). \tag{8.13}$$

For the inclusion "$\subset$," if $x \in (0,1)$ and $Y(x) \le y_0$, then $x \le F(Y(x)) \le F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely if $x \in (0,1)$ and $x \le F(y_0)$ then (by definition of $Y(x)$) $y_0 \ge Y(x)$.

From the identity in Eq. (8.13), it follows that $Y$ is measurable and

$$(Y_* m)((-\infty, y_0)) = m\left(Y^{-1}(-\infty, y_0)\right) = m((0, F(y_0)] \cap (0,1)) = F(y_0).$$

Therefore, $Law(Y) = \mu_F$ as desired. ∎

# 9

# Integration Theory

In this chapter, we will greatly extend the "simple" integral or expectation which was developed in Section 5.4 above. Recall there that if $(\Omega, \mathcal{B}, \mu)$ was measurable space and $\varphi : \Omega \to [0, \infty)$ was a measurable simple function, then we let

$$\mathbb{E}_\mu \varphi := \sum_{\lambda \in [0,\infty)} \lambda \mu \left(\varphi = \lambda\right).$$

The conventions being use here is that $0 \cdot \mu\left(\varphi = 0\right) = 0$ even when $\mu\left(\varphi = 0\right) = \infty$. This convention is necessary in order to make the integral linear – at a minimum we will want $\mathbb{E}_\mu\left[0\right] = 0$. Please be careful not blindly apply the $0 \cdot \infty = 0$ convention in other circumstances.

## 9.1 Integrals of positive functions

**Definition 9.1.** *Let $L^+ = L^+\left(\mathcal{B}\right) = \{f : \Omega \to [0, \infty] : f$ is measurable$\}$. Define*

$$\int_\Omega f\left(\omega\right) d\mu\left(\omega\right) = \int_\Omega f d\mu := \sup \left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\right\}.$$

*We say the $f \in L^+$ is **integrable** if $\int_\Omega f d\mu < \infty$. If $A \in \mathcal{B}$, let*

$$\int_A f\left(\omega\right) d\mu\left(\omega\right) = \int_A f d\mu := \int_\Omega 1_A f \ d\mu.$$

*We also use the notation,*

$$\mathbb{E}f = \int_\Omega f d\mu \text{ and } \mathbb{E}\left[f : A\right] := \int_A f d\mu.$$

*Remark 9.2.* Because of item 4. of Proposition 5.27, if $\varphi$ is a non-negative simple function, $\int_\Omega \varphi d\mu = \mathbb{E}_\mu \varphi$ so that $\int_\Omega$ is an extension of $\mathbb{E}_\mu$.

**Lemma 9.3.** *Let $f, g \in L^+\left(\mathcal{B}\right)$. Then:*

*1. if $\lambda \ge 0$, then*

$$\int_\Omega \lambda f d\mu = \lambda \int_\Omega f d\mu$$

*wherein $\lambda \int_\Omega f d\mu \equiv 0$ if $\lambda = 0$, even if $\int_\Omega f d\mu = \infty$.*

*2. if $0 \le f \le g$, then*

$$\int_\Omega f d\mu \le \int_\Omega g d\mu. \tag{9.1}$$

*3. For all $\varepsilon > 0$ and $p > 0$,*

$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon^p} \int_\Omega f^p 1_{\{f \ge \varepsilon\}} d\mu \le \frac{1}{\varepsilon^p} \int_\Omega f^p d\mu. \tag{9.2}$$

*The inequality in Eq. (9.2) is called Chebyshev's Inequality for $p = 1$ and Markov's inequality for $p = 2$.*

*4. If $\int_\Omega f d\mu < \infty$ then $\mu(f = \infty) = 0$ (i.e. $f < \infty$ a.e.) and the set $\{f > 0\}$ is $\sigma$ – finite.*

**Proof.** 1. We may assume $\lambda > 0$ in which case,

$$\begin{aligned}
\int_\Omega \lambda f d\mu &= \sup \left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le \lambda f\right\} \\
&= \sup \left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \lambda^{-1}\varphi \le f\right\} \\
&= \sup \left\{\mathbb{E}_\mu\left[\lambda \psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \sup \left\{\lambda \mathbb{E}_\mu\left[\psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \lambda \int_\Omega f d\mu.
\end{aligned}$$

2. Since

$$\left\{\varphi \text{ is simple and } \varphi \le f\right\} \subset \left\{\varphi \text{ is simple and } \varphi \le g\right\},$$

Eq. (9.1) follows from the definition of the integral.

3. Since $1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \frac{1}{\varepsilon} f \le \frac{1}{\varepsilon} f$ we have

$$1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \le \left(\frac{1}{\varepsilon} f\right)^p$$

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \ge \varepsilon) = \int_\Omega 1_{\{f \ge \varepsilon\}} d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega 1_{\{f \ge \varepsilon\}} f^p d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega f^p d\mu.$$

4. If $\mu(f=\infty)>0$, then $\varphi_n := n1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \le f$ for all $n$ and hence

$$n\mu(f=\infty) = \mathbb{E}_\mu(\varphi_n) \le \int_\Omega f d\mu$$

for all $n$. Letting $n\to\infty$ shows $\int_\Omega f d\mu = \infty$. Thus if $\int_\Omega f d\mu < \infty$ then $\mu(f=\infty)=0$.

Moreover,

$$\{f>0\} = \cup_{n=1}^\infty \{f>1/n\}$$

with $\mu(f>1/n) \le n\int_\Omega f d\mu < \infty$ for each $n$. ■

**Theorem 9.4 (Monotone Convergence Theorem).** *Suppose $f_n \in L^+$ is a sequence of functions such that $f_n \uparrow f$ ($f$ is necessarily in $L^+$) then*

$$\int f_n \uparrow \int f \text{ as } n\to\infty.$$

**Proof.** Since $f_n \le f_m \le f$, for all $n \le m < \infty$,

$$\int f_n \le \int f_m \le \int f$$

from which if follows $\int f_n$ is increasing in $n$ and

$$\lim_{n\to\infty}\int f_n \le \int f. \tag{9.3}$$

For the opposite inequality, let $\varphi : \Omega \to [0,\infty)$ be a simple function such that $0\le\varphi\le f$, $\alpha\in(0,1)$ and $\Omega_n := \{f_n \ge \alpha\varphi\}$.[1] Notice that $\Omega_n \uparrow \Omega$ and $f_n \ge \alpha 1_{\Omega_n}\cdot\varphi$ and so by definition of $\int f_n$,

$$\int f_n \ge \mathbb{E}_\mu[\alpha 1_{\Omega_n}\varphi] = \alpha\mathbb{E}_\mu[1_{\Omega_n}\varphi]. \tag{9.4}$$

Then using the identity

$$1_{\Omega_n}\varphi = 1_{\Omega_n}\sum_{y>0} y 1_{\{\varphi=y\}} = \sum_{y>0} y 1_{\{\varphi=y\}\cap\Omega_n},$$

and the linearity of $\mathbb{E}_\mu$ we have,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}_\mu[1_{\Omega_n}\varphi] &= \lim_{n\to\infty}\sum_{y>0} y\cdot\mu(\Omega_n\cap\{\varphi=y\}) \\ &= \sum_{y>0} y \lim_{n\to\infty}\mu(\Omega_n\cap\{\varphi=y\}) \text{ (finite sum)} \\ &= \sum_{y>0} y\mu(\{\varphi=y\}) = \mathbb{E}_\mu[\varphi],\end{aligned}$$

wherein we have used the continuity of $\mu$ under increasing unions for the third equality. This identity allows us to let $n\to\infty$ in Eq. (9.4) to conclude $\lim_{n\to\infty}\int f_n \ge \alpha\mathbb{E}_\mu[\varphi]$ and since $\alpha\in(0,1)$ was arbitrary we may further conclude, $\mathbb{E}_\mu[\varphi] \le \lim_{n\to\infty}\int f_n$. The latter inequality being true for all simple functions $\varphi$ with $\varphi\le f$ then implies that

$$\int f = \sup_{0\le\varphi\le f}\mathbb{E}_\mu[\varphi] \le \lim_{n\to\infty}\int f_n,$$

which combined with Eq. (9.3) proves the theorem. ■

*Remark 9.5.* The definition $\int f d\mu$ makes sense for **all** functions $f:\Omega\to[0,\infty]$ and not just measurable functions. However, the measurability of the $f_n$ (and hence $f=\lim_{n\to\infty} f_n$) was needed in the proof of the monotone convergence theorem in order for $\Omega_n := \{f_n \ge \alpha\varphi\}$ to be measurable.

*Remark 9.6 ("Explicit" Integral Formula).* Given $f:\Omega\to[0,\infty]$ measurable, we know from the approximation Theorem 8.39 $\varphi_n\uparrow f$ where

$$\varphi_n := \sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\left\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right\}} + 2^n 1_{\{f>2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned}\int_\Omega f d\mu &= \lim_{n\to\infty}\int_\Omega \varphi_n d\mu \\ &= \lim_{n\to\infty}\left[\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}\mu\left(\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right) + 2^n\mu(f>2^n)\right].\end{aligned}$$

**Corollary 9.7.** *If $f_n\in L^+$ is a sequence of functions then*

$$\int\sum_{n=1}^\infty f_n = \sum_{n=1}^\infty\int f_n.$$

*In particular, if $\sum_{n=1}^\infty\int f_n<\infty$ then $\sum_{n=1}^\infty f_n<\infty$ a.e.*

[1] Notice that in order for $\Omega_n$ to be measurable we must assume that $f_n$ is measurable here.

**Proof.** First off we show that

$$\int (f_1+f_2) = \int f_1 + \int f_2$$

by choosing non-negative simple function $\varphi_n$ and $\psi_n$ such that $\varphi_n \uparrow f_1$ and $\psi_n \uparrow f_2$. Then $(\varphi_n+\psi_n)$ is simple as well and $(\varphi_n+\psi_n)\uparrow(f_1+f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int (f_1+f_2) &= \lim_{n\to\infty}\int(\varphi_n+\psi_n) = \lim_{n\to\infty}\left(\int\varphi_n + \int\psi_n\right)\\ &= \lim_{n\to\infty}\int\varphi_n + \lim_{n\to\infty}\int\psi_n = \int f_1 + \int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum_{n=1}^{N} f_n$ and $g = \sum_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\begin{aligned}\sum_{n=1}^{\infty}\int f_n &:= \lim_{N\to\infty}\sum_{n=1}^{N}\int f_n = \lim_{N\to\infty}\int\sum_{n=1}^{N} f_n\\ &= \lim_{N\to\infty}\int g_N = \int g =: \int\sum_{n=1}^{\infty} f_n.\end{aligned}$$

■

*Example 9.8 (Sums as Integrals I).* Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, $\mu(A) = \#(A)$ for $A \subset \Omega$ is the counting measure on $\mathcal{B}$, and $f : \mathbb{N} \to [0,\infty]$ is a function. Since

$$f = \sum_{n=1}^{\infty} f(n)\, 1_{\{n\}},$$

it follows from Corollary 9.7 that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty}\int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^{\infty} f(n)\,\mu(\{n\}) = \sum_{n=1}^{\infty} f(n).$$

Thus the integral relative to counting measure is simply the infinite sum.

**Lemma 9.9 (Sums as Integrals II*).** *Let $\Omega$ be a set and $\rho : \Omega \to [0,\infty]$ be a function, let $\mu = \sum_{\omega\in\Omega}\rho(\omega)\delta_\omega$ on $\mathcal{B} = 2^{\Omega}$, i.e.*

$$\mu(A) = \sum_{\omega\in A}\rho(\omega).$$

*If $f : \Omega \to [0,\infty]$ is a function (which is necessarily measurable), then*

$$\int_\Omega f d\mu = \sum_\Omega f\rho.$$

**Proof.** Suppose that $\varphi : \Omega \to [0,\infty)$ is a simple function, then $\varphi = \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}$ and

$$\begin{aligned}\sum_\Omega \varphi\rho &= \sum_{\omega\in\Omega}\rho(\omega)\sum_{z\in[0,\infty)} z1_{\{\varphi=z\}}(\omega) = \sum_{z\in[0,\infty)} z\sum_{\omega\in\Omega}\rho(\omega)1_{\{\varphi=z\}}(\omega)\\ &= \sum_{z\in[0,\infty)} z\mu(\{\varphi=z\}) = \int_\Omega \varphi d\mu.\end{aligned}$$

So if $\varphi : \Omega \to [0,\infty)$ is a simple function such that $\varphi \le f$, then

$$\int_\Omega \varphi d\mu = \sum_\Omega \varphi\rho \le \sum_\Omega f\rho.$$

Taking the sup over $\varphi$ in this last equation then shows that

$$\int_\Omega f d\mu \le \sum_\Omega f\rho.$$

For the reverse inequality, let $\Lambda \subset_f \Omega$ be a finite set and $N \in (0,\infty)$. Set $f^N(\omega) = \min\{N, f(\omega)\}$ and let $\varphi_{N,\Lambda}$ be the simple function given by $\varphi_{N,\Lambda}(\omega) := 1_\Lambda(\omega) f^N(\omega)$. Because $\varphi_{N,\Lambda}(\omega) \le f(\omega)$,

$$\sum_\Lambda f^N\rho = \sum_\Omega \varphi_{N,\Lambda}\rho = \int_\Omega \varphi_{N,\Lambda} d\mu \le \int_\Omega f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_\Lambda f\rho \le \int_\Omega f d\mu.$$

Since $\Lambda$ is arbitrary, this implies

$$\sum_\Omega f\rho \le \int_\Omega f d\mu.$$

■

**Exercise 9.1.** Suppose that $\mu_n : \mathcal{B} \to [0,\infty]$ are measures on $\mathcal{B}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in $n$ for all $A \in \mathcal{B}$. Prove that $\mu : \mathcal{B} \to [0,\infty]$ defined by $\mu(A) := \lim_{n\to\infty}\mu_n(A)$ is also a measure.

**Proposition 9.10.** *Suppose that $f \ge 0$ is a measurable function. Then $\int_\Omega f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \ge 0$ are measurable functions such that $f \le g$ a.e. then $\int f d\mu \le \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

**Proof.** If $f=0$ a.e. and $\varphi \le f$ is a simple function then $\varphi = 0$ a.e. This implies that $\mu(\varphi^{-1}(\{y\})) = 0$ for all $y>0$ and hence $\int_\Omega \varphi d\mu = 0$ and therefore $\int_\Omega f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 9.3),

$$\mu(f \ge 1/n) \le n \int f d\mu = 0 \text{ for all } n.$$

Therefore, $\mu(f>0) \le \sum_{n=1}^\infty \mu(f \ge 1/n) = 0$, i.e. $f=0$ a.e.

For the second assertion let $E$ be the exceptional set where $f>g$, i.e.

$$E := \{\omega \in \Omega : f(\omega) > g(\omega)\}.$$

By assumption $E$ is a null set and $1_{E^c} f \le 1_{E^c} g$ everywhere. Because $g = 1_{E^c} g + 1_E g$ and $1_E g = 0$ a.e.,

$$\int g d\mu = \int 1_{E^c} g d\mu + \int 1_E g d\mu = \int 1_{E^c} g d\mu$$

and similarly $\int f d\mu = \int 1_{E^c} f d\mu$. Since $1_{E^c} f \le 1_{E^c} g$ everywhere,

$$\int f d\mu = \int 1_{E^c} f d\mu \le \int 1_{E^c} g d\mu = \int g d\mu.$$

■

**Corollary 9.11.** *Suppose that $\{f_n\}$ is a sequence of non-negative measurable functions and $f$ is a measurable function such that $f_n \uparrow f$ off a null set, then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

**Proof.** Let $E \subset \Omega$ be a null set such that $f_n 1_{E^c} \uparrow f 1_{E^c}$ as $n \to \infty$. Then by the monotone convergence theorem and Proposition 9.10,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

**Lemma 9.12 (Fatou's Lemma).** *If $f_n : \Omega \to [0,\infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n$$

**Proof.** Define $g_k := \inf_{n \ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\int g_k \le \int f_n \text{ for all } n \ge k$$

and therefore

$$\int g_k \le \liminf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \liminf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \le \liminf_{n\to\infty} \int f_n.$$

■

The following Corollary and the next lemma are simple applications of Corollary 9.7.

**Corollary 9.13.** *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{B}$ is a collection of sets such that $\mu(A_i \cap A_j) = 0$ for all $i \ne j$, then*

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n).$$

**Proof.** Since

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \int_\Omega 1_{\cup_{n=1}^\infty A_n} d\mu \text{ and}$$
$$\sum_{n=1}^\infty \mu(A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} d\mu$$

it suffices to show

$$\sum_{n=1}^\infty 1_{A_n} = 1_{\cup_{n=1}^\infty A_n} \ \mu - \text{a.e.} \tag{9.5}$$

Now $\sum_{n=1}^\infty 1_{A_n} \ge 1_{\cup_{n=1}^\infty A_n}$ and $\sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)$ iff $\omega \in A_i \cap A_j$ for some $i \ne j$, that is

$$\left\{\omega : \sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)\right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (9.5) and hence the corollary. ■

**Lemma 9.14 (The First Borell – Cantelli Lemma).** *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

$$\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^\infty \bigcup_{n \ge N} A_n.$$

*If $\sum_{n=1}^\infty \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.*

**Proof.** (First Proof.) Let us first observe that

$$\{A_n \text{ i.o.}\} = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\right\}.$$

Hence if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \int_\Omega 1_{A_n} \, d\mu = \int_\Omega \sum_{n=1}^{\infty} 1_{A_n} \, d\mu$$

implies that $\sum_{n=1}^{\infty} 1_{A_n}(\omega) < \infty$ for $\mu$ - a.e. $\omega$. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$.

(Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned} \mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} A_n\right) \\ &\leq \lim_{N\to\infty} \sum_{n\geq N} \mu(A_n) \end{aligned}$$

and the last limit is zero since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ■

*Example 9.15.* Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space (i.e. $P(\Omega) = 1$) and $X_n : \Omega \to \{0, 1\}$ are Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $P(X_n = 1 \text{ i.o.}) = 0$ and hence $P(X_n = 0 \text{ a.a.}) = 1$. In particular, $P(\lim_{n\to\infty} X_n = 0) = 1$.

## 9.2 Integrals of Complex Valued Functions

**Definition 9.16.** *A measurable function $f : \Omega \to \bar{\mathbb{R}}$ is **integrable** if $f_+ := f 1_{\{f\geq 0\}}$ and $f_- = -f \; 1_{\{f\leq 0\}}$ are **integrable**. We write $\mathrm{L}^1(\mu; \mathbb{R})$ for the space of real valued integrable functions. For $f \in \mathrm{L}^1(\mu; \mathbb{R})$, let*

$$\int_\Omega f d\mu = \int_\Omega f_+ d\mu - \int_\Omega f_- d\mu.$$

*To shorten notation in this chapter we may simply write $\int f d\mu$ or even $\int f$ for $\int_\Omega f d\mu$.*

**Convention:** If $f, g : \Omega \to \bar{\mathbb{R}}$ are two measurable functions, let $f + g$ denote the collection of measurable functions $h : \Omega \to \bar{\mathbb{R}}$ such that $h(\omega) = f(\omega) + g(\omega)$ whenever $f(\omega) + g(\omega)$ is well defined, i.e. is not of the form $\infty - \infty$ or $-\infty + \infty$. We use a similar convention for $f - g$. Notice that if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $h_1, h_2 \in f + g$, then $h_1 = h_2$ a.e. because $|f| < \infty$ and $|g| < \infty$ a.e.

**Notation 9.17 (Abuse of notation)** *We will sometimes denote the integral $\int_\Omega f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{B}$.*

*Remark 9.18.* Since

$$f_\pm \leq |f| \leq f_+ + f_-,$$

a measurable function $f$ is **integrable** iff $\int |f| \; d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu; \mathbb{R}) := \left\{f : \Omega \to \bar{\mathbb{R}} : \; f \text{ is measurable and } \int_\Omega |f| \; d\mu < \infty\right\}.$$

If $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $f = g$ a.e. then $f_\pm = g_\pm$ a.e. and so it follows from Proposition 9.10 that $\int f d\mu = \int g d\mu$. In particular if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ we may define

$$\int_\Omega (f + g) \, d\mu = \int_\Omega h d\mu$$

where $h$ is any element of $f + g$.

**Proposition 9.19.** *The map*

$$f \in \mathrm{L}^1(\mu; \mathbb{R}) \to \int_\Omega f d\mu \in \mathbb{R}$$

*is linear and has the monotonicity property: $\int f d\mu \leq \int g d\mu$ for all $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ such that $f \leq g$ a.e.*

**Proof.** Let $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $a, b \in \mathbb{R}$. By modifying $f$ and $g$ on a null set, we may assume that $f, g$ are real valued functions. We have $af + bg \in \mathrm{L}^1(\mu; \mathbb{R})$ because

$$|af + bg| \leq |a| \, |f| + |b| \, |g| \in \mathrm{L}^1(\mu; \mathbb{R}).$$

If $a < 0$, then

$$(af)_+ = -af_- \text{ and } (af)_- = -af_+$$

so that

$$\int af = -a \int f_- + a \int f_+ = a(\int f_+ - \int f_-) = a \int f.$$

A similar calculation works for $a > 0$ and the case $a = 0$ is trivial so we have shown that

$$\int af = a \int f.$$

Now set $h = f + g$. Since $h = h_+ - h_-$,

$$h_+ - h_- = f_+ - f_- + g_+ - g_-$$

or

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Therefore,

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+$$

and hence

$$\int h = \int h_+ - \int h_- = \int f_+ + \int g_+ - \int f_- - \int g_- = \int f + \int g.$$

Finally if $f_+ - f_- = f \le g = g_+ - g_-$ then $f_+ + g_- \le g_+ + f_-$ which implies that

$$\int f_+ + \int g_- \le \int g_+ + \int f_-$$

or equivalently that

$$\int f = \int f_+ - \int f_- \le \int g_+ - \int g_- = \int g.$$

The monotonicity property is also a consequence of the linearity of the integral, the fact that $f \le g$ a.e. implies $0 \le g - f$ a.e. and Proposition 9.10. ■

**Definition 9.20.** *A measurable function $f : \Omega \to \mathbb{C}$ is* ***integrable*** *if $\int_\Omega |f| \ d\mu < \infty$. Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : \Omega \to \mathbb{C} : \ f \text{ is measurable and } \int_\Omega |f| \ d\mu < \infty \right\}.$$

*denote the complex valued integrable functions. Because,* $\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|) \le |f| \le \sqrt{2}\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|)$, $\int |f| \ d\mu < \infty$ *iff*

$$\int |\mathrm{Re}\, f| \, d\mu + \int |\mathrm{Im}\, f| \, d\mu < \infty.$$

*For $f \in \mathrm{L}^1(\mu;\mathbb{C})$ define*

$$\int f \ d\mu = \int \mathrm{Re}\, f \ d\mu + i \int \mathrm{Im}\, f \ d\mu.$$

It is routine to show the integral is still linear on $\mathrm{L}^1(\mu;\mathbb{C})$ (prove!). In the remainder of this section, let $\mathrm{L}^1(\mu)$ be either $\mathrm{L}^1(\mu;\mathbb{C})$ or $\mathrm{L}^1(\mu;\mathbb{R})$. If $A \in \mathcal{B}$ and $f \in \mathrm{L}^1(\mu;\mathbb{C})$ or $f : \Omega \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_\Omega 1_A f d\mu.$$

**Proposition 9.21.** *Suppose that $f \in \mathrm{L}^1(\mu;\mathbb{C})$, then*

$$\left| \int_\Omega f d\mu \right| \le \int_\Omega |f| \, d\mu. \tag{9.6}$$

**Proof.** Start by writing $\int_\Omega f \ d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_\Omega f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta} \int_\Omega f \ d\mu = \int_\Omega e^{-i\theta} f \ d\mu = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i \int_\Omega \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_\Omega \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 9.10,

$$\left| \int_\Omega f d\mu \right| = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_\Omega \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_\Omega |f| \, d\mu.$$

■

**Proposition 9.22.** *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \ne 0\}$ is $\sigma$ – finite, in fact $\{|f| \ge \frac{1}{n}\} \uparrow \{f \ne 0\}$ and $\mu(|f| \ge \frac{1}{n}) < \infty$ for all $n$.*
2. *The following are equivalent*
   a) *$\int_E f = \int_E g$ for all $E \in \mathcal{B}$*
   b) *$\int_\Omega |f - g| = 0$*
   c) *$f = g$ a.e.*

**Proof.** 1. By Chebyshev's inequality, Lemma 9.3,

$$\mu(|f| \ge \frac{1}{n}) \le n \int_\Omega |f| \, d\mu < \infty$$

for all $n$.

2. (a) $\Longrightarrow$ (c) Notice that

$$\int_E f = \int_E g \Leftrightarrow \int_E (f - g) = 0$$

for all $E \in \mathcal{B}$. Taking $E = \{\mathrm{Re}(f - g) > 0\}$ and using $1_E \mathrm{Re}(f - g) \ge 0$, we learn that

$$0 = \mathrm{Re} \int_E (f - g) d\mu = \int 1_E \mathrm{Re}(f - g) \Longrightarrow 1_E \mathrm{Re}(f - g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

$$\mu(\{\mathrm{Re}(f-g)>0\})=\mu(E)=0.$$

Similar $\mu(\mathrm{Re}(f-g)<0)=0$ so that $\mathrm{Re}(f-g)=0$ a.e. Similarly, $\mathrm{Im}(f-g)=0$ a.e and hence $f-g=0$ a.e., i.e. $f=g$ a.e.

(c) $\implies$ (b) is clear and so is (b) $\implies$ (a) since

$$\left|\int_E f-\int_E g\right|\le\int|f-g|=0.$$

■

**Lemma 9.23 (Integral Comparison I).** *Suppose that $h\in \mathrm{L}^1(\mu)$ satisfies*

$$\int_A hd\mu\ge 0 \text{ for all } A\in\mathcal{B}, \tag{9.7}$$

*then $h\ge 0$ a.e.*

**Proof.** Since by assumption,

$$0=\mathrm{Im}\int_A hd\mu=\int_A \mathrm{Im}\,hd\mu \text{ for all } A\in\mathcal{B},$$

we may apply Proposition 9.22 to conclude that $\mathrm{Im}\,h=0$ a.e. Thus we may now assume that $h$ is real valued. Taking $A=\{h<0\}$ in Eq. (9.7) implies

$$\int_\Omega 1_A|h|\,d\mu=\int_\Omega -1_A hd\mu=-\int_A hd\mu\le 0.$$

However $1_A|h|\ge 0$ and therefore it follows that $\int_\Omega 1_A|h|\,d\mu=0$ and so Proposition 9.22 implies $1_A|h|=0$ a.e. which then implies $0=\mu(A)=\mu(h<0)=0$. ■

**Lemma 9.24 (Integral Comparison II).** *Suppose $(\Omega,\mathcal{B},\mu)$ is a $\sigma$ – finite measure space (i.e. there exists $\Omega_n\in\mathcal{B}$ such that $\Omega_n\uparrow\Omega$ and $\mu(\Omega_n)<\infty$ for all $n$) and $f,g:\Omega\to[0,\infty]$ are $\mathcal{B}$ – measurable functions. Then $f\ge g$ a.e. iff*

$$\int_A fd\mu\ge\int_A gd\mu \text{ for all } A\in\mathcal{B}. \tag{9.8}$$

*In particular $f=g$ a.e. iff equality holds in Eq. (9.8).*

**Proof.** It was already shown in Proposition 9.10 that $f\ge g$ a.e. implies Eq. (9.8). For the converse assertion, let $B_n:=\{f\le n1_{\Omega_n}\}$. Then from Eq. (9.8),

$$\infty>n\mu(\Omega_n)\ge\int f1_{B_n}d\mu\ge\int g1_{B_n}d\mu$$

from which it follows that both $f1_{B_n}$ and $g1_{B_n}$ are in $\mathrm{L}^1(\mu)$ and hence $h:=f1_{B_n}-g1_{B_n}\in\mathrm{L}^1(\mu)$. Using Eq. (9.8) again we know that

$$\int_A h=\int f1_{B_n\cap A}-\int g1_{B_n\cap A}\ge 0 \text{ for all } A\in\mathcal{B}.$$

An application of Lemma 9.23 implies $h\ge 0$ a.e., i.e. $f1_{B_n}\ge g1_{B_n}$ a.e. Since $B_n\uparrow\{f<\infty\}$, we may conclude that

$$f1_{\{f<\infty\}}=\lim_{n\to\infty}f1_{B_n}\ge\lim_{n\to\infty}g1_{B_n}=g1_{\{f<\infty\}} \text{ a.e.}$$

Since $f\ge g$ whenever $f=\infty$, we have shown $f\ge g$ a.e.

If equality holds in Eq. (9.8), then we know that $g\le f$ and $f\le g$ a.e., i.e. $f=g$ a.e. ■

Notice that we can not drop the $\sigma$ – finiteness assumption in Lemma 9.24. For example, let $\mu$ be the measure on $\mathcal{B}$ such that $\mu(A)=\infty$ when $A\neq\emptyset$, $g=3$, and $f=2$. Then equality holds (both sides are infinite unless $A=\emptyset$ when they are both zero) in Eq. (9.8) holds even though $f<g$ **everywhere.**

**Definition 9.25.** *Let $(\Omega,\mathcal{B},\mu)$ be a measure space and $L^1(\mu)=L^1(\Omega,\mathcal{B},\mu)$ denote the set of $\mathrm{L}^1(\mu)$ functions modulo the equivalence relation; $f\sim g$ iff $f=g$ a.e. We make this into a normed space using the norm*

$$\|f-g\|_{L^1}=\int|f-g|\,d\mu$$

*and into a metric space using $\rho_1(f,g)=\|f-g\|_{L^1}$.*

**Warning:** in the future we will often not make much of a distinction between $L^1(\mu)$ and $\mathrm{L}^1(\mu)$. On occasion this can be dangerous and this danger will be pointed out when necessary.

*Remark 9.26.* More generally we may define $L^p(\mu)=L^p(\Omega,\mathcal{B},\mu)$ for $p\in[1,\infty)$ as the set of measurable functions $f$ such that

$$\int_\Omega|f|^p\,d\mu<\infty$$

modulo the equivalence relation; $f\sim g$ iff $f=g$ a.e.

We will see in later that

$$\|f\|_{L^p}=\left(\int|f|^p\,d\mu\right)^{1/p} \text{ for } f\in L^p(\mu)$$

is a norm and $(L^p(\mu),\|\cdot\|_{L^p})$ is a Banach space in this norm and in particular,

$$\|f+g\|_p\le\|f\|_p+\|g\|_p \text{ for all } f,g\in L^p(\mu).$$

**Theorem 9.27 (Dominated Convergence Theorem).** *Suppose $f_n, g_n, g \in \mathrm{L}^1(\mu)$, $f_n \to f$ a.e., $|f_n| \le g_n \in \mathrm{L}^1(\mu)$, $g_n \to g$ a.e. and $\int_\Omega g_n d\mu \to \int_\Omega g d\mu$. Then $f \in \mathrm{L}^1(\mu)$ and*

$$\int_\Omega f d\mu = \lim_{h\to\infty} \int_\Omega f_n d\mu.$$

*(In most typical applications of this theorem $g_n = g \in \mathrm{L}^1(\mu)$ for all $n$.)*

**Proof.** Notice that $|f| = \lim_{n\to\infty} |f_n| \le \lim_{n\to\infty} |g_n| \le g$ a.e. so that $f \in \mathrm{L}^1(\mu)$. By considering the real and imaginary parts of $f$ separately, it suffices to prove the theorem in the case where $f$ is real. By Fatou's Lemma,

$$\begin{aligned}\int_\Omega (g \pm f) d\mu &= \int_\Omega \liminf_{n\to\infty} (g_n \pm f_n)\, d\mu \le \liminf_{n\to\infty} \int_\Omega (g_n \pm f_n)\, d\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu\right) \\ &= \int_\Omega g d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu\right)\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_\Omega g d\mu \pm \int_\Omega f d\mu \le \int_\Omega g d\mu + \begin{cases} \liminf_{n\to\infty} \int_\Omega f_n d\mu \\ -\limsup_{n\to\infty} \int_\Omega f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \le \int_\Omega f d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

This shows that $\lim_{n\to\infty} \int_\Omega f_n d\mu$ exists and is equal to $\int_\Omega f d\mu$. ■

**Exercise 9.2.** Give another proof of Proposition 9.21 by first proving Eq. (9.6) with $f$ being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 8.39 along with the dominated convergence Theorem 9.27 to handle the general case.

**Corollary 9.28.** *Let $\{f_n\}_{n=1}^\infty \subset \mathrm{L}^1(\mu)$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$, then $\sum_{n=1}^\infty f_n$ is convergent a.e. and*

$$\int_\Omega \left(\sum_{n=1}^\infty f_n\right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

**Proof.** The condition $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \le \sum_{n=1}^N |f_n| \le \sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu).$$

So by the dominated convergence theorem,

$$\begin{aligned}\int_\Omega \left(\sum_{n=1}^\infty f_n\right) d\mu &= \int_\Omega \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty} \int_\Omega S_N d\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int_\Omega f_n d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.\end{aligned}$$

■

*Example 9.29 (Sums as integrals).* Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, $\mu$ is counting measure on $\mathcal{B}$ (see Example 9.8), and $f : \mathbb{N} \to \mathbb{C}$ is a function. From Example 9.8 we have $f \in L^1(\mu)$ iff $\sum_{n=1}^\infty |f(n)| < \infty$, i.e. iff the sum, $\sum_{n=1}^\infty f(n)$ is absolutely convergent. Moreover, if $f \in L^1(\mu)$, we may again write

$$f = \sum_{n=1}^\infty f(n)\, 1_{\{n\}}$$

and then use Corollary 9.28 to conclude that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^\infty \int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^\infty f(n)\, \mu(\{n\}) = \sum_{n=1}^\infty f(n).$$

So again the integral relative to counting measure is simply the infinite sum **provided** the sum is absolutely convergent.

However if $f(n) = (-1)^n \frac{1}{n}$, then

$$\sum_{n=1}^\infty f(n) := \lim_{N\to\infty} \sum_{n=1}^N f(n)$$

is perfectly well defined while $\int_{\mathbb{N}} f d\mu$ is **not**. In fact in this case we have,

$$\int_{\mathbb{N}} f_\pm d\mu = \infty.$$

The point is that when we write $\sum_{n=1}^\infty f(n)$ the ordering of the terms in the sum may matter. On the other hand, $\int_{\mathbb{N}} f d\mu$ knows nothing about the integer ordering.

The following corollary will be routinely be used in the sequel – often without explicit mention.

**Corollary 9.30 (Differentiation Under the Integral).** *Suppose that $J \subset \mathbb{R}$ is an open interval and $f : J \times \Omega \to \mathbb{C}$ is a function such that*

1. *$\omega \to f(t,\omega)$ is measurable for each $t \in J$.*
2. *$f(t_0,\cdot) \in L^1(\mu)$ for some $t_0 \in J$.*
3. *$\frac{\partial f}{\partial t}(t,\omega)$ exists for all $(t,\omega)$.*
4. *There is a function $g \in \mathrm{L}^1(\mu)$ such that $\left|\frac{\partial f}{\partial t}(t,\cdot)\right| \le g$ for each $t \in J$.*

*Then $f(t,\cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$ (i.e. $\int_\Omega |f(t,\omega)|\,d\mu(\omega) < \infty$), $t \to \int_\Omega f(t,\omega)d\mu(\omega)$ is a differentiable function on $J$, and*

$$\frac{d}{dt}\int_\Omega f(t,\omega)d\mu(\omega) = \int_\Omega \frac{\partial f}{\partial t}(t,\omega)d\mu(\omega).$$

**Proof.** By considering the real and imaginary parts of $f$ separately, we may assume that $f$ is real. Also notice that

$$\frac{\partial f}{\partial t}(t,\omega) = \lim_{n\to\infty} n(f(t+n^{-1},\omega) - f(t,\omega))$$

and therefore, for $\omega \to \frac{\partial f}{\partial t}(t,\omega)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t,\omega) - f(t_0,\omega)| \le g(\omega)\,|t - t_0| \text{ for all } t \in J \tag{9.9}$$

and hence

$$|f(t,\omega)| \le |f(t,\omega) - f(t_0,\omega)| + |f(t_0,\omega)| \le g(\omega)\,|t-t_0| + |f(t_0,\omega)|\,.$$

This shows $f(t,\cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$. Let $G(t) := \int_\Omega f(t,\omega)d\mu(\omega)$, then

$$\frac{G(t) - G(t_0)}{t - t_0} = \int_\Omega \frac{f(t,\omega) - f(t_0,\omega)}{t - t_0} d\mu(\omega).$$

By assumption,

$$\lim_{t\to t_0} \frac{f(t,\omega) - f(t_0,\omega)}{t - t_0} = \frac{\partial f}{\partial t}(t,\omega) \text{ for all } \omega \in \Omega$$

and by Eq. (9.9),

$$\left|\frac{f(t,\omega) - f(t_0,\omega)}{t - t_0}\right| \le g(\omega) \text{ for all } t \in J \text{ and } \omega \in \Omega.$$



Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned}\lim_{n\to\infty} \frac{G(t_n) - G(t_0)}{t_n - t_0} &= \lim_{n\to\infty}\int_\Omega \frac{f(t_n,\omega) - f(t_0,\omega)}{t_n - t_0} d\mu(\omega)\\ &= \int_\Omega \lim_{n\to\infty} \frac{f(t_n,\omega) - f(t_0,\omega)}{t_n - t_0} d\mu(\omega)\\ &= \int_\Omega \frac{\partial f}{\partial t}(t_0,\omega)d\mu(\omega)\end{aligned}$$

for **all** sequences $t_n \in J \setminus \{t_0\}$ such that $t_n \to t_0$. Therefore, $\dot{G}(t_0) = \lim_{t\to t_0} \frac{G(t)-G(t_0)}{t-t_0}$ exists and

$$\dot{G}(t_0) = \int_\Omega \frac{\partial f}{\partial t}(t_0,\omega)d\mu(\omega).$$

■

**Corollary 9.31.** *Suppose that $\{a_n\}_{n=0}^\infty \subset \mathbb{C}$ is a sequence of complex numbers such that series*

$$f(z) := \sum_{n=0}^\infty a_n (z - z_0)^n$$

*is convergent for $|z - z_0| < R$, where $R$ is some positive number. Then $f : D(z_0,R) \to \mathbb{C}$ is complex differentiable on $D(z_0,R)$ and*

$$f'(z) = \sum_{n=0}^\infty n a_n (z-z_0)^{n-1} = \sum_{n=1}^\infty n a_n (z - z_0)^{n-1}. \tag{9.10}$$

*By induction it follows that $f^{(k)}$ exists for all $k$ and that*

$$f^{(k)}(z) = \sum_{n=0}^\infty n(n-1)\dots(n-k+1)a_n(z-z_0)^{n-1}.$$

**Proof.** Let $\rho < R$ be given and choose $r \in (\rho, R)$. Since $z = z_0 + r \in D(z_0,R)$, by assumption the series $\sum_{n=0}^\infty a_n r^n$ is convergent and in particular $M := \sup_n |a_n r^n| < \infty$. We now apply Corollary 9.30 with $X = \mathbb{N}\cup\{0\}$, $\mu$ being counting measure, $\Omega = D(z_0,\rho)$ and $g(z,n) := a_n(z-z_0)^n$. Since

$$\begin{aligned}|g'(z,n)| &= |n a_n (z-z_0)^{n-1}| \le n\,|a_n|\,\rho^{n-1}\\ &\le \frac{1}{r} n \left(\frac{\rho}{r}\right)^{n-1} |a_n|\,r^n \le \frac{1}{r} n \left(\frac{\rho}{r}\right)^{n-1} M\end{aligned}$$

and the function $G(n) := \frac{M}{r}n\left(\frac{\varrho}{r}\right)^{n-1}$ is summable (by the Ratio test for example), we may use $G$ as our dominating function. It then follows from Corollary 9.30

$$f(z) = \int_X g(z,n)d\mu(n) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$$

is complex differentiable with the differential given as in Eq. (9.10). ∎

**Definition 9.32 (Moment Generating Function).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The* ***moment generating function*** *of $X$ is $M_X : \mathbb{R} \to [0,\infty]$ defined by*

$$M_X(t) := \mathbb{E}\left[e^{tX}\right].$$

**Proposition 9.33.** *Suppose there exists $\varepsilon > 0$ such that $\mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty$, then $M_X(t)$ is a smooth function of $t \in (-\varepsilon, \varepsilon)$ and*

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbb{E}X^n \text{ if } |t| \le \varepsilon. \tag{9.11}$$

*In particular,*

$$\mathbb{E}X^n = \left(\frac{d}{dt}\right)^n |_{t=0} M_X(t) \text{ for all } n \in \mathbb{N}_0. \tag{9.12}$$

**Proof.** If $|t| \le \varepsilon$, then

$$\mathbb{E}\left[\sum_{n=0}^{\infty}\frac{|t|^n}{n!}|X|^n\right] \le \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{\varepsilon^n}{n!}|X|^n\right] = \mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty.$$

it $e^{tX} \le e^{\varepsilon|X|}$ for all $|t| \le \varepsilon$. Hence it follows from Corollary 9.28 that, for $|t| \le \varepsilon$,

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{t^n}{n!}X^n\right] = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}X^n.$$

Equation (9.12) now is a consequence of Corollary 9.31. ∎

**Exercise 9.3.** Let $d \in \mathbb{N}$, $\Omega = \mathbb{N}_0^d$, $\mathcal{B} = 2^\Omega$, $\mu : \mathcal{B} \to \mathbb{N}_0 \cup \{\infty\}$ be counting measure on $\Omega$, and for $x \in \mathbb{R}^d$ and $\omega \in \Omega$, let $x^\omega := x_1^{\omega_1} \dots x_n^{\omega_n}$. Further suppose that $f : \Omega \to \mathbb{C}$ is function and $r_i > 0$ for $1 \le i \le d$ such that

$$\sum_{\omega\in\Omega}|f(\omega)|r^\omega = \int_\Omega |f(\omega)|r^\omega d\mu(\omega) < \infty,$$

where $r := (r_1, \dots, r_d)$. Show;

1. There is a constant, $C < \infty$ such that $|f(\omega)| \le \frac{C}{r^\omega}$ for all $\omega \in \Omega$.
2. Let
$$U := \left\{x \in \mathbb{R}^d : |x_i| < r_i \ \forall\ i\right\} \text{ and } \bar{U} = \left\{x \in \mathbb{R}^d : |x_i| \le r_i \ \forall\ i\right\}$$
Show $\sum_{\omega\in\Omega}|f(\omega)x^\omega| < \infty$ for all $x \in \bar{U}$ and the function, $F : U \to \mathbb{R}$ defined by
$$F(x) = \sum_{\omega\in\Omega} f(\omega)x^\omega \text{ is continuous on } \bar{U}.$$
3. Show, for all $x \in U$ and $1 \le i \le d$, that
$$\frac{\partial}{\partial x_i}F(x) = \sum_{\omega\in\Omega}\omega_i f(\omega)x^{\omega - e_i}$$
where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ is the $i^{\text{th}}$ – standard basis vector on $\mathbb{R}^d$.
4. For any $\alpha \in \Omega$, let $\partial^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$ and $\alpha! := \prod_{i=1}^d \alpha_i!$ Explain why we may now conclude that
$$\partial^\alpha F(x) = \sum_{\omega\in\Omega}\alpha! f(\omega)x^{\omega-\alpha} \text{ for all } x \in U. \tag{9.13}$$
5. Conclude that $f(\alpha) = \frac{(\partial\alpha F)(0)}{\alpha!}$ for all $\alpha \in \Omega$.
6. If $g : \Omega \to \mathbb{C}$ is another function such that $\sum_{\omega\in\Omega} g(\omega)x^\omega = \sum_{\omega\in\Omega} f(\omega)x^\omega$ for $x$ in a neighborhood of $0 \in \mathbb{R}^d$, then $g(\omega) = f(\omega)$ for all $\omega \in \Omega$.

### 9.2.1 Square Integrable Random Variables and Correlations

Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space. We say that $X : \Omega \to \mathbb{R}$ is **integrable** if $X \in L^1(P)$ and **square integrable** if $X \in L^2(P)$. When $X$ is integrable we let $a_X := \mathbb{E}X$ be the **mean** of $X$.

Now suppose that $X, Y : \Omega \to \mathbb{R}$ are two square integrable random variables. Since

$$0 \le |X-Y|^2 = |X|^2 + |Y|^2 - 2|X||Y|,$$

it follows that

$$|XY| \le \frac{1}{2}|X|^2 + \frac{1}{2}|Y|^2 \in L^1(P).$$

In particular by taking $Y = 1$, we learn that $|X| \le \frac{1}{2}\left(1 + \left|X^2\right|\right)$ which shows that every square integrable random variable is also integrable.

**Definition 9.34.** *The* ***covariance,*** $\operatorname{Cov}(X,Y)$, *of two square integrable random variables, $X$ and $Y$, is defined by*

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X-a_X)(Y-a_Y)] = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

*where $a_X := \mathbb{E}X$ and $a_Y := \mathbb{E}Y$. The **variance** of $X$,*

$$\mathrm{Var}(X) := \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \tag{9.14}$$

*We say that $X$ and $Y$ are **uncorrelated** if $\mathrm{Cov}(X,Y) = 0$, i.e. $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n \subset L^2(P)$ are **uncorrelated** iff $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$.*

It follows from Eq. (9.14) that

$$\mathrm{Var}(X) \le \mathbb{E}\left[X^2\right] \text{ for all } X \in L^2(P). \tag{9.15}$$

**Lemma 9.35.** *The covariance function, $\mathrm{Cov}(X,Y)$ is bilinear in $X$ and $Y$ and $\mathrm{Cov}(X,Y) = 0$ if either $X$ or $Y$ is constant. For any constant $k$, $\mathrm{Var}(X+k) = \mathrm{Var}(X)$ and $\mathrm{Var}(kX) = k^2 \mathrm{Var}(X)$. If $\{X_k\}_{k=1}^n$ are uncorrelated $L^2(P)$ – random variables, then*

$$\mathrm{Var}(S_n) = \sum_{k=1}^n \mathrm{Var}(X_k).$$

**Proof.** We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove $\mathrm{Var}(X+k) = \mathrm{Var}(X)$;

$$\begin{aligned}\mathrm{Var}(X+k) &= \mathrm{Cov}(X+k, X+k) = \mathrm{Cov}(X+k, X) + \mathrm{Cov}(X+k, k) \\ &= \mathrm{Cov}(X+k, X) = \mathrm{Cov}(X,X) + \mathrm{Cov}(k, X) \\ &= \mathrm{Cov}(X,X) = \mathrm{Var}(X),\end{aligned}$$

wherein we have used the bilinearity of $\mathrm{Cov}(\cdot,\cdot)$ and the property that $\mathrm{Cov}(Y,k) = 0$ whenever $k$ is a constant. ■

**Exercise 9.4 (A Weak Law of Large Numbers).** Assume $\{X_n\}_{n=1}^\infty$ is a sequence if uncorrelated square integrable random variables which are identically distributed, i.e. $X_n \overset{d}{=} X_m$ for all $m, n \in \mathbb{N}$. Let $S_n := \sum_{k=1}^n X_k$, $\mu := \mathbb{E}X_k$ and $\sigma^2 := \mathrm{Var}(X_k)$ (these are independent of $k$). Show;

$$\begin{aligned}\mathbb{E}\left[\frac{S_n}{n}\right] &= \mu, \\ \mathbb{E}\left(\frac{S_n}{n} - \mu\right)^2 &= \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}, \text{ and} \\ P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) &\le \frac{\sigma^2}{n\varepsilon^2}\end{aligned}$$

for all $\varepsilon > 0$ and $n \in \mathbb{N}$. (Compare this with Exercise 5.19.)

### 9.2.2 Some Discrete Distributions

**Definition 9.36 (Generating Function).** *Suppose that $N : \Omega \to \mathbb{N}_0$ is an integer valued random variable on a probability space, $(\Omega, \mathcal{B}, P)$. The generating function associated to $N$ is defined by*

$$G_N(z) := \mathbb{E}\left[z^N\right] = \sum_{n=0}^\infty P(N=n) z^n \text{ for } |z| \le 1. \tag{9.16}$$

By Corollary 9.31, it follows that $P(N=n) = \frac{1}{n!} G_N^{(n)}(0)$ so that $G_N$ can be used to completely recover the distribution of $N$.

**Proposition 9.37 (Generating Functions).** *The generating function satisfies,*

$$G_N^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right] \text{ for } |z| < 1$$

*and*

$$G^{(k)}(1) = \lim_{z \uparrow 1} G^{(k)}(z) = \mathbb{E}[N(N-1)\dots(N-k+1)],$$

*where it is possible that one and hence both sides of this equation are infinite. In particular, $G'(1) := \lim_{z\uparrow 1} G'(z) = \mathbb{E}N$ and if $\mathbb{E}N^2 < \infty$,*

$$\mathrm{Var}(N) = G''(1) + G'(1) - [G'(1)]^2. \tag{9.17}$$

**Proof.** By Corollary 9.31 for $|z| < 1$,

$$\begin{aligned}G_N^{(k)}(z) &= \sum_{n=0}^\infty P(N=n) \cdot n(n-1)\dots(n-k+1) z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right].\end{aligned} \tag{9.18}$$

Since, for $z \in (0,1)$,

$$0 \le N(N-1)\dots(N-k+1) z^{N-k} \uparrow N(N-1)\dots(N-k+1) \text{ as } z \uparrow 1,$$

we may apply the MCT to pass to the limit as $z \uparrow 1$ in Eq. (9.18) to find,

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}[N(N-1)\dots(N-k+1)].$$

■

**Exercise 9.5 (Some Discrete Distributions).** Let $p \in (0,1]$ and $\lambda > 0$. In the four parts below, the distribution of $N$ will be described. You should work out the generating function, $G_N(z)$, in each case and use it to verify the given formulas for $\mathbb{E}N$ and $\mathrm{Var}(N)$.

1. Bernoulli($p$) : $P(N=1)=p$ and $P(N=0)=1-p$. You should find $\mathbb{E}N=p$ and $\operatorname{Var}(N)=p-p^2$.
2. Binomial($n,p$) : $P(N=k)=\binom{n}{k}p^k(1-p)^{n-k}$ for $k=0,1,\dots,n$. ($P(N=k)$ is the probability of $k$ successes in a sequence of $n$ independent yes/no experiments with probability of success being $p$.) You should find $\mathbb{E}N=np$ and $\operatorname{Var}(N)=n\left(p-p^2\right)$.
3. Geometric($p$) : $P(N=k)=p(1-p)^{k-1}$ for $k\in\mathbb{N}$. ($P(N=k)$ is the probability that the $k^{\text{th}}$ – trial is the first time of success out a sequence of independent trials with probability of success being $p$.) You should find $\mathbb{E}N=1/p$ and $\operatorname{Var}(N)=\frac{1-p}{p^2}$.
4. Poisson($\lambda$) : $P(N=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ for all $k\in\mathbb{N}_0$. You should find $\mathbb{E}N=\lambda=\operatorname{Var}(N)$.

**Exercise 9.6.** Let $S_{n,p}\overset{d}{=}\text{Binomial}(n,p)$, $k\in\mathbb{N}$, $p_n=\lambda_n/n$ where $\lambda_n\to\lambda>0$ as $n\to\infty$. Show that

$$\lim_{n\to\infty}P(S_{n,p_n}=k)=\frac{\lambda^k}{k!}e^{-\lambda}=P(\text{Poisson}(\lambda)=k).$$

Thus we see that for $p=O(1/n)$ and $k$ not too large relative to $n$ that for large $n$,

$$P(\text{Binomial}(n,p)=k)\cong P(\text{Poisson}(pn)=k)=\frac{(pn)^k}{k!}e^{-pn}.$$

(We will come back to the Poisson distribution and the related Poisson process later on.)

## 9.3 Integration on $\mathbb{R}$

**Notation 9.38** *If $m$ is Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$, $f$ is a non-negative Borel measurable function and $a<b$ with $a,b\in\bar{\mathbb{R}}$, we will often write $\int_a^b f(x)\,dx$ or $\int_a^b f dm$ for $\int_{(a,b]\cap\mathbb{R}} f dm$.*

*Example 9.39.* Suppose $-\infty<a<b<\infty$, $f\in C([a,b],\mathbb{R})$ and $m$ be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi=\{a=a_0<a_1<\dots<a_n=b\},$$

let

$$\operatorname{mesh}(\pi):=\max\{|a_j-a_{j-1}|:j=1,\dots,n\}$$

and

$$f_\pi(x):=\sum_{l=0}^{n-1}f(a_l)\,1_{(a_l,a_{l+1}]}(x).$$

Then

$$\int_a^b f_\pi\,dm=\sum_{l=0}^{n-1}f(a_l)\,m((a_l,a_{l+1}])=\sum_{l=0}^{n-1}f(a_l)(a_{l+1}-a_l)$$

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty}\operatorname{mesh}(\pi_k)=0$, we know that

$$\lim_{k\to\infty}\int_a^b f_{\pi_k}\,dm=\int_a^b f(x)\,dx \tag{9.19}$$

where the latter integral is the Riemann integral. Using the (uniform) continuity of $f$ on $[a,b]$, it easily follows that $\lim_{k\to\infty}f_{\pi_k}(x)=f(x)$ and that $|f_{\pi_k}(x)|\le g(x):=M1_{(a,b]}(x)$ for all $x\in(a,b]$ where $M:=\max_{x\in[a,b]}|f(x)|<\infty$. Since $\int_{\mathbb{R}}g dm=M(b-a)<\infty$, we may apply D.C.T. to conclude,

$$\lim_{k\to\infty}\int_a^b f_{\pi_k}\,dm=\int_a^b\lim_{k\to\infty}f_{\pi_k}\,dm=\int_a^b f\,dm.$$

This equation with Eq. (9.19) shows

$$\int_a^b f\,dm=\int_a^b f(x)\,dx$$

whenever $f\in C([a,b],\mathbb{R})$, i.e. the Lebesgue and the Riemann integral agree on continuous functions. See Theorem 9.82 below for a more general statement along these lines.

**Theorem 9.40 (The Fundamental Theorem of Calculus).** *Suppose $-\infty<a<b<\infty$, $f\in C((a,b),\mathbb{R})\cap L^1((a,b),m)$ and $F(x):=\int_a^x f(y)dm(y)$. Then*

1. *$F\in C([a,b],\mathbb{R})\cap C^1((a,b),\mathbb{R})$.*
2. *$F'(x)=f(x)$ for all $x\in(a,b)$.*
3. *If $G\in C([a,b],\mathbb{R})\cap C^1((a,b),\mathbb{R})$ is an anti-derivative of $f$ on $(a,b)$ (i.e. $f=G'|_{(a,b)}$) then*
$$\int_a^b f(x)dm(x)=G(b)-G(a).$$

**Proof.** Since $F(x):=\int_{\mathbb{R}}1_{(a,x)}(y)f(y)dm(y)$, $\lim_{x\to z}1_{(a,x)}(y)=1_{(a,z)}(y)$ for $m$ – a.e. $y$ and $\left|1_{(a,x)}(y)f(y)\right|\le 1_{(a,b)}(y)\,|f(y)|$ is an $L^1$ – function, it follows from the dominated convergence Theorem 9.27 that $F$ is continuous on $[a,b]$. Simple manipulations show,

$$\left|\frac{F(x+h)-F(x)}{h}-f(x)\right| = \frac{1}{|h|}\begin{cases}\left|\int_x^{x+h}[f(y)-f(x)]\,dm(y)\right| & \text{if } h>0\\ \left|\int_{x+h}^{x}[f(y)-f(x)]\,dm(y)\right| & \text{if } h<0\end{cases}$$
$$\le \frac{1}{|h|}\begin{cases}\int_x^{x+h}|f(y)-f(x)|\,dm(y) & \text{if } h>0\\ \int_{x+h}^{x}|f(y)-f(x)|\,dm(y) & \text{if } h<0\end{cases}$$
$$\le \sup\{|f(y)-f(x)| : y\in[x-|h|, x+|h|]\}$$

and the latter expression, by the continuity of $f$, goes to zero as $h\to 0$ . This shows $F'=f$ on $(a,b)$.

For the converse direction, we have by assumption that $G'(x)=F'(x)$ for $x\in(a,b)$. Therefore by the mean value theorem, $F-G=C$ for some constant $C$. Hence

$$\int_a^b f(x)dm(x) = F(b) = F(b)-F(a)$$
$$=(G(b)+C)-(G(a)+C)=G(b)-G(a).$$

■

We can use the above results to integrate some non-Riemann integrable functions:

*Example 9.41.* For all $\lambda>0$,

$$\int_0^\infty e^{-\lambda x}dm(x)=\lambda^{-1} \text{ and } \int_{\mathbb{R}}\frac{1}{1+x^2}dm(x)=\pi.$$

The proof of these identities are similar. By the monotone convergence theorem, Example 9.39 and the fundamental theorem of calculus for Riemann integrals (or Theorem 9.40 below),

$$\int_0^\infty e^{-\lambda x}dm(x)=\lim_{N\to\infty}\int_0^N e^{-\lambda x}dm(x)=\lim_{N\to\infty}\int_0^N e^{-\lambda x}dx$$
$$=-\lim_{N\to\infty}\frac{1}{\lambda}e^{-\lambda x}|_0^N=\lambda^{-1}$$

and

$$\int_{\mathbb{R}}\frac{1}{1+x^2}dm(x)=\lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dm(x)=\lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dx$$
$$=\lim_{N\to\infty}\left[\tan^{-1}(N)-\tan^{-1}(-N)\right]=\pi.$$

Let us also consider the functions $x^{-p}$. Using the MCT and the fundamental theorem of calculus,

$$\int_{(0,1]}\frac{1}{x^p}\,dm(x)=\lim_{n\to\infty}\int_0^1 1_{(\frac{1}{n},1]}(x)\frac{1}{x^p}dm(x)$$
$$=\lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x^p}dx=\lim_{n\to\infty}\left.\frac{x^{-p+1}}{1-p}\right|_{1/n}^1$$
$$=\begin{cases}\frac{1}{1-p} & \text{if } p<1\\ \infty & \text{if } p>1\end{cases}$$

If $p=1$ we find

$$\int_{(0,1]}\frac{1}{x^p}\,dm(x)=\lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x}dx=\lim_{n\to\infty}\ln(x)|_{1/n}^1=\infty.$$

**Exercise 9.7.** Show

$$\int_1^\infty\frac{1}{x^p}dm\,(x)=\begin{cases}\infty & \text{if } p\le 1\\ \frac{1}{p-1} & \text{if } p>1\end{cases}.$$

*Example 9.42 (Integration of Power Series).* Suppose $R>0$ and $\{a_n\}_{n=0}^\infty$ is a sequence of complex numbers such that $\sum_{n=0}^\infty|a_n|\,r^n<\infty$ for all $r\in(0,R)$. Then

$$\int_\alpha^\beta\left(\sum_{n=0}^\infty a_nx^n\right)dm(x)=\sum_{n=0}^\infty a_n\int_\alpha^\beta x^ndm(x)=\sum_{n=0}^\infty a_n\frac{\beta^{n+1}-\alpha^{n+1}}{n+1}$$

for all $-R<\alpha<\beta<R$. Indeed this follows from Corollary 9.28 since

$$\sum_{n=0}^\infty\int_\alpha^\beta|a_n|\,|x|^n\,dm(x)\le\sum_{n=0}^\infty\left(\int_0^{|\beta|}|a_n|\,|x|^n\,dm(x)+\int_0^{|\alpha|}|a_n|\,|x|^n\,dm(x)\right)$$
$$\le\sum_{n=0}^\infty|a_n|\,\frac{|\beta|^{n+1}+|\alpha|^{n+1}}{n+1}\le 2r\sum_{n=0}^\infty|a_n|\,r^n<\infty$$

where $r=\max(|\beta|,|\alpha|)$.

*Example 9.43.* Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q}\cap[0,1]$ and define

$$f(x)=\sum_{n=1}^\infty 2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x-r_n|}}=5 \text{ if } x=r_n.$$

Since, By Theorem 9.40,

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{|x-r_n|}}\,dx &= \int_{r_n}^1 \frac{1}{\sqrt{x-r_n}}dx + \int_0^{r_n} \frac{1}{\sqrt{r_n - x}}dx \\ &= 2\sqrt{x-r_n}|_{r_n}^1 - 2\sqrt{r_n - x}|_0^{r_n} = 2\left(\sqrt{1-r_n} - \sqrt{r_n}\right) \\ &\le 4,\end{aligned}$$

we find

$$\int_{[0,1]} f(x)dm(x) = \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x-r_n|}}dx \le \sum_{n=1}^{\infty} 2^{-n}4 = 4 < \infty.$$

In particular, $m(f=\infty)=0$, i.e. that $f<\infty$ for almost every $x \in [0,1]$ and this implies that

$$\sum_{n=1}^{\infty} 2^{-n} \frac{1}{\sqrt{|x-r_n|}} < \infty \text{ for a.e. } x \in [0,1].$$

This result is somewhat surprising since the singularities of the summands form a dense subset of $[0,1]$.

*Example 9.44.* The following limit holds,

$$\lim_{n\to\infty} \int_0^n \left(1-\frac{x}{n}\right)^n dm(x) = 1. \tag{9.20}$$

**DCT Proof.** To verify this, let $f_n(x) := \left(1-\frac{x}{n}\right)^n 1_{[0,n]}(x)$. Then $\lim_{n\to\infty} f_n(x) = e^{-x}$ for all $x \ge 0$. Moreover by simple calculus[2]

$$1-x \le e^{-x} \text{ for all } x \in \mathbb{R}.$$

Therefore, for $x<n$, we have

$$0 \le 1-\frac{x}{n} \le e^{-x/n} \implies \left(1-\frac{x}{n}\right)^n \le \left[e^{-x/n}\right]^n = e^{-x},$$

from which it follows that

$$0 \le f_n(x) \le e^{-x} \text{ for all } x \ge 0.$$

From Example 9.41, we know

$$\int_0^\infty e^{-x}dm(x) = 1 < \infty,$$

so that $e^{-x}$ is an integrable function on $[0,\infty)$. Hence by the dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty} \int_0^n \left(1-\frac{x}{n}\right)^n dm(x) &= \lim_{n\to\infty} \int_0^\infty f_n(x)dm(x) \\ &= \int_0^\infty \lim_{n\to\infty} f_n(x)dm(x) = \int_0^\infty e^{-x}dm(x) = 1.\end{aligned}$$

**MCT Proof.** The limit in Eq. (9.20) may also be computed using the monotone convergence theorem. To do this we must show that $n \to f_n(x)$ is increasing in $n$ for each $x$ and for this it suffices to consider $n > x$. But for $n > x$,

$$\begin{aligned}\frac{d}{dn}\ln f_n(x) &= \frac{d}{dn}\left[n\ln\left(1-\frac{x}{n}\right)\right] = \ln\left(1-\frac{x}{n}\right) + \frac{n}{1-\frac{x}{n}}\frac{x}{n^2} \\ &= \ln\left(1-\frac{x}{n}\right) + \frac{\frac{x}{n}}{1-\frac{x}{n}} = h(x/n)\end{aligned}$$

where, for $0 \le y < 1$,

$$h(y) := \ln(1-y) + \frac{y}{1-y}.$$

Since $h(0)=0$ and

$$h'(y) = -\frac{1}{1-y} + \frac{1}{1-y} + \frac{y}{(1-y)^2} > 0$$

it follows that $h \ge 0$. Thus we have shown, $f_n(x) \uparrow e^{-x}$ as $n\to\infty$ as claimed.

*Example 9.45.* Suppose that $f_n(x) := n1_{(0,\frac{1}{n}]}(x)$ for $n \in \mathbb{N}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$ while

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,dx = \lim_{n\to\infty} 1 = 1 \ne 0 = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\,dx.$$

The problem is that the best dominating function we can take is

$$g(x) = \sup_n f_n(x) = \sum_{n=1}^{\infty} n\cdot 1_{\left(\frac{1}{n+1},\frac{1}{n}\right]}(x).$$

Notice that

$$\int_{\mathbb{R}} g(x)\,dx = \sum_{n=1}^{\infty} n\cdot\left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty.$$

[2] Since $y = 1-x$ is the tangent line to $y = e^{-x}$ at $x=0$ and $e^{-x}$ is convex up, it follows that $1-x \le e^{-x}$ for all $x \in \mathbb{R}$.

*Example 9.46 (Jordan's Lemma).* In this example, let us consider the limit;

$$\lim_{n\to\infty}\int_0^\pi \cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta.$$

Let

$$f_n(\theta) := 1_{(0,\pi]}(\theta)\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}.$$

Then

$$|f_n| \le 1_{(0,\pi]} \in L^1(m)$$

and

$$\lim_{n\to\infty} f_n(\theta) = 1_{(0,\pi]}(\theta)\,1_{\{\pi\}}(\theta) = 1_{\{\pi\}}(\theta).$$

Therefore by the D.C.T.,

$$\lim_{n\to\infty}\int_0^\pi \cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta = \int_{\mathbb{R}} 1_{\{\pi\}}(\theta)\,dm(\theta) = m(\{\pi\}) = 0.$$

*Example 9.47.* Recall from Example 9.41 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x}dm(x) \text{ for all } \lambda > 0.$$

Let $\varepsilon > 0$. For $\lambda \ge 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

$$0 \le \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} = x^n e^{-\lambda x} \le C_n(\varepsilon)e^{-\varepsilon x}.$$

Using this fact, Corollary 9.30 and induction gives

$$n!\lambda^{-n-1} = \left(-\frac{d}{d\lambda}\right)^n \lambda^{-1} = \int_{[0,\infty)}\left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x}dm(x)$$
$$= \int_{[0,\infty)} x^n e^{-\lambda x}dm(x).$$

That is

$$n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x}dm(x). \tag{9.21}$$

*Remark 9.48.* Corollary 9.30 may be generalized by allowing the hypothesis to hold for $x \in X \setminus E$ where $E \in \mathcal{B}$ is a **fixed** null set, i.e. $E$ must be independent of $t$. Consider what happens if we formally apply Corollary 9.30 to $g(t) := \int_0^\infty 1_{x\le t}dm(x)$,

$$\dot{g}(t) = \frac{d}{dt}\int_0^\infty 1_{x\le t}dm(x) \overset{?}{=} \int_0^\infty \frac{\partial}{\partial t}1_{x\le t}dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t}1_{x\le t} = 0$ unless $t = x$ in which case it is not defined. On the other hand $g(t) = t$ so that $\dot{g}(t) = 1$. (The reader should decide which hypothesis of Corollary 9.30 has been violated in this example.)

**Exercise 9.8 (Folland 2.28 on p. 60.).** Compute the following limits and justify your calculations:

1. $\lim_{n\to\infty}\int_0^\infty \frac{\sin(\frac{x}{n})}{(1+\frac{x}{n})^n}\,dx.$
2. $\lim_{n\to\infty}\int_0^1 \frac{1+nx^2}{(1+x^2)^n}\,dx$
3. $\lim_{n\to\infty}\int_0^\infty \frac{n\sin(x/n)}{x(1+x^2)}\,dx$
4. For all $a \in \mathbb{R}$ compute,

$$f(a) := \lim_{n\to\infty}\int_a^\infty n(1+n^2x^2)^{-1}dx.$$

**Exercise 9.9 (Integration by Parts).** Suppose that $f, g : \mathbb{R} \to \mathbb{R}$ are two continuously differentiable functions such that $f'g$, $fg'$, and $fg$ are all Lebesgue integrable functions on $\mathbb{R}$. Prove the following integration by parts formula;

$$\int_{\mathbb{R}} f'(x)\cdot g(x)\,dx = -\int_{\mathbb{R}} f(x)\cdot g'(x)\,dx. \tag{9.22}$$

Similarly show that if Suppose that $f, g : [0,\infty) \to [0,\infty)$ are two continuously differentiable functions such that $f'g$, $fg'$, and $fg$ are all Lebesgue integrable functions on $[0,\infty)$, then

$$\int_0^\infty f'(x)\cdot g(x)\,dx = -f(0)\,g(0) - \int_0^\infty f(x)\cdot g'(x)\,dx. \tag{9.23}$$

**Outline:** 1. First notice that Eq. (9.22) holds if $f(x) = 0$ for $|x| \ge N$ for some $N < \infty$ by undergraduate calculus.

2. Let $\psi : \mathbb{R} \to [0,1]$ be a continuously differentiable function such that $\psi(x) = 1$ if $|x| \le 1$ and $\psi(x) = 0$ if $|x| \ge 2$. For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ Write out the identity in Eq. (9.22) with $f(x)$ being replaced by $f(x)\psi_\varepsilon(x)$.

3. Now use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in the identity you found in step 2.

4. A similar outline works to prove Eq. (9.23).

**Definition 9.49 (Gamma Function).** *The* ***Gamma function****,* $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ *is defined by*

$$\Gamma(x) := \int_0^\infty u^{x-1}e^{-u}du \tag{9.24}$$

*(The reader should check that* $\Gamma(x) < \infty$ *for all* $x > 0$*.)*

Here are some of the more basic properties of this function.

*Example 9.50 ($\Gamma$ – function properties).* Let $\Gamma$ be the gamma function, then;

1. $\Gamma(1) = 1$ as is easily verified.
2. $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$ as follows by integration by parts;
$$\Gamma(x+1) = \int_0^\infty e^{-u}\ u^{x+1}\ \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\ e^{-u}\right) du$$
$$= x \int_0^\infty u^{x-1}\ e^{-u}\ du = x\ \Gamma(x).$$
In particular, it follows from items 1. and 2. and induction that
$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{9.25}$$
(Equation (9.25) was also proved in Eq. (9.21).)
3. $\Gamma(1/2) = \sqrt{\pi}$. This last assertion is a bit trickier. One proof is to make use of the fact (proved below in Lemma 11.29) that
$$\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{9.26}$$
Taking $a = 1$ and making the change of variables, $u = r^2$ below implies,
$$\sqrt{\pi} = \int_{-\infty}^\infty e^{-r^2} dr = 2\int_0^\infty u^{-1/2} e^{-u} du = \Gamma(1/2).$$
$$\Gamma(1/2) = 2\int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr$$
$$= I_1(1) = \sqrt{\pi}.$$
4. A simple induction argument using items 2. and 3. now shows that
$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi}$$
where $(-1)!! := 1$ and $(2n-1)!! = (2n-1)(2n-3)\dots 3 \cdot 1$ for $n \in \mathbb{N}$.Letting the mesh of $\Pi$ tend to zero using the uniform continuity of $f$ then shows $\lambda(f) = \mu(f)$.

## 9.4 Densities and Change of Variables Theorems

**Exercise 9.10 (Measures and Densities).** Let $(X, \mathcal{M}, \mu)$ be a measure space and $\rho : X \to [0, \infty]$ be a measurable function. For $A \in \mathcal{M}$, set $\nu(A) := \int_A \rho d\mu$.

1. Show $\nu : \mathcal{M} \to [0, \infty]$ is a measure.
2. Let $f : X \to [0, \infty]$ be a measurable function, show
$$\int_X f d\nu = \int_X f \rho d\mu. \tag{9.27}$$
**Hint:** first prove the relationship for characteristic functions, then for simple functions, and then for general positive measurable functions.
3. Show that a measurable function $f : X \to \mathbb{C}$ is in $L^1(\nu)$ iff $|f|\rho \in L^1(\mu)$ and if $f \in L^1(\nu)$ then Eq. (9.27) still holds.

**Notation 9.51** *It is customary to informally describe $\nu$ defined in Exercise 9.10 by writing $d\nu = \rho d\mu$.*

**Exercise 9.11 (Abstract Change of Variables Formula).** Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Recall that $\nu = f_*\mu : \mathcal{F} \to [0, \infty]$ defined by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$ is a measure on $\mathcal{F}$.

1. Show
$$\int_Y g d\nu = \int_X (g \circ f) d\mu \tag{9.28}$$
for all measurable functions $g : Y \to [0, \infty]$. **Hint:** see the hint from Exercise 9.10.
2. Show a measurable function $g : Y \to \mathbb{C}$ is in $L^1(\nu)$ iff $g \circ f \in L^1(\mu)$ and that Eq. (9.28) holds for all $g \in \mathrm{L}^1(\nu)$.

*Example 9.52.* Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{X_i\}_{i=1}^n$ are random variables on $\Omega$ with $\nu := \mathrm{Law}_P(X_1, \dots, X_n)$, then
$$\mathbb{E}[g(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g\ d\nu$$
for all $g : \mathbb{R}^n \to \mathbb{R}$ which are Borel measurable and either bounded or non-negative. This follows directly from Exercise 9.11 with $f := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\mu = P$.

*Remark 9.53.* As a special case of Example 9.52, suppose that $X$ is a random variable on a probability space, $(\Omega, \mathcal{B}, P)$, and $F(x) := P(X \le x)$. Then

$$\mathbb{E}\left[f\left(X\right)\right]=\int_{\mathbb{R}}f\left(x\right)dF\left(x\right) \tag{9.29}$$

where $dF(x)$ is shorthand for $d\mu_F(x)$ and $\mu_F$ is the unique probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\mu_F((-\infty,x])=F(x)$ for all $x\in\mathbb{R}$. Moreover if $F:\mathbb{R}\to[0,1]$ happens to be $C^1$-function, then

$$d\mu_F\left(x\right)=F'\left(x\right)dm\left(x\right) \tag{9.30}$$

and Eq. (9.29) may be written as

$$\mathbb{E}\left[f\left(X\right)\right]=\int_{\mathbb{R}}f\left(x\right)F'\left(x\right)dm\left(x\right). \tag{9.31}$$

To verify Eq. (9.30) it suffices to observe, by the fundamental theorem of calculus, that

$$\mu_F\left((a,b]\right)=F\left(b\right)-F\left(a\right)=\int_a^bF'\left(x\right)dx=\int_{(a,b]}F'dm.$$

From this equation we may deduce that $\mu_F(A)=\int_AF'dm$ for all $A\in\mathcal{B}_{\mathbb{R}}$. Equation 9.31 now follows from Exercise 9.10.

**Exercise 9.12.** Let $F:\mathbb{R}\to\mathbb{R}$ be a $C^1$-function such that $F'(x)>0$ for all $x\in\mathbb{R}$ and $\lim_{x\to\pm\infty}F(x)=\pm\infty$. (Notice that $F$ is strictly increasing so that $F^{-1}:\mathbb{R}\to\mathbb{R}$ exists and moreover, by the inverse function theorem that $F^{-1}$ is a $C^1$ – function.) Let $m$ be Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$ and

$$\nu(A)=m(F(A))=m(\left(F^{-1}\right)^{-1}(A))=\left(F_*^{-1}m\right)(A)$$

for all $A\in\mathcal{B}_{\mathbb{R}}$. Show $d\nu=F'dm$. Use this result to prove the change of variable formula,

$$\int_{\mathbb{R}}h\circ F\cdot F'dm=\int_{\mathbb{R}}hdm \tag{9.32}$$

which is valid for all Borel measurable functions $h:\mathbb{R}\to[0,\infty]$.

**Hint:** Start by showing $d\nu=F'dm$ on sets of the form $A=(a,b]$ with $a,b\in\mathbb{R}$ and $a<b$. Then use the uniqueness assertions in Exercise 6.3 to conclude $d\nu=F'dm$ on all of $\mathcal{B}_{\mathbb{R}}$. To prove Eq. (9.32) apply Exercise 9.11 with $g=h\circ F$ and $f=F^{-1}$.

## 9.5 Riesz Markov Theorem for $[0,1]^d$ and $\mathbb{R}^d$

Before getting to the main theorem let us begin with a few basic generalities. For the moment suppose $(X,\rho)$ be a metric space. For example, it will be sufficient to take $X$ to be a subset of $\mathbb{R}^d$ and $\rho$ to be the usual Euclidean metric. Let $C_c(X)$ denote the compactly supported real valued continuous functions on $X$. If $X$ is compact we simply write $C(X)$ for $C_c(X)$. For $f\in C_c(X)$ we let $\|f\|_\infty:=\max_{x\in X}|f(x)|$.

**Definition 9.54.** *A positive linear functional, $I$, on $C_c(X)$ is a linear map, $I:C_c(X)\to\mathbb{R}$ such that $I(f)\ge0$ whenever $f\ge0$.*

*Example 9.55.* If $\mu$ is any measure on $(X,\mathcal{B}_X)$ with $\mu(K)<\infty$ for all compact subset $K\subset X$, then $I(f):=\mu(f):=\int_Xfd\mu$ is a positive linear functional on $C_c(X)$. The Riesz Markov theorem states under certain assumptions on $(X,\rho)$ that the converse result is true as well.

**Proposition 9.56.** *If $I$ is a positive linear functional on $C_c(X)$ and $K$ is a compact subset of $X$, then there exists $C_K<\infty$ such that $|I(f)|\le C_K\|f\|_\infty$ for all $f\in C_c(X)$ with* $\operatorname{supp}(f)\subset K$.

**Proof.** By Urysohn's Lemma for metric spaces (see Lemma 2.22 or also **??**), there exists $\varphi\in C_c(X,[0,1])$ such that $\varphi=1$ on $K$. Then for all $f\in C_c(X,\mathbb{R})$ such that $\operatorname{supp}(f)\subset K$, $|f|\le\|f\|_\infty\varphi$ or equivalently $\|f\|_\infty\varphi\pm f\ge0$. Hence $\|f\|_\infty I(\varphi)\pm I(f)\ge0$ or equivalently, $|I(f)|\le\|f\|_\infty I(\varphi)$. Letting $C_K:=I(\varphi)$, we have shown that $|I(f)|\le C_K\|f\|_\infty$ for all $f\in C_c(X,\mathbb{R})$ with $\operatorname{supp}(f)\subset K$. For general $f\in C_c(X,\mathbb{C})$ with $\operatorname{supp}(f)\subset K$, choose $|\alpha|=1$ such that $\alpha I(f)\ge0$. Then

$$|I\left(f\right)|=\alpha I(f)=I(\alpha\ f)=I(\operatorname{Re}(\alpha f))\le C_K\|\operatorname{Re}\left(\alpha f\right)\|_\infty\le C_K\|f\|_\infty.$$

■

**Notation 9.57** *Let $C(X)_\downarrow=\{f_\infty:X\to[-\infty,\infty)|\ \exists\ f_n\in C(X)\ \ni\ f_n\downarrow f_\infty\}$ where $f_n\downarrow f_\infty$ means $f_n(x)\ge f_{n+1}(x)$ for all $n\in\mathbb{N}$ and $x\in X$ and $f_\infty(x)=\lim_{n\to\infty}f_n(x)\in[-\infty,\infty)$ for all $x\in X$.*

*Example 9.58.* If $F$ is a closed subset of $X$, then $1_F\in C(X)_\downarrow$. To prove this first observe that $\rho_F(x):=\inf\{\rho(x,y):y\in F\}\ge0$ is a continuous function such that $\rho_F(x)=0$ iff $x\in F$. Now let $h_n:[0,\infty)\to[0,1]$ be defined by $h_n(t)=\max\{0,1-nt\}$ so that $h_n\downarrow1_{\{0\}}$. It then follows that $f_n:=h_n\circ\rho_F\in C(X)$ with $f_n\downarrow1_{\{0\}}\circ\rho_F=1_F$.

**Theorem 9.59.** *Suppose that $X$ is compact and $\lambda$ is a positive linear functional on $C(X)$. Then $\lambda$ satisfies;*

1. *If $f_n,f\in C(X)$ and $f_n\downarrow f_\infty\in C(X)_\downarrow$ with $f_\infty\le f$, then $\lim_{n\to\infty}\lambda(f_n)\le\lambda(f)$.*
2. *If $f_n,g_n\in C(X)$ such that $f_n\downarrow f_\infty$ and $g_n\downarrow g_\infty$ (pointwise) with $f_\infty\le g_\infty$, then $\lim_{n\to\infty}\lambda(f_n)\le\lim_{n\to\infty}\lambda(g_n)$. In particular if $f_n\downarrow f_\infty$ and $g_n\downarrow f_\infty$, then $\lim_{n\to\infty}$ then $\lim_{n\to\infty}\lambda(f_n)=\lim_{n\to\infty}\lambda(g_n)$.*

3. *Because of item 2. we may extend $\lambda$ to a function on $C(X)_\downarrow$ by setting $\lambda(f_\infty) := \lim_{n\to\infty}\lambda(f_n)$ whenever $f_n \in C(X)$ with $f_n \downarrow f_\infty$. The extension satisfies;*
   *a) $\lambda(f) \le \lambda(g)$ for all $f,g \in C(X)_\downarrow$ with $f \le g$ and*
   *b) for all $f,g \in C(X)_\downarrow$ and $a \ge 0$, $f + ag \in C(X)_\downarrow$ and*
$$\lambda(f+ag) = \lambda(f) + a\lambda(g). \tag{9.33}$$

**Proof.** We take each item in turn.

1. If $f_n, f \in C(X)$ and $f_\infty \in C(X)_\downarrow$ are as in item 1., then $f_n \le f_n \vee f \downarrow f$ and so by Dini's theorem it follows that $f_n \vee f \to f$ uniformly on $X$. Because $\lambda$ is a bounded linear functional (see Propositions 9.56) it follows that
$$\lim_{n\to\infty}\lambda(f_n) \le \lim_{n\to\infty}\lambda(f_n \vee f) = \lambda(f).$$
2. If $f_n, g_n \in C(X)$ and $f_\infty, g_\infty \in C(X)_\downarrow$ are as in item 2., then for each $m \in \mathbb{N}$,
$$f_n \le f_n \vee g_m \downarrow f \vee g_m = g_m \text{ as } n \to \infty.$$
Hence by item 1., $\lim_{n\to\infty}\lambda(f_n) \le \lambda(g_m)$ and the result follows by letting $m \to \infty$ in this last inequality.
3. The monotinicity assertion of item 3a. follows directly from item 2. The proof 3b. is straight forward and will be left to the reader.

■

*Remark 9.60.* Because of Theorem 9.59 and Example 9.58, if $(X,\rho)$ is a compact metric space and $\lambda \in C(X)^*$ is a positive linear functional, then $\lambda(1_F)$ is well defined for all $F$ closed subsets, $F \subset X$. Moreover if $F_1, F_2$ are closed subsets of $X$ such that $F_1 \subset F_2$, then $\lambda(1_{F_1}) \le \lambda(1_{F_2})$. It is fact true it this level of generality that there exists a Borel measure, $\mu$, on $(X, \mathcal{B}_X)$ such that $\mu(F) = \lambda(1_F)$ for closed subsets $F \subset X$. In this section we are going to verify this assertion when $X = [0,1]^d$ for $d \in \mathbb{N}$.

**Definition 9.61.** *To each $b \in \mathbb{R}$ and $\varepsilon > 0$ let $\varphi_{b,\varepsilon} \in C(\mathbb{R},[0,1])$ be defined by*
$$\varphi_{b,\varepsilon}(x) = \begin{cases} 1 & \text{if } \quad x \le b \\ 1-(x-b)/\varepsilon & \text{if } b \le x \le b+\varepsilon \\ 0 & \text{if } \quad x \ge b+\varepsilon. \end{cases} \tag{9.34}$$

The key property of these functions are that $\varphi_{b,\varepsilon} \downarrow 1_{(-\infty,b]}$ as $\varepsilon \downarrow 0$.

**Theorem 9.62 (Riesz Markov Theorem for an Interval).** *Let $X = [0,1]$ and $\lambda \in C(X)^*$ be a positive linear functional. Then there exists a unique Borel measure, $\mu$, on $\mathcal{B}_X$ such that*
$$\lambda(f) = \int_0^1 f d\mu \text{ for all } f \in C(X). \tag{9.35}$$

**Proof.** The uniqueness is an easy consequence of the multiplicative systems theorem and hence we will concentrate on existence. We break the existence proof up into a number of steps.

1. For $b \in [0,1]$, let $F(b) := \lambda\left(1_{[0,b]}\right)$ which is a well defined non-decreasing function on $[0,1]$ by Remark 9.60. It is also true that $F$ is right continuous. To prove this choose $\delta_n$ so that
$$0 < \delta_n < \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}. \tag{9.36}$$
Then
$$1_{[0,b+\frac{1}{n}]} \le f_n := \varphi_{b+\frac{1}{n},\delta_n}|_{[0,1]} \downarrow 1_{[0,b]} \text{ as } n \to \infty$$
and hence
$$F(b+) = \lim_{n\to\infty} F\left(b+\frac{1}{n}\right) \le \lim_{n\to\infty}\lambda(f_n) = F(b) \le F(b+).$$
2. By item 1. and Corollary 6.59, there exists a unique measure $\mu$ on $\mathcal{B}_{[0,1]}$ such that $F(b) = \mu([0,b])$ for all $b \in [0,1]$.
3. If $a,b \in [0,1]$ with $a < b$, then $\lambda\left(1_{[a,b]}\right) \le \mu([a,b])$. Since
$$\mu([0,b]) = F(b) = \lambda\left(1_{[0,b]}\right)$$
we may assume that $0 < a \le b \le 1$. For any $0 < \tilde{a} < a$ we have,
$$1_{[0,\tilde{a}]} + 1_{[a,b]} \le 1_{[0,b]}$$
and therefore
$$F(\tilde{a}) + \lambda\left(1_{[a,b]}\right) = \lambda\left(1_{[0,\tilde{a}]} + 1_{[a,b]}\right) \le \lambda\left(1_{[0,b]}\right) = F(b).$$
From this equation it follows that
$$\lambda\left(1_{[a,b]}\right) \le F(b) - F(\tilde{a}) = \mu((\tilde{a},b]).$$
Letting $\tilde{a} \uparrow a$ gives the desired result, namely that $\lambda\left(1_{[a,b]}\right) \le \mu([a,b])$.
4. Since $\sum_{x\in[0,1]}\mu(\{x\}) \le \mu([0,1]) < \infty$, it follows that $U := \{x \in [0,1] : \mu(\{x\}) > 0\}$ is at most a countable set.
5. (**Verification of** Eq. (9.35).) First assume $f \in C(X)$ is non-negative, i.e. $f \ge 0$. Let
$$\pi = \{0 = a_0 < a_1 < \cdots < a_n = 1\}$$
be a partition of $[0,1]$ where we always assume that $a_j \in U^c$ for $1 \le j < n$. Let
$$c_i^\pi := \max\{|f(x)| : a_i \le x \le a_{i+1}\} \text{ for } 0 \le i < n,$$

and define $f_\pi$ and $\hat{f}_\pi$ by

$$f_\pi = c_0 1_{[0,a_1]} + \sum_{i=1}^{n-1} c_i 1_{(a_i,a_{i+1}]} \text{ and}$$
$$\hat{f}_\pi = c_0 1_{[0,a_1]} + \sum_{i=1}^{n-1} c_i 1_{[a_i,a_{i+1}]} \in C([0,1])_\downarrow .$$

It is easy to see that $f \le f_\pi \le \hat{f}_\pi$ and that $f_\pi \to f$ uniformly on $X$. Hence it follows that

$$\begin{aligned}\lambda(f) \le \lambda\left(\hat{f}_\pi\right) &= c_0\lambda\left(1_{[0,a_1]}\right) + \sum_{i=1}^{n-1} c_i \lambda\left(1_{[a_i,a_{i+1}]}\right) \\ &\le c_0\mu([0,a_1]) + \sum_{i=1}^{n-1} c_i\mu([a_i,a_{i+1}]) \\ &= c_0\mu([0,a_1]) + \sum_{i=1}^{n-1} c_i\mu((a_i,a_{i+1}]) \\ &= \mu(f_\pi) \to \mu(f) \text{ as mesh}(\pi) \to 0\end{aligned}$$

wherein we have used $\mu(\{a_i\}) = 0$ for all $1 \le i < n$ in the second to last inequality and the dominated convergence theorem to take the limit.
Thus we have shown that $\lambda(f) \le \mu(f)$ for all $f \ge 0$. Since $\lambda(1) = \mu(1)$, if $f \in C(X)$ and $M = \|f\|_\infty$ we find that

$$M\lambda(1) + \lambda(f) = \lambda(f+M) \le \mu(f+M) = \mu(f) + M\mu(1)$$

and thus $\lambda(f) \le \mu(f)$ for all $f \in C(X)$. Replacing $f$ by $-f$ implies that $\mu(f) \le \lambda(f)$ and therefore that $\lambda(f) = \mu(f)$.

■

The next theorem is the multi-dimensional extension of Theorem 9.62. The proof will follow the same pattern as its one dimensional cousin with a few complications due to cumbersome nature of multi-dimensional cumulative distribution functions.

**Theorem 9.63 (Multi-Dimensional Riesz Markov Theorem).** *Let $X = [0,1]^d$ and $\lambda \in C(X)^*$ be a positive linear functional. Then there exists a unique Borel measure, $\mu$, on $\mathcal{B}_X$ such that*

$$\lambda(f) = \mu(f) := \int_X f d\mu \text{ for all } f \in C(X). \tag{9.37}$$

**Proof.** For $\mathbf{b} \in [0,1]^d$ and $\varepsilon > 0$ let $\varphi_{\mathbf{b},\varepsilon}(x) := \prod_{i=1}^d \varphi_{b_i,\varepsilon}(x_i)$ so that $\varphi_{\mathbf{b},\varepsilon}|_X \in C(X)$ and $\varphi_{\mathbf{b},\varepsilon} \downarrow 1_{[0,\mathbf{b}]}$ as $\varepsilon \downarrow 0$ which again shows $1_{[0,\mathbf{b}]} \in C(X)_\downarrow$. We will carry the proof in a number of steps.

1. We start by defining $F : [0,1]^d \to [0,\infty)$ by $F(\mathbf{b}) := \lambda\left(1_{[0,\mathbf{b}]}\right)$ for all $\mathbf{b} \in [0,1]^d$ which exists by Remark 9.60. Let $\mu_F$ be the unique finitely additive measure on $\mathcal{A}\left([0,1]^d\right)$ such that
$$\mu_F(\{a,b]) = \sum_{\gamma \subset \{1,\dots,d\}} \varepsilon_\gamma (-1)^{|\gamma|} F(a_\gamma \times b_{\gamma^c}). \tag{9.38}$$
Notice that if $\mathbf{a},\mathbf{b} \in [0,1]^d$ with $\mathbf{a} \le \mathbf{b}$, then $1_{[0,\mathbf{a}]} \le 1_{[0,\mathbf{b}]}$ and so $F(\mathbf{a}) = \lambda\left(1_{[0,\mathbf{a}]}\right) \le \lambda\left(1_{[0,\mathbf{b}]}\right) = F(\mathbf{b})$.
In what follows we extend $F$ to $[0,\infty)^d$ by setting $F(b) := F(b_1 \wedge 1, \dots, b_d \wedge 1)$.
2. Next we will show $F$ is right continuous and $\mu_F \ge 0$ and hence there exists a unique extension, $\mu$, of $\mu_F$ to $\mathcal{B}_X$. We now prove these claims.
   a) **Right continuity of** $F$. We must show if $\mathbf{b}_n \in [0,1]^d$ with $\mathbf{a} \le \mathbf{b}_n$ and $\lim_{n\to\infty} \mathbf{b}_n = \mathbf{a}$, then $\lim_{n\to\infty} F(\mathbf{b}_n) = F(\mathbf{a})$. If we let $\mathbf{c}_m := \mathbf{a} + \frac{1}{m}\mathbf{1}$ for $m \in \mathbb{N}$, we will have $[0,\mathbf{b}_n] \subset [0,\mathbf{c}_m]$ for all sufficiently large $n$ and hence
   $$\limsup_{n\to\infty} F(\mathbf{b}_n) = \limsup_{n\to\infty} \lambda\left(1_{[0,\mathbf{b}_n]}\right) \le \lambda\left(1_{[0,\mathbf{c}_m]}\right) = F(\mathbf{c}_m).$$
   Therefore
   $$F(\mathbf{a}) \le \liminf_{n\to\infty} F(\mathbf{b}_n) \le \limsup_{n\to\infty} F(\mathbf{b}_n) \le \lim_{m\to\infty} F(\mathbf{c}_m)$$
   and so to finish the proof of right continuity it suffices to show $\lim_{n\to\infty} F(\mathbf{c}_n) \le F(\mathbf{a})$.
   If we let $\delta_n > 0$ be as in Eq. (9.36), then
   $$1_{[0,\mathbf{c}_n]} \le \varphi_{\mathbf{c}_n,\delta_n} \downarrow 1_{[0,\mathbf{a}]}$$
   and therefore
   $$F(\mathbf{c}_n) \le \lambda(\varphi_{\mathbf{c}_n,\delta_n}) \downarrow \lambda\left(1_{[0,\mathbf{a}]}\right) = F(\mathbf{a})$$
   which shows $\lim_{n\to\infty} F(\mathbf{c}_n) \le F(\mathbf{a})$.
   b) $\mu_F$ **is positive.** Let $a,b \in [0,1]^n$. From Eq. (5.43) we have
   $$\sum_{|\gamma| \text{ even}} w_\gamma 1_{[0,a_\gamma \times b_{\gamma^c}]} \ge \sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0,a_\gamma \times b_{\gamma^c}]}$$

and therefore using Theorem 9.59,

$$\sum_{|\gamma| \text{ even}} w_\gamma F\left(a_\gamma \times b_{\gamma^c}\right) = \lambda\left(\sum_{|\gamma| \text{ even}} w_\gamma \left(1_{[0,a_\gamma \times b_{\gamma^c}]}\right)\right)$$
$$\geq \lambda\left(\sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0,a_\gamma \times b_{\gamma^c}]}\right) = \sum_{|\gamma| \text{ odd}} w_\gamma F\left(a_\gamma \times b_{\gamma^c}\right).$$

The inequality combined with Eq. (9.38) shows $\mu_F(\{a,b])\geq 0$.

3. Next we show that if $a,b\in[0,1]^n$ with $a<b$ then $\lambda\left(1_{[a,b]}\right)\leq\mu_F([a,b])$. To prove this let $\tilde{a}_i=0$ and and $\{a_i,b_i]=[0,b_i]$ if $a_i=0$ and $\tilde{a}_i\in(0,a_i)$ and $\{a_i,b_i]=(a_i,b_i]$ if $a_i>0$. If we further let $\xi_\gamma:=\prod_{i\in\gamma}1_{a_i\neq 0}=\prod_{i\in\gamma}1_{\tilde{a}_i\neq 0}$, then we find,

$$\begin{aligned}1_{[a,b]}(x) &= \prod_{i=1}^d 1_{[a_i,b_i]}(x_i) = \prod_{i=1}^d \left[1_{[0,b_i]}(x_i) - 1_{[0,a_i)}(x_i)\right] \\ &= \prod_{i=1}^d \left[1_{[0,b_i]}(x_i) - 1_{a_i\neq 0}\cdot 1_{[0,a_i)}(x_i)\right] \\ &\leq \prod_{i=1}^d \left[1_{[0,b_i]}(x_i) - 1_{a_i\neq 0}\cdot 1_{[0,\tilde{a}_i]}(x_i)\right] \\ &= \sum_{\gamma\subset\{1,2,\dots,d\}} \xi_\gamma (-1)^{|\gamma|}\mathbf{1}_{\tilde{a}_\gamma\times b_{\gamma^c}}(x).\end{aligned}$$

This inequality may be rewritten as

$$1_{[a,b]} + \sum_{|\gamma| \text{ odd}} \xi_\gamma \mathbf{1}_{\tilde{a}_\gamma\times b_{\gamma^c}} \leq \sum_{|\gamma| \text{ even}} \xi_\gamma \mathbf{1}_{\tilde{a}_\gamma\times b_{\gamma^c}}.$$

Applying $\lambda$ to the last inequality and solving the result for $\lambda\left(1_{[a,b]}\right)$ shows

$$\begin{aligned}\lambda\left(1_{[a,b]}\right) &\leq \sum_{\gamma\subset\{1,2,\dots,d\}} \xi_\gamma (-1)^{|\gamma|}\lambda\left(\mathbf{1}_{\tilde{a}_\gamma\times b_{\gamma^c}}\right) \\ &= \sum_{\gamma\subset\{1,2,\dots,d\}} \xi_\gamma (-1)^{|\gamma|} F\left(\tilde{a}_\gamma\times b_{\gamma^c}\right) = \mu_F(\{\tilde{a},b])\end{aligned}$$

Letting $\tilde{a}\uparrow a$ in this inequality then implies $\lambda\left(1_{[a,b]}\right)\leq\mu_F([a,b])$.

4. (**Verification of** Eq. (9.37).) Let $\pi=\{0=s_0<s_1<\dots<s_n=1\}$ be partition of $[0,1]$, for $s=s_k\in\pi\setminus\{1\}$ let $s_+:=s_{k+1}$,and let $\Pi$ be the partition of $[0,1]^n$ consisting of rectangles $Q$ of the form $\{\mathbf{a},\mathbf{a}_+]$ where $\mathbf{a}=(a_1,\dots,a_d)\in[\pi\setminus\{1\}]^n$, $\mathbf{a}_+=\left((a_1)_+,\dots,(a_d)_+\right)$, and $\{\mathbf{a},\mathbf{a}_+]=\prod_{i=1}^d\{a_i,(a_i)_+]$ with $\{a_i,(a_i)_+]=(a_i,(a_i)_+]$ if $a_i>0$ and $\{a_i,(a_i)_+]=\left[a_i,(a_i)_+\right]$ if $a_i=0$. Let us first assume that $f\in C(X)$ is a non-negative function and for $Q\in\Pi$ let $c_Q:=\max_{x\in\bar{Q}}f(x)$ and define $f^\pi$ and $\hat{f}^\pi$ by

$$f^\pi := \sum_{Q\in\Pi} c_Q 1_Q \text{ and } \hat{f}^\pi := \sum_{Q\in\Pi} c_Q 1_{\bar{Q}}.$$

Since $0\leq f\leq f^\pi\leq\hat{f}^\pi\in C(X)_\downarrow$, we conclude that

$$0\leq\lambda(f)\leq\lambda\left(\hat{f}^\pi\right)=\sum_{Q\in\Pi}c_Q\lambda\left(1_{\bar{Q}}\right)\leq\sum_{Q\in\Pi}c_Q\mu_F\left(\bar{Q}\right) \tag{9.39}$$

wherein the last inequality we have used step 6.
Since

$$\sum_{u\in[0,1]}\mu\left(\left\{x\in[0,1]^d: x_i=u\right\}\right)\leq\mu\left([0,1]^d\right)<\infty$$

we conclude that

$$U:=\left\{u\in[0,1]:\max_i\mu\left(\left\{x\in[0,1]^d: x_i=u\right\}\right)>0\right\}$$

is at most countable. Let us now always suppose that we have chosen $\pi$ so that $\pi\setminus\{0,1\}\subset U^c$ in which case $\mu_F\left(\bar{Q}\right)=\mu_F(Q)$ for all $Q\in\Pi$. Thus under previous restriction on $\pi$ we the inequality in Eq. (9.39) becomes

$$0\leq\lambda(f)\leq\sum_{Q\in\Pi}c_Q\mu_F(Q)=\mu(f^\pi).$$

Since $f^\pi\to f$ boundedly as $|\pi|\to 0$ with $\pi\setminus\{0,1\}\subset U^c$ as above, we may pass to the limit in the previous inequality using DCT to find $0\leq\lambda(f)\leq\mu(f)$ for all $f\geq 0$.
For arbitrary $f\in C(X)$ choose $M>0$ so that $M+f\geq 0$ and therefore,

$$M\lambda(1)+\lambda(f)=\lambda(M+f)\leq\mu(M+f)=M\mu(1)+\mu(f).$$

As $\mu(1)=F(\mathbf{1})=\lambda(\mathbf{1})$, we conclude that $\lambda(f)\leq\mu(f)$ for all $f\in C(X)$. Replacing $f$ by $-f$ then shows $\mu(f)\leq\lambda(f)$ for all $f\in C(X)$ and hence it follows that $\lambda(f)=\mu(f)$.

■

**Theorem 9.64.** *Suppose that* $\lambda\in C_c\left((0,1)^d\right)^*$ *is a positive linear functional. Then there exists a* $K$ *– finite measure* $\mu$ *on* $\left((0,1)^d,\mathcal{B}_{(0,1)^d}\right)$ *such that* $\lambda(f)=\mu(f)$ *for all* $f\in C_c\left((0,1)^d\right)$.

**Proof.** Given $\varphi \in C_c\left((0,1)^d,[0,1]\right)$, let $\lambda_\varphi(f) := \lambda(\varphi\cdot f)$ for all $f \in C_c\left([0,1]^d\right)$. As $\lambda_\varphi$ is a positive linear functional on $C\left([0,1]^d\right)$ we may apply Theorem 9.63 to find a unique measure $\mu_\varphi$ on $\left([0,1]^d,\mathcal{B}_{[0,1]^d}\right)$ such that $\lambda_\varphi(f) = \mu_\varphi(f)$ for all $f \in C\left([0,1]^d\right)$. We now complete the proof in a number of steps.

1. If $\psi \in C_c\left((0,1)^d,[0,1]\right)$ with $\varphi \le \psi$ then $\mu_\varphi \le \mu_\psi$. To see this is the case first observe that if $F$ is a closed subset of $[0,1]^d$ and $f_n \in C\left([0,1]^d\right)$ are chosen so that $f_n \downarrow 1_F$, then
$$\mu_\varphi(F) = \lim_{n\to\infty}\lambda_\varphi(f_n) = \lim_{n\to\infty}\lambda(\varphi f_n) = \lambda(\varphi\cdot 1_F). \tag{9.40}$$
As $\varphi\cdot 1_F \le \psi\cdot 1_F$ it follows
$$\mu_\varphi(F) = \lambda(\varphi\cdot 1_F) \le \lambda(\psi\cdot 1_F) = \mu_\psi(F).$$
Then using the regularity properties of $\mu_\varphi$ and $\mu_\psi$ (see Exercise 6.12 and/or 6.10) if $A \in \mathcal{B}_{[0,1]^d}$ we have,
$$\begin{aligned}\mu_\varphi(A) &= \sup\{\mu_\varphi(F) : F \subset A \ \&\ F \text{ closed}\}\\ &\le \sup\{\mu_\psi(F) : F\subset A \ \&\ F \text{ closed}\} = \mu_\psi(A).\end{aligned} \tag{9.41}$$
2. If $A \in \mathcal{B}_{[0,1]^d}$ with $A \subset \{\varphi = \psi\}$, then $\mu_\varphi(A) = \mu_\psi(A)$. Indeed if $F$ is a closed subset of $A$, then $\varphi\cdot 1_F = \psi\cdot 1_F$ and hence from Eq. (9.40) it follows that $\mu_\varphi(F) = \mu_\psi(F)$ and in this case the inequality in Eq. (9.41) becomes an equality.
3. Let $Q_n := \left[\frac{1}{n}, 1-\frac{1}{n}\right]^d$ for $n \ge 3$ and let $\varphi_n \in C_c\left((0,1)^d,[0,1]\right)$ be chosen so that $1_{Q_n} \le \varphi_n$ and $\operatorname{supp}(\varphi_n) \subset Q_{n+1}$ and hence $\varphi_n \uparrow 1_{(0,1)^d}$. We then define the measure $\mu$ on $\mathcal{B}_{(0,1)^d}$ by
$$\mu(A) = \uparrow\lim_{n\to\infty}\mu_{\varphi_n}(A) \ \ \forall\ A \in \mathcal{B}_{(0,1)^d}.$$
4. An easy consequence of item 2. is that if $f$ is a bounded $\mathcal{B}_{(0,1)^d}$ – measurable function with compact support in $(0,1)^d$, then $\mu(f) = \mu_{\varphi_n}(f)$ for all $n$ such that $\operatorname{supp}(f) \subset Q_n$. In particular if $f \in C_c\left((0,1)^d\right)$ with $\operatorname{supp}(f)\subset Q_n$, then
$$\mu(f) = \mu_{\varphi_n}(f) = \lambda_{\varphi_n}(f) = \lambda(\varphi_n f) = \lambda(f)$$
and the proof is complete.

■

**Corollary 9.65.** *Suppose that $\lambda \in C_c\left(\mathbb{R}^d\right)^*$ is a positive linear functional. Then there exists a $K$ – finite measure $\mu$ on $\left(\mathbb{R}^d,\mathcal{B}_{\mathbb{R}^d}\right)$ such that $\lambda(f) = \mu(f)$ for all $f \in C_c\left(\mathbb{R}^d\right)$.*

**Proof.** The map, $\psi : (0,1) \to \mathbb{R}$ be defined by $\psi(t) = \cot^{-1}(t/\pi)$ is a homeomorphism and hence so is $\psi : (0,1)^d \to \mathbb{R}^d$ where
$$\psi(t_1,\dots,t_d) := (\psi(t_1),\dots,\psi(t_d)).$$
Using this homeomorphic identification of $(0,1)^d$ with $\mathbb{R}^d$ we may easily translate the statement in Theorem 9.64 to the statement in the corollary. ■

### 9.5.1 Bone yards to the proof of Theorems 9.62 9.63

**Proof.**

■

## 9.6 Some Common Continuous Distributions

*Example 9.66 (Uniform Distribution).* Suppose that $X$ has the uniform distribution in $[0,b]$ for some $b \in (0,\infty)$, i.e. $X_*P = \frac{1}{b}\cdot m$ on $[0,b]$. More explicitly,

$$\mathbb{E}\left[f(X)\right] = \frac{1}{b}\int_0^b f(x)\,dx \text{ for all bounded measurable } f.$$

The moment generating function for $X$ is;

$$\begin{aligned} M_X(t) &= \frac{1}{b}\int_0^b e^{tx}dx = \frac{1}{bt}\left(e^{tb}-1\right) \\ &= \sum_{n=1}^\infty \frac{1}{n!}(bt)^{n-1} = \sum_{n=0}^\infty \frac{b^n}{(n+1)!}t^n. \end{aligned}$$

On the other hand (see Proposition 9.33),

$$M_X(t) = \sum_{n=0}^\infty \frac{t^n}{n!}\mathbb{E}X^n.$$

Thus it follows that

$$\mathbb{E}X^n = \frac{b^n}{n+1}.$$

Of course this may be calculated directly just as easily,

$$\mathbb{E}X^n = \frac{1}{b}\int_0^b x^n dx = \frac{1}{b(n+1)}x^{n+1}|_0^b = \frac{b^n}{n+1}.$$

**Definition 9.67.** *A random variable $T \ge 0$ is said to be* ***exponential with parameter*** *$\lambda \in [0,\infty)$ provided, $P(T>t) = e^{-\lambda t}$ for all $t \ge 0$. We will write $T \overset{d}{=} E(\lambda)$ for short.*

If $\lambda > 0$, we have

$$P(T>t) = e^{-\lambda t} = \int_t^\infty \lambda e^{-\lambda\tau}d\tau$$

from which it follows that $P(T \in (t,t+dt)) = \lambda 1_{t\ge 0}e^{-\lambda t}dt$. Applying Corollary 9.30 repeatedly implies,

$$\mathbb{E}T = \int_0^\infty \tau\lambda e^{-\lambda\tau}d\tau = \lambda\left(-\frac{d}{d\lambda}\right)\int_0^\infty e^{-\lambda\tau}d\tau = \lambda\left(-\frac{d}{d\lambda}\right)\lambda^{-1} = \lambda^{-1}$$

and more generally that

$$\mathbb{E}T^k = \int_0^\infty \tau^k e^{-\lambda\tau}\lambda d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k\int_0^\infty e^{-\lambda\tau}d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k\lambda^{-1} = k!\lambda^{-k}. \tag{9.42}$$

In particular we see that

$$\mathrm{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \tag{9.43}$$

Alternatively we may compute the moment generating function for $T$,

$$\begin{aligned} M_T(a) &:= \mathbb{E}\left[e^{aT}\right] = \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau}d\tau \\ &= \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau}d\tau = \frac{\lambda}{\lambda-a} = \frac{1}{1-a\lambda^{-1}} \end{aligned} \tag{9.44}$$

which is valid for $a < \lambda$. On the other hand (see Proposition 9.33), we know that

$$\mathbb{E}\left[e^{aT}\right] = \sum_{n=0}^\infty \frac{a^n}{n!}\mathbb{E}\left[T^n\right] \text{ for } |a| < \lambda. \tag{9.45}$$

Comparing this with Eq. (9.44) again shows that Eq. (9.42) is valid.

Here is yet another way to understand and generalize Eq. (9.44). We simply make the change of variables, $u = \lambda\tau$ in the integral in Eq. (9.42) to learn,

$$\mathbb{E}T^k = \lambda^{-k}\int_0^\infty u^k e^{-u}d\tau = \lambda^{-k}\Gamma(k+1).$$

This last equation is valid for all $k \in (-1,\infty)$ – in particular $k$ need not be an integer.

**Theorem 9.68 (Memoryless property).** *A random variable, $T \in (0,\infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$P(T>s+t|T>s) = P(T>t) \text{ for all } s,t \ge 0,$$

*where as usual, $P(A|B) := P(A\cap B)/P(B)$ when $p(B) > 0$. (Note that $T \overset{d}{=} E(0)$ means that $P(T>t) = e^{0t} = 1$ for all $t > 0$ and therefore that $T = \infty$ a.s.)*

**Proof.** (The following proof is taken from [32].) Suppose first that $T \overset{d}{=} E(\lambda)$ for some $\lambda > 0$. Then

$$P(T > s+t|T > s) = \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t).$$

For the converse, let $g(t) := P(T > t)$, then by assumption,

$$\frac{g(t+s)}{g(s)} = P(T > s+t|T > s) = P(T > t) = g(t)$$

whenever $g(s) \neq 0$ and $g(t)$ is a decreasing function. Therefore if $g(s) = 0$ for some $s > 0$ then $g(t) = 0$ for all $t > s$. Thus it follows that

$$g(t+s) = g(t)\,g(s) \text{ for all } s,t \geq 0.$$

Since $T > 0$, we know that $g(1/n) = P(T > 1/n) > 0$ for some $n$ and therefore, $g(1) = g(1/n)^n > 0$ and we may write $g(1) = e^{-\lambda}$ for some $0 \leq \lambda < \infty$.

Observe for $p,q \in \mathbb{N}$, $g(p/q) = g(1/q)^p$ and taking $p = q$ then shows, $e^{-\lambda} = g(1) = g(1/q)^q$. Therefore, $g(p/q) = e^{-\lambda p/q}$ so that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{Q}_+ := \mathbb{Q} \cap \mathbb{R}_+$. Given $r,s \in \mathbb{Q}_+$ and $t \in \mathbb{R}$ such that $r \leq t \leq s$ we have, since $g$ is decreasing, that

$$e^{-\lambda r} = g(r) \geq g(t) \geq g(s) = e^{-\lambda s}.$$

Hence letting $s \uparrow t$ and $r \downarrow t$ in the above equations shows that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$ and therefore $T \overset{d}{=} E(\lambda)$. ■

**Exercise 9.13 (Gamma Distributions).** Let $X$ be a positive random variable. For $k,\theta > 0$, we say that $X \overset{d}{=}$Gamma$(k,\theta)$ if

$$(X_*P)(dx) = f(x;k,\theta)\,dx \text{ for } x > 0,$$

where

$$f(x;k,\theta) := x^{k-1}\frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \text{ for } x > 0, \text{ and } k,\theta > 0.$$

Find the moment generating function (see Definition 9.32), $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ for $t < \theta^{-1}$. Differentiate your result in $t$ to show

$$\mathbb{E}[X^m] = k(k+1)\dots(k+m-1)\,\theta^m \text{ for all } m \in \mathbb{N}_0.$$

In particular, $\mathbb{E}[X] = k\theta$ and $\operatorname{Var}(X) = k\theta^2$. (Notice that when $k = 1$ and $\theta = \lambda^{-1}$, $X \overset{d}{=} E(\lambda)$.)

### 9.6.1 Normal (Gaussian) Random Variables

**Definition 9.69 (Normal / Gaussian Random Variables).** *A random variable, $Y$, **is normal with mean** $\mu$ **standard deviation** $\sigma^2$ iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{9.46}$$

*We will abbreviate this by writing $Y \overset{d}{=} N(\mu,\sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ we will simply write $N$ for $N(0,1)$ and if $Y \overset{d}{=} N$, we will say $Y$ is a **standard normal** random variable.*

Observe that Eq. (9.46) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$. Also observe that $Y \overset{d}{=} N(\mu,\sigma^2)$ is equivalent to $Y \overset{d}{=} \sigma N + \mu$. Indeed, by making the change of variable, $y = \sigma x + \mu$, we find

$$\begin{aligned}\mathbb{E}[f(\sigma N + \mu)] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(\sigma x + \mu)\,e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}\frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} f(y)\,e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy.\end{aligned}$$

Lastly the constant, $\left(2\pi\sigma^2\right)^{-1/2}$ is chosen so that

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}y^2}dy = 1,$$

see Example 9.50 and Lemma 11.29.

**Exercise 9.14.** Suppose that $X \overset{d}{=} N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ is a $C^1$ − function such that $Xf(X)$, $f'(X)$ and $f(X)$ are all integrable random variables. Show

$$\begin{aligned}\mathbb{E}[Xf(X)] &= -\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)\,\frac{d}{dx}e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f'(x)\,e^{-\frac{1}{2}x^2}dx = \mathbb{E}[f'(X)].\end{aligned}$$

*Example 9.70.* Suppose that $X \overset{d}{=} N(0,1)$ and define $\alpha_k := \mathbb{E}\left[X^{2k}\right]$ for all $k \in \mathbb{N}_0$. By Exercise 9.14,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1} \cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

Hence it follows that

$$\alpha_1 = \alpha_0 = 1,\ \alpha_2 = 3\alpha_1 = 3,\ \alpha_3 = 5 \cdot 3$$

and by a simple induction argument,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!, \tag{9.47}$$

where $(-1)!! := 0$. Actually we can use the $\Gamma$ – function to say more. Namely for any $\beta > -1$,

$$\mathbb{E}\,|X|^{\beta} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^{\beta}\, e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^{\beta} e^{-\frac{1}{2}x^2} dx.$$

Now make the change of variables, $y = x^2/2$ (i.e. $x = \sqrt{2y}$ and $dx = \frac{1}{\sqrt{2}} y^{-1/2} dy$) to learn,

$$\begin{aligned} \mathbb{E}\,|X|^{\beta} &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} (2y)^{\beta/2}\, e^{-y} y^{-1/2} dy \\ &= \frac{1}{\sqrt{\pi}} 2^{\beta/2} \int_0^{\infty} y^{(\beta+1)/2} e^{-y} y^{-1} dy = \frac{1}{\sqrt{\pi}} 2^{\beta/2} \Gamma\left(\frac{\beta+1}{2}\right). \end{aligned} \tag{9.48}$$

**Exercise 9.15.** Suppose that $X \overset{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = \exp\left(-\lambda^2/2\right). \tag{9.49}$$

**Hint:** Use Corollary 9.30 to show, $f'(\lambda) = i\mathbb{E}\left[X e^{i\lambda X}\right]$ and then use Exercise 9.14 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

**Exercise 9.16.** Suppose that $X \overset{d}{=} N(0,1)$ and $t \in \mathbb{R}$. Show $\mathbb{E}\left[e^{tX}\right] = \exp\left(t^2/2\right)$. (You could follow the hint in Exercise 9.15 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

**Exercise 9.17.** Use Exercise 9.16 and Proposition 9.33 to give another proof that $\mathbb{E}X^{2k} = (2k-1)!!$ when $X \overset{d}{=} N(0,1)$.

**Exercise 9.18.** Let $X \overset{d}{=} N(0,1)$ and $\alpha \in \mathbb{R}$, find $\rho : \mathbb{R}_+ \to \mathbb{R}_+ := (0,\infty)$ such that

$$\mathbb{E}\left[f(|X|^{\alpha})\right] = \int_{\mathbb{R}_+} f(x)\,\rho(x)\,dx$$

for all continuous functions, $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support in $\mathbb{R}_+$.

**Lemma 9.71 (Gaussian tail estimates).** *Suppose that $X$ is a standard normal random variable, i.e.*

$$P(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}},$$

*then for all $x \geq 0$,*

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right) \leq \frac{1}{2} e^{-x^2/2}. \tag{9.50}$$

*Moreover (see [35, Lemma 2.5]),*

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \tag{9.51}$$

*which combined with Eq. (9.50) proves Mill's ratio (see [18]);*

$$\lim_{x\to\infty} \frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}} = 1. \tag{9.52}$$

**Proof.** See Figure 9.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right),$$

the red is the plot of $\frac{1}{2} e^{-x^2/2}$, and the blue is the plot of

$$\max\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right).$$

The formal proof of these estimates for the reader who is not convinced by Figure 9.1 is given below.

We begin by observing that

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy \leq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \frac{y}{x} e^{-y^2/2} dy \\ &\leq -\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-y^2/2}\Big|_x^{-\infty} = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \end{aligned} \tag{9.53}$$

If we only want to prove Mill's ratio (9.52), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,

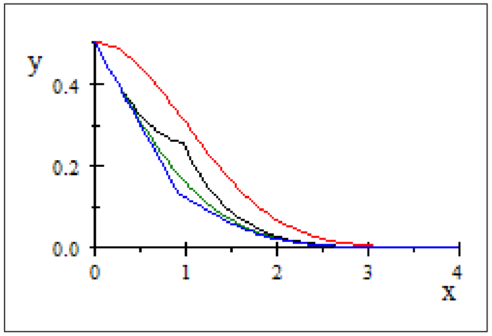



**Fig. 9.1.** Plots of $P(X \geq x)$ and its estimates.

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \\ &\geq \frac{1}{\sqrt{2\pi}} \int_x^{\alpha x} \frac{y}{\alpha x} e^{-y^2/2} dy = -\frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-y^2/2}|_{y=x}^{y=\alpha x} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-x^2/2} \left[1 - e^{-\alpha^2 x^2/2}\right] \end{aligned}$$

from which it follows,

$$\liminf_{x\to\infty} \left[\sqrt{2\pi} x e^{x^2/2} \cdot P(X \geq x)\right] \geq 1/\alpha \uparrow 1 \text{ as } \alpha \downarrow 1.$$

The estimate in Eq. (9.53) shows $\limsup_{x\to\infty} \left[\sqrt{2\pi} x e^{x^2/2} \cdot P(X \geq x)\right] \leq 1$.

To get more precise estimates, we begin by observing,

$$\begin{aligned} P(X \geq x) &= \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy \qquad (9.54) \\ &\leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-x^2/2} dy \leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x. \end{aligned}$$

This equation along with Eq. (9.53) gives the first equality in Eq. (9.50). To prove the second equality observe that $\sqrt{2\pi} > 2$, so

$$\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2} \text{ if } x \geq 1.$$

For $x \leq 1$ we must show,

$$\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2}$$

or equivalently that $f(x) := e^{x^2/2} - \sqrt{\frac{2}{\pi}} x \leq 1$ for $0 \leq x \leq 1$. Since $f$ is convex $\left(f''(x) = \left(x^2+1\right) e^{x^2/2} > 0\right)$, $f(0) = 1$ and $f(1) \cong 0.85 < 1$, it follows that $f \leq 1$ on $[0,1]$. This proves the second inequality in Eq. (9.50).

It follows from Eq. (9.54) that

$$\begin{aligned} P(X \geq x) &= \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy \\ &\geq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x 1 dy = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} x \text{ for all } x \geq 0. \end{aligned}$$

So to finish the proof of Eq. (9.51) we must show,

$$\begin{aligned} f(x) &:= \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} - \left(1 + x^2\right) P(X \geq x) \\ &= \frac{1}{\sqrt{2\pi}} \left[x e^{-x^2/2} - \left(1 + x^2\right) \int_x^\infty e^{-y^2/2} dy\right] \leq 0 \text{ for all } 0 \leq x < \infty. \end{aligned}$$

This follows by observing that $f(0) = -1/2 < 0$, $\lim_{x\uparrow\infty} f(x) = 0$ and

$$\begin{aligned} f'(x) &= \frac{1}{\sqrt{2\pi}} \left[e^{-x^2/2}\left(1 - x^2\right) - 2xP(X \geq x) + \left(1 + x^2\right) e^{-x^2/2}\right] \\ &= 2\left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2} - xP(X \geq y)\right) \geq 0, \end{aligned}$$

where the last inequality is a consequence Eq. (9.50). ■

## 9.7 Stirling's Formula

On occasion one is faced with estimating an integral of the form, $\int_J e^{-G(t)} dt$, where $J = (a,b) \subset \mathbb{R}$ and $G(t)$ is a $C^1$ – function with a unique (for simplicity) global minimum at some point $t_0 \in J$. The idea is that the majority contribution of the integral will often come from some neighborhood, $(t_0 - \alpha, t_0 + \alpha)$, of $t_0$. Moreover, it may happen that $G(t)$ can be well approximated on this neighborhood by its Taylor expansion to order 2;

$$G(t) \cong G(t_0) + \frac{1}{2} \ddot{G}(t_0)(t - t_0)^2.$$

Notice that the linear term is zero since $t_0$ is a minimum and therefore $\dot{G}(t_0) = 0$. We will further assume that $\ddot{G}(t_0) \neq 0$ and hence $\ddot{G}(t_0) > 0$. Under these hypothesis we will have,

$$\int_J e^{-G(t)}dt \cong e^{-G(t_0)} \int_{|t-t_0|<\alpha} \exp\left(-\frac{1}{2}\ddot{G}(t_0)(t-t_0)^2\right) dt.$$

Making the change of variables, $s = \sqrt{\ddot{G}(t_0)}(t-t_0)$, in the above integral then gives,

$$\begin{aligned}\int_J e^{-G(t)}dt &\cong \frac{1}{\sqrt{\ddot{G}(t_0)}} e^{-G(t_0)} \int_{|s|<\sqrt{\ddot{G}(t_0)}\cdot\alpha} e^{-\frac{1}{2}s^2} ds \\ &= \frac{1}{\sqrt{\ddot{G}(t_0)}} e^{-G(t_0)} \left[\sqrt{2\pi} - \int_{\sqrt{\ddot{G}(t_0)}\cdot\alpha}^{\infty} e^{-\frac{1}{2}s^2} ds\right] \\ &= \frac{1}{\sqrt{\ddot{G}(t_0)}} e^{-G(t_0)} \left[\sqrt{2\pi} - O\left(\frac{1}{\sqrt{\ddot{G}(t_0)}\cdot\alpha} e^{-\frac{1}{2}\ddot{G}(t_0)\cdot\alpha^2}\right)\right].\end{aligned}$$

If $\alpha$ is sufficiently large, for example if $\sqrt{\ddot{G}(t_0)}\cdot\alpha = 3$, then the error term is about 0.0037 and we should be able to conclude that

$$\int_J e^{-G(t)}dt \cong \sqrt{\frac{2\pi}{\ddot{G}(t_0)}} e^{-G(t_0)}. \tag{9.55}$$

The proof of the next theorem (Stirling's formula for the Gamma function) will illustrate these ideas and what one has to do to carry them out rigorously.

**Theorem 9.72 (Stirling's formula).** *The Gamma function (see Definition 9.49), satisfies Stirling's formula,*

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{\sqrt{2\pi}e^{-x}x^{x+1/2}} = 1. \tag{9.56}$$

*In particular, if $n \in \mathbb{N}$, we have*

$$n! = \Gamma(n+1) \sim \sqrt{2\pi}e^{-n}n^{n+1/2}$$

*where we write $a_n \sim b_n$ to mean, $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. (See Example 9.77 below for a slightly cruder but more elementary estimate of $n!$)*

**Proof.** (The following proof is an elaboration of the proof found on page 236-237 in Krantz's Real Analysis and Foundations.) We begin with the formula for $\Gamma(x+1)$;

$$\Gamma(x+1) = \int_0^\infty e^{-t}t^x dt = \int_0^\infty e^{-G_x(t)}dt, \tag{9.57}$$

where

$$G_x(t) := t - x\ln t.$$

Then $\dot{G}_x(t) = 1 - x/t$, $\ddot{G}_x(t) = x/t^2$, $G_x$ has a global minimum (since $\ddot{G}_x > 0$) at $t_0 = x$ where

$$G_x(x) = x - x\ln x \text{ and } \ddot{G}_x(x) = 1/x.$$

So if Eq. (9.55) is valid in this case we should expect,

$$\Gamma(x+1) \cong \sqrt{2\pi x}e^{-(x-x\ln x)} = \sqrt{2\pi}e^{-x}x^{x+1/2}$$

which would give Stirling's formula. The rest of the proof will be spent on rigorously justifying the approximations involved.

Let us begin by making the change of variables $s = \sqrt{\ddot{G}(t_0)}(t-t_0) = \frac{1}{\sqrt{x}}(t-x)$ as suggested above. Then

$$\begin{aligned}G_x(t) - G_x(x) &= (t-x) - x\ln(t/x) = \sqrt{x}s - x\ln\left(\frac{x+\sqrt{x}s}{x}\right) \\ &= x\left[\frac{s}{\sqrt{x}} - \ln\left(1+\frac{s}{\sqrt{x}}\right)\right] = s^2 q\left(\frac{s}{\sqrt{x}}\right)\end{aligned}$$

where

$$q(u) := \frac{1}{u^2}[u - \ln(1+u)] \text{ for } u > -1 \text{ with } q(0) := \frac{1}{2}.$$

Setting $q(0) = 1/2$ makes $q$ a continuous and in fact smooth function on $(-1,\infty)$, see Figure 9.2. Using the power series expansion for $\ln(1+u)$ we find,

$$q(u) = \frac{1}{2} + \sum_{k=3}^\infty \frac{(-u)^{k-2}}{k} \text{ for } |u| < 1. \tag{9.58}$$

Making the change of variables, $t = x + \sqrt{x}s$ in the second integral in Eq. (9.57) yields,

$$\Gamma(x+1) = e^{-(x-x\ln x)}\sqrt{x}\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds = x^{x+1/2}e^{-x}\cdot I(x),$$

where

$$I(x) = \int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds = \int_{-\infty}^\infty 1_{s\geq-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds. \tag{9.59}$$

From Eq. (9.58) it follows that $\lim_{u\to0} q(u) = 1/2$ and therefore,

$$\int_{-\infty}^\infty \lim_{x\to\infty}\left[1_{s\geq-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\right]ds = \int_{-\infty}^\infty e^{-\frac{1}{2}s^2}ds = \sqrt{2\pi}. \tag{9.60}$$

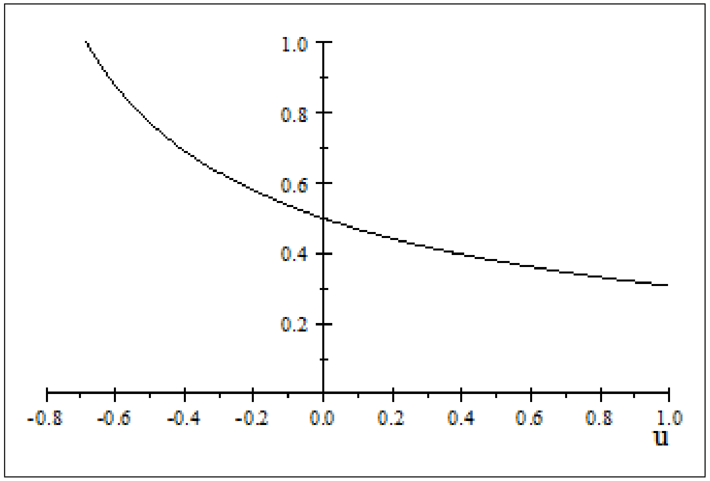



**Fig. 9.2.** Plot of $q(u)$.

So if there exists a dominating function, $F \in L^1(\mathbb{R}, m)$, such that

$$1_{s \geq -\sqrt{x}} \cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} \leq F(s) \text{ for all } s \in \mathbb{R} \text{ and } x \geq 1,$$

we can apply the DCT to learn that $\lim_{x\to\infty} I(x) = \sqrt{2\pi}$ which will complete the proof of Stirling's formula.

We now construct the desired function $F$. From Eq. (9.58) it follows that $q(u) \geq 1/2$ for $-1 < u \leq 0$. Since $u - \ln(1+u) > 0$ for $u \neq 0$ ($u - \ln(1+u)$ is convex and has a minimum of 0 at $u = 0$) we may conclude that $q(u) > 0$ for all $u > -1$ therefore by compactness (on $[0, M]$), $\min_{-1<u\leq M} q(u) = \varepsilon(M) > 0$ for all $M \in (0, \infty)$, see Remark 9.73 for more explicit estimates. Lastly, since $\frac{1}{u}\ln(1+u) \to 0$ as $u \to \infty$, there exists $M < \infty$ ($M = 3$ would due) such that $\frac{1}{u}\ln(1+u) \leq \frac{1}{2}$ for $u \geq M$ and hence,

$$q(u) = \frac{1}{u}\left[1 - \frac{1}{u}\ln(1+u)\right] \geq \frac{1}{2u} \text{ for } u \geq M.$$

So there exists $\varepsilon > 0$ and $M < \infty$ such that (for all $x \geq 1$),

$$\begin{aligned} 1_{s\geq-\sqrt{x}} e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-\sqrt{x}s/2} \\ &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-s/2} \\ &\leq e^{-\varepsilon s^2} + e^{-|s|/2} =: F(s) \in L^1(\mathbb{R}, ds). \end{aligned}$$

■

We will sometimes use the following variant of Eq. (9.56);

$$\lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x} = 1 \tag{9.61}$$

To prove this let $x$ go to $x - 1$ in Eq. (9.56) in order to find,

$$1 = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{2\pi} e^{-x} \cdot e \cdot (x-1)^{x-1/2}} = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x \cdot \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^x}$$

which gives Eq. (9.61) since

$$\lim_{x\to\infty} \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^x = 1.$$

*Remark 9.73 (Estimating $q(u)$ by Taylor's Theorem).* Another way to estimate $q(u)$ is to use Taylor's theorem with integral remainder. In general if $h$ is $C^2$ – function on $[0, 1]$, then by the fundamental theorem of calculus and integration by parts,

$$\begin{aligned} h(1) - h(0) &= \int_0^1 \dot{h}(t)\, dt = -\int_0^1 \dot{h}(t)\, d(1-t) \\ &= -\dot{h}(t)(1-t)|_0^1 + \int_0^1 \ddot{h}(t)(1-t)\, dt \\ &= \dot{h}(0) + \frac{1}{2}\int_0^1 \ddot{h}(t)\, d\nu(t) \end{aligned} \tag{9.62}$$

where $d\nu(t) := 2(1-t)\,dt$ which is a probability measure on $[0, 1]$. Applying this to $h(t) = F(a + t(b-a))$ for a $C^2$ – function on an interval of points between $a$ and $b$ in $\mathbb{R}$ then implies,

$$F(b) - F(a) = (b-a)\dot{F}(a) + \frac{1}{2}(b-a)^2 \int_0^1 \ddot{F}(a + t(b-a))\, d\nu(t). \tag{9.63}$$

(Similar formulas hold to any order.) Applying this result with $F(x) = x - \ln(1+x)$, $a = 0$, and $b = u \in (-1, \infty)$ gives,

$$u - \ln(1+u) = \frac{1}{2}u^2 \int_0^1 \frac{1}{(1+tu)^2}\, d\nu(t),$$

i.e.

$$q(u) = \frac{1}{2}\int_0^1 \frac{1}{(1+tu)^2}\, d\nu(t).$$

From this expression for $q(u)$ it now easily follows that

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+0)^2} d\nu(t) = \frac{1}{2} \text{ if } -1 < u \leq 0$$

and

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+u)^2} d\nu(t) = \frac{1}{2(1+u)^2}.$$

So an explicit formula for $\varepsilon(M)$ is $\varepsilon(M) = (1+M)^{-2}/2$.

### 9.7.1 Two applications of Stirling's formula

In this subsection suppose $x \in (0,1)$ and $S_n \overset{d}{=} \text{Binomial}(n,x)$ for all $n \in \mathbb{N}$, i.e.

$$P_x(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k} \text{ for } 0 \leq k \leq n. \tag{9.64}$$

Recall that $\mathbb{E}S_n = nx$ and $\text{Var}(S_n) = n\sigma^2$ where $\sigma^2 := x(1-x)$. The weak law of large numbers states (Exercise 5.19) that

$$P\left(\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\sigma^2$$

and therefore, $\frac{S_n}{n}$ is concentrating near its mean value, $x$, for $n$ large, i.e. $S_n \cong nx$ for $n$ large. The next central limit theorem describes the fluctuations of $S_n$ about $nx$.

**Theorem 9.74 (De Moivre-Laplace Central Limit Theorem).** *For all* $-\infty < a < b < \infty$,

$$\begin{aligned}\lim_{n\to\infty} P\left(a \leq \frac{S_n - nx}{\sigma\sqrt{n}} \leq b\right) &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2} dy \\ &= P(a \leq N \leq b)\end{aligned}$$

*where* $N \overset{d}{=} N(0,1)$. *Informally,* $\frac{S_n - nx}{\sigma\sqrt{n}} \overset{d}{\cong} N$ *or equivalently,* $S_n \overset{d}{\cong} nx + \sigma\sqrt{n}\cdot N$ *which if valid in a neighborhood of* $nx$ *whose length is order* $\sqrt{n}$.

**Proof.** (We are not going to cover all the technical details in this proof as we will give much more general versions of this theorem later.) Starting with the definition of the Binomial distribution we have,

$$\begin{aligned}p_n &:= P\left(a \leq \frac{S_n - nx}{\sigma\sqrt{n}} \leq b\right) = P\left(S_n \in nx + \sigma\sqrt{n}[a,b]\right) \\ &= \sum_{k \in nx + \sigma\sqrt{n}[a,b]} P(S_n = k) \\ &= \sum_{k \in nx + \sigma\sqrt{n}[a,b]} \binom{n}{k} x^k (1-x)^{n-k}.\end{aligned}$$

Letting $k = nx + \sigma\sqrt{n}y_k$, i.e. $y_k = (k - nx)/\sigma\sqrt{n}$ we see that $\Delta y_k = y_{k+1} - y_k = 1/(\sigma\sqrt{n})$. Therefore we may write $p_n$ as

$$p_n = \sum_{y_k \in [a,b]} \sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \Delta y_k. \tag{9.65}$$

So to finish the proof we need to show, for $k = O(\sqrt{n})$ $(y_k = O(1))$, that

$$\sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \text{ as } n \to \infty \tag{9.66}$$

in which case the sum in Eq. (9.65) may be well approximated by the "Riemann sum;"

$$p_n \sim \sum_{y_k \in [a,b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \Delta y_k \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2} dy \text{ as } n \to \infty.$$

By Stirling's formula,

$$\begin{aligned}\sigma\sqrt{n}\binom{n}{k} &= \sigma\sqrt{n}\frac{1}{k!}\frac{n!}{(n-k)!} \sim \frac{\sigma\sqrt{n}}{\sqrt{2\pi}}\frac{n^{n+1/2}}{k^{k+1/2}(n-k)^{n-k+1/2}} \\ &= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(\frac{k}{n}\right)^{k+1/2}\left(1-\frac{k}{n}\right)^{n-k+1/2}} \\ &= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k+1/2}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k+1/2}} \\ &\sim \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\sqrt{x(1-x)}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}.\end{aligned}$$

In order to shorten the notation, let $z_k := \frac{\sigma}{\sqrt{n}}y_k = O\left(n^{-1/2}\right)$ so that $k = nx + nz_k = n(x + z_k)$. In this notation we have shown,

$$\sqrt{2\pi}\sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k} \sim \frac{x^k(1-x)^{n-k}}{(x+z_k)^k(1-x-z_k)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^k\left(1-\frac{1}{1-x}z_k\right)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^{n(x+z_k)}\left(1-\frac{1}{1-x}z_k\right)^{n(1-x-z_k)}} =: q(n,k). \tag{9.67}$$

Taking logarithms and using Taylor's theorem we learn

$$n(x+z_k)\ln\left(1+\frac{1}{x}z_k\right)$$
$$= n(x+z_k)\left(\frac{1}{x}z_k - \frac{1}{2x^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= nz_k + \frac{n}{2x}z_k^2 + O\left(n^{-3/2}\right) \text{ and}$$
$$n(1-x-z_k)\ln\left(1-\frac{1}{1-x}z_k\right)$$
$$= n(1-x-z_k)\left(-\frac{1}{1-x}z_k - \frac{1}{2(1-x)^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= -nz_k + \frac{n}{2(1-x)}z_k^2 + O\left(n^{-3/2}\right).$$

and then adding these expressions shows,

$$-\ln q(n,k) = \frac{n}{2}z_k^2\left(\frac{1}{x}+\frac{1}{1-x}\right) + O\left(n^{-3/2}\right)$$
$$= \frac{n}{2\sigma^2}z_k^2 + O\left(n^{-3/2}\right) = \frac{1}{2}y_k^2 + O\left(n^{-3/2}\right).$$

Combining this with Eq. (9.67) shows,

$$\sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_k^2 + O\left(n^{-3/2}\right)\right)$$

which gives the desired estimate in Eq. (9.66). ∎

The previous central limit theorem has shown that

$$\frac{S_n}{n} \overset{d}{\cong} x + \frac{\sigma}{\sqrt{n}}N$$

which implies the major fluctuations of $S_n/n$ occur within intervals about $x$ of length $O\left(\frac{1}{\sqrt{n}}\right)$. The next result aims to understand the rare events where $S_n/n$ makes a "large" deviation from its mean value, $x$ – in this case a large deviation is something of size $O(1)$ as $n\to\infty$.

**Theorem 9.75 (Binomial Large Deviation Bounds).** *Let us continue to use the notation in Theorem 9.74. Then for all $y\in(0,x)$,*

$$\lim_{n\to\infty}\frac{1}{n}\ln P_x\left(\frac{S_n}{n}\le y\right) = y\ln\frac{x}{y} + (1-y)\ln\frac{1-x}{1-y}.$$

*Roughly speaking,*

$$P_x\left(\frac{S_n}{n}\le y\right) \approx e^{-nI_x(y)}$$

*where $I_x(y)$ is the "rate function,"*

$$I_x(y) := y\ln\frac{y}{x} + (1-y)\ln\frac{1-y}{1-x},$$

*see Figure 9.3 for the graph of $I_{1/2}$.*

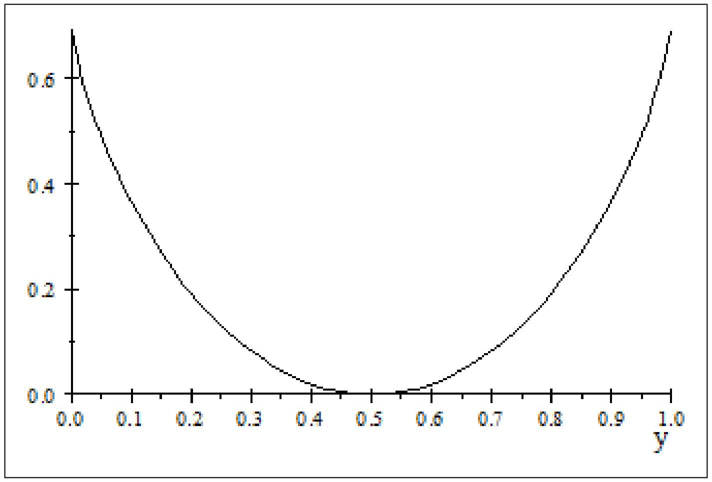



**Fig. 9.3.** A plot of the rate function, $I_{1/2}$.

**Proof.** By definition of the binomial distribution,

$$P_x\left(\frac{S_n}{n}\le y\right) = P_x(S_n\le ny) = \sum_{k\le ny}\binom{n}{k}x^k(1-x)^{n-k}.$$

If $a_k \geq 0$, then we have the following crude estimates on $\sum_{k=0}^{m-1} a_k$,

$$\max_{k<m} a_k \leq \sum_{k=0}^{m-1} a_k \leq m \cdot \max_{k<m} a_k. \tag{9.68}$$

In order to apply this with $a_k = \binom{n}{k} x^k (1-x)^{n-k}$ and $m = [ny]$, we need to find the maximum of the $a_k$ for $0 \leq k \leq ny$. This is easy to do since $a_k$ is increasing for $0 \leq k \leq ny$ as we now show. Consider,

$$\begin{aligned}\frac{a_{k+1}}{a_k} &= \frac{\binom{n}{k+1} x^{k+1} (1-x)^{n-k-1}}{\binom{n}{k} x^k (1-x)^{n-k}} \\ &= \frac{k! \, (n-k)! \cdot x}{(k+1)! \cdot (n-k-1)! \cdot (1-x)} \\ &= \frac{(n-k) \cdot x}{(k+1) \cdot (1-x)}.\end{aligned}$$

Therefore, where the latter expression is greater than or equal to 1 iff

$$\begin{aligned}\frac{a_{k+1}}{a_k} \geq 1 &\iff (n-k) \cdot x \geq (k+1) \cdot (1-x) \\ &\iff nx \geq k+1-x \iff k < (n-1)\, x - 1.\end{aligned}$$

Thus for $k < (n-1)\, x - 1$ we may conclude that $\binom{n}{k} x^k (1-x)^{n-k}$ is increasing in $k$.

Thus the crude bound in Eq. (9.68) implies,

$$\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]} \leq P_x\left(\frac{S_n}{n} \leq y\right) \leq [ny] \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}$$

or equivalently,

$$\begin{aligned}&\frac{1}{n} \ln\left[\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right] \\ &\qquad \leq \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \leq y\right) \\ &\qquad\qquad \leq \frac{1}{n} \ln\left[(ny) \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right].\end{aligned}$$

By Stirling's formula, for $k$ such that $k$ and $n-k$ is large we have,

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi}} \frac{n^{n+1/2}}{k^{k+1/2} \cdot (n-k)^{n-k+1/2}} = \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k}{n}\right)^{k+1/2} \cdot \left(1-\frac{k}{n}\right)^{n-k+1/2}}$$

and therefore,

$$\frac{1}{n} \ln \binom{n}{k} \sim -\frac{k}{n} \ln\left(\frac{k}{n}\right) - \left(1-\frac{k}{n}\right) \ln\left(1-\frac{k}{n}\right).$$

So taking $k = [ny]$, we learn that

$$\lim_{n\to\infty} \frac{1}{n} \ln \binom{n}{[ny]} = -y \ln y - (1-y) \ln (1-y)$$

and therefore,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \leq y\right) &= -y \ln y - (1-y) \ln (1-y) + y \ln x + (1-y) \ln (1-x) \\ &= y \ln \frac{x}{y} + (1-y) \ln\left(\frac{1-x}{1-y}\right).\end{aligned}$$

■

As a consistency check it is worth noting, by Jensen's inequality described below, that

$$-I_x(y) = y \ln \frac{x}{y} + (1-y) \ln\left(\frac{1-x}{1-y}\right) \leq \ln\left(y \frac{x}{y} + (1-y) \frac{1-x}{1-y}\right) = \ln(1) = 0.$$

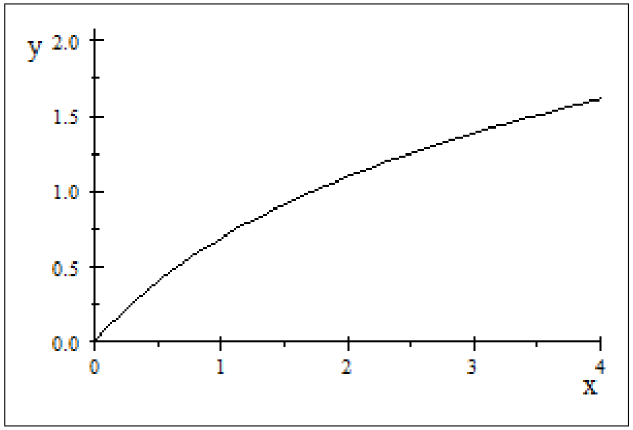
This must be the case since

$$-I_x(y) = \lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \leq y\right) \leq \lim_{n\to\infty} \frac{1}{n} \ln 1 = 0.$$

### 9.7.2 A primitive Stirling type approximation

**Theorem 9.76.** *Suppose that $f : (0,\infty) \to \mathbb{R}$ is an increasing concave down function (like $f(x) = \ln x$) and let $s_n := \sum_{k=1}^n f(k)$, then*

$$\begin{aligned}s_n - \frac{1}{2}(f(n) + f(1)) &\leq \int_1^n f(x)\, dx \\ &\leq s_n - \frac{1}{2}[f(n+1) + 2f(1)] + \frac{1}{2} f(2) \\ &\leq s_n - \frac{1}{2}[f(n) + 2f(1)] + \frac{1}{2} f(2).\end{aligned}$$




**Proof.** On the interval, $[k-1,k]$, we have that $f(x)$ is larger than the straight line segment joining $(k-1, f(k-1))$ and $(k, f(k))$ and thus

$$\frac{1}{2}(f(k)+f(k-1)) \leq \int_{k-1}^{k} f(x)\,dx.$$

Summing this equation on $k=2,\ldots,n$ shows,

$$s_n - \frac{1}{2}(f(n)+f(1)) = \sum_{k=2}^{n} \frac{1}{2}(f(k)+f(k-1)) \leq \sum_{k=2}^{n} \int_{k-1}^{k} f(x)\,dx = \int_1^n f(x)\,dx.$$

For the upper bound on the integral we observe that $f(x) \leq f(k) - f'(k)(x-k)$ for all $x$ and therefore,

$$\int_{k-1}^{k} f(x)\,dx \leq \int_{k-1}^{k} [f(k) - f'(k)(x-k)]\,dx = f(k) - \frac{1}{2}f'(k).$$

Summing this equation on $k=2,\ldots,n$ then implies,

$$\int_1^n f(x)\,dx \leq \sum_{k=2}^{n} f(k) - \frac{1}{2}\sum_{k=2}^{n} f'(k).$$

Since $f''(x) \leq 0$, $f'(x)$ is decreasing and therefore $f'(x) \leq f'(k-1)$ for $x \in [k-1,k]$ and integrating this equation over $[k-1,k]$ gives

$$f(k) - f(k-1) \leq f'(k-1).$$

Summing the result on $k=3,\ldots,n+1$ then shows,

$$f(n+1) - f(2) \leq \sum_{k=2}^{n} f'(k)$$

and thus ti follows that

$$\begin{aligned}\int_1^n f(x)\,dx &\leq \sum_{k=2}^{n} f(k) - \frac{1}{2}(f(n+1) - f(2)) \\ &= s_n - \frac{1}{2}[f(n+1) + 2f(1)] + \frac{1}{2}f(2) \\ &\leq s_n - \frac{1}{2}[f(n) + 2f(1)] + \frac{1}{2}f(2)\end{aligned}$$

■

*Example 9.77 (Approximating $n!$).* Let us take $f(n) = \ln n$ and recall that

$$\int_1^n \ln x dx = n\ln n - n + 1.$$

Thus we may conlcud that

$$s_n - \frac{1}{2}\ln n \leq n\ln n - n + 1 \leq s_n - \frac{1}{2}\ln n + \frac{1}{2}\ln 2.$$

Thus it follows that

$$\left(n+\frac{1}{2}\right)\ln n - n + 1 - \ln\sqrt{2} \leq s_n \leq \left(n+\frac{1}{2}\right)\ln n - n + 1.$$

Exponentiating this identity then implies,

$$\frac{e}{\sqrt{2}} \cdot e^{-n} n^{n+1/2} \leq n! \leq e \cdot e^{-n} n^{n+1/2}$$

which compares well with Strirling's formula (Theorem 9.72) which states,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Observe that

$$\frac{e}{\sqrt{2}} \cong 1.9221 \leq \sqrt{2\pi} \cong 2.506 \leq e \cong 2.7183.$$

## 9.8 Comparison of the Lebesgue and the Riemann Integral*

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a,b] \to \mathbb{R}$ be a bounded function. A partition of $[a,b]$ is a finite subset $\pi \subset [a,b]$ containing $\{a,b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{9.69}$$

of $[a,b]$ let

$$\text{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1,\dots,n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \quad m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1},t_j]}, \quad g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1},t_j]} \text{ and}$$

$$S_\pi f = \sum M_j(t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j(t_j - t_{j-1}).$$

Notice that

$$S_\pi f = \int_a^b G_\pi dm \text{ and } s_\pi f = \int_a^b g_\pi dm.$$

The upper and lower Riemann integrals are defined respectively by

$$\overline{\int_a^b} f(x)dx = \inf_\pi S_\pi f \text{ and } \underline{\int_b^a} f(x)dx = \sup_\pi\ s_\pi f.$$

**Definition 9.78.** *The function $f$ is* ***Riemann integrable*** *iff $\overline{\int_a^b} f = \underline{\int_a^b} f \in \mathbb{R}$ and which case the Riemann integral $\int_a^b f$ is defined to be the common value:*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

The proof of the following Lemma is left to the reader as Exercise 9.29.

**Lemma 9.79.** *If $\pi'$ and $\pi$ are two partitions of $[a,b]$ and $\pi \subset \pi'$ then*

$$G_\pi \ge G_{\pi'} \ge f \ge g_{\pi'} \ge g_\pi \text{ and}$$
$$S_\pi f \ge S_{\pi'} f \ge s_{\pi'} f \ge s_\pi f.$$

*There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that* $\text{mesh}(\pi_k) \downarrow 0$ *and*

$$S_{\pi_k} f \downarrow \overline{\int_a^b} f \text{ and } s_{\pi_k} f \uparrow \underline{\int_a^b} f \text{ as } k \to \infty.$$

If we let

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{9.70}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} gdm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{9.71}$$

and

$$\int_{[a,b]} Gdm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{9.72}$$

**Notation 9.80** *For $x \in [a,b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0}\ \sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \text{ and}$$
$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0}\ \inf\ \{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\}.$$

**Lemma 9.81.** *The functions $H, h : [a,b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \le f(x) \le H(x)$ for all $x \in [a,b]$ and $h(x) = H(x)$ iff $f$ is continuous at $x$.*
2. *If $\{\pi_k\}_{k=1}^\infty$ is any increasing sequence of partitions such that* $\text{mesh}(\pi_k) \downarrow 0$ *and $G$ and $g$ are defined as in Eq. (9.70), then*
$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall\ x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{9.73}$$
*(Note $\pi$ is a countable set.)*
3. *$H$ and $h$ are Borel measurable.*

**Proof.** Let $G_k := G_{\pi_k} \downarrow G$ and $g_k := g_{\pi_k} \uparrow g$.

1. It is clear that $h(x) \le f(x) \le H(x)$ for all $x$ and $H(x) = h(x)$ iff $\lim_{y\to x} f(y)$ exists and is equal to $f(x)$. That is $H(x) = h(x)$ iff $f$ is continuous at $x$.
2. For $x \notin \pi$,
$$G_k(x) \ge H(x) \ge f(x) \ge h(x) \ge g_k(x)\ \forall\ k$$
and letting $k \to \infty$ in this equation implies
$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x)\ \forall\ x \notin \pi. \tag{9.74}$$
Moreover, given $\varepsilon > 0$ and $x \notin \pi$,
$$\sup\{f(y) : |y-x| \le \varepsilon,\ y \in [a,b]\} \ge G_k(x)$$
for all $k$ large enough, since eventually $G_k(x)$ is the supremum of $f(y)$ over some interval contained in $[x-\varepsilon, x+\varepsilon]$. Again letting $k \to \infty$ implies $\sup_{|y-x|\le\varepsilon} f(y) \ge G(x)$ and therefore, that
$$H(x) = \limsup_{y\to x} f(y) \ge G(x)$$

for all $x \notin \pi$. Combining this equation with Eq. (9.74) then implies $H(x) = G(x)$ if $x \notin \pi$. A similar argument shows that $h(x) = g(x)$ if $x \notin \pi$ and hence Eq. (9.73) is proved.
3. The functions $G$ and $g$ are limits of measurable functions and hence measurable. Since $H = G$ and $h = g$ except possibly on the countable set $\pi$, both $H$ and $h$ are also Borel measurable. (You justify this statement.)

■

**Theorem 9.82.** *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{9.75}$$

*and the following statements are equivalent:*

1. *$H(x) = h(x)$ for $m$ -a.e. $x$,*
2. *the set*
$$E := \{x \in [a,b] : f \text{ is discontinuous at } x\}$$
*is an $\bar{m}$ – null set.*
3. *$f$ is Riemann integrable.*

*If $f$ is Riemann integrable then $f$ is Lebesgue measurable[3], i.e. $f$ is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue $\sigma$ – algebra and $\mathcal{B}$ is the Borel $\sigma$ – algebra on $[a,b]$. Moreover if we let $\bar{m}$ denote the completion of $m$, then*

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{9.76}$$

**Proof.** Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 9.79 and let $G$ and $g$ be defined as in Lemma 9.81. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (9.75) is a consequence of Eqs. (9.71) and (9.72). From Eq. (9.75), $f$ is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \le f \le H$ this happens iff $h(x) = H(x)$ for $m$ - a.e. $x$. Since $E = \{x : H(x) \ne h(x)\}$, this last condition is equivalent to $E$ being a $m$ – null set. In light of these results and Eq. (9.73), the remaining assertions including Eq. (9.76) are now consequences of Lemma 9.85. ■

**Notation 9.83** *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

[3] $f$ need not be Borel measurable.



## 9.9 Measurability on Complete Measure Spaces*

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

**Proposition 9.84.** *Suppose that $(X, \mathcal{B}, \mu)$ is a complete measure space[4] and $f : X \to \mathbb{R}$ is measurable.*

1. *If $g : X \to \mathbb{R}$ is a function such that $f(x) = g(x)$ for $\mu$ – a.e. $x$, then $g$ is measurable.*
2. *If $f_n : X \to \mathbb{R}$ are measurable and $f : X \to \mathbb{R}$ is a function such that $\lim_{n\to\infty} f_n = f$, $\mu$ - a.e., then $f$ is measurable as well.*

**Proof.** 1. Let $E = \{x : f(x) \ne g(x)\}$ which is assumed to be in $\mathcal{B}$ and $\mu(E) = 0$. Then $g = 1_{E^c} f + 1_E g$ since $f = g$ on $E^c$. Now $1_{E^c} f$ is measurable so $g$ will be measurable if we show $1_E g$ is measurable. For this consider,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{9.77}$$

Since $(1_E g)^{-1}(B) \subset E$ if $0 \notin B$ and $\mu(E) = 0$, it follow by completeness of $\mathcal{B}$ that $(1_E g)^{-1}(B) \in \mathcal{B}$ if $0 \notin B$. Therefore Eq. (9.77) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \ne f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, $g$ is measurable. Because $f = g$ on $E^c$ and $\mu(E) = 0$, $f = g$ a.e. so by part 1. $f$ is also measurable. ■

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0,1,2\}$, $\mathcal{B} = \{\{0\}, \{1,2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet $g$ is not measurable.

**Lemma 9.85.** *Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\bar{\mathcal{M}}$ is the completion of $\mathcal{M}$ relative to $\mu$ and $\bar{\mu}$ is the extension of $\mu$ to $\bar{\mathcal{M}}$. Then a function $f : X \to \mathbb{R}$ is $(\bar{\mathcal{M}}, \mathcal{B} = \mathcal{B}_\mathbb{R})$ – measurable iff there exists a function $g : X \to \mathbb{R}$ that is $(\mathcal{M}, \mathcal{B})$ – measurable such $E = \{x : f(x) \ne g(x)\} \in \bar{\mathcal{M}}$ and $\bar{\mu}(E) = 0$, i.e. $f(x) = g(x)$ for $\bar{\mu}$ – a.e. $x$. Moreover for such a pair $f$ and $g$, $f \in L^1(\bar{\mu})$ iff $g \in L^1(\mu)$ and in which case*

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

**Proof.** Suppose first that such a function $g$ exists so that $\bar{\mu}(E) = 0$. Since $g$ is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 9.84 that $f$ is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if $f$ is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may

[4] Recall this means that if $N \subset X$ is a set such that $N \subset A \in \mathcal{M}$ and $\mu(A) = 0$, then $N \in \mathcal{M}$ as well.

assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (9.78) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{9.78}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (9.78) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 8.39 it holds for all $\mathcal{M}$ – measurable functions $g : X \to [0, \infty]$. The rest of the assertions follow in the standard way by considering $(\operatorname{Re} g)_\pm$ and $(\operatorname{Im} g)_\pm$. ■

## 9.10 More Exercises

**Exercise 9.19.** Let $\mu$ be a measure on an algebra $\mathcal{A} \subset 2^X$, then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathcal{A}$.

**Exercise 9.20 (From problem 12 on p. 27 of Folland.).** Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $A, B \in \mathcal{M}$ let $\rho(A, B) = \mu(A \triangle B)$ where $A \triangle B = (A \setminus B) \cup (B \setminus A)$. It is clear that $\rho(A, B) = \rho(B, A)$. Show:

1. $\rho$ satisfies the triangle inequality:
$$\rho(A, C) \leq \rho(A, B) + \rho(B, C) \text{ for all } A, B, C \in \mathcal{M}.$$
2. Define $A \sim B$ iff $\mu(A \triangle B) = 0$ and notice that $\rho(A, B) = 0$ iff $A \sim B$. Show "$\sim$" is an equivalence relation.
3. Let $\mathcal{M}/\sim$ denote $\mathcal{M}$ modulo the equivalence relation, $\sim$, and let $[A] := \{B \in \mathcal{M} : B \sim A\}$. Show that $\bar{\rho}([A], [B]) := \rho(A, B)$ is gives a well defined metric on $\mathcal{M}/\sim$.
4. Similarly show $\tilde{\mu}([A]) = \mu(A)$ is a well defined function on $\mathcal{M}/\sim$ and show $\tilde{\mu} : (\mathcal{M}/\sim) \to \mathbb{R}_+$ is $\bar{\rho}$ – continuous.

**Exercise 9.21.** Suppose that $\mu_n : \mathcal{M} \to [0, \infty]$ are measures on $\mathcal{M}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in $n$ for all $A \in \mathcal{M}$. Prove that $\mu : \mathcal{M} \to [0, \infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

**Exercise 9.22.** Now suppose that $\Lambda$ is some index set and for each $\lambda \in \Lambda$, $\mu_\lambda : \mathcal{M} \to [0, \infty]$ is a measure on $\mathcal{M}$. Define $\mu : \mathcal{M} \to [0, \infty]$ by $\mu(A) = \sum_{\lambda \in \Lambda} \mu_\lambda(A)$ for each $A \in \mathcal{M}$. Show that $\mu$ is also a measure.

**Exercise 9.23.** Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show

$$\mu(\{A_n \text{ a.a.}\}) \leq \liminf_{n\to\infty} \mu(A_n)$$

and if $\mu(\cup_{m \geq n} A_m) < \infty$ for some $n$, then

$$\mu(\{A_n \text{ i.o.}\}) \geq \limsup_{n\to\infty} \mu(A_n).$$

**Exercise 9.24 (Folland 2.13 on p. 52.).** Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of non-negative measurable functions such that $f_n \to f$ pointwise and

$$\lim_{n\to\infty} \int f_n = \int f < \infty.$$

Then

$$\int_E f = \lim_{n\to\infty} \int_E f_n$$

for all measurable sets $E \in \mathcal{M}$. The conclusion need not hold if $\lim_{n\to\infty} \int f_n = \int f$. **Hint:** "Fatou times two."

**Exercise 9.25.** Give examples of measurable functions $\{f_n\}$ on $\mathbb{R}$ such that $f_n$ decreases to 0 uniformly yet $\int f_n dm = \infty$ for all $n$. Also give an example of a sequence of measurable functions $\{g_n\}$ on $[0, 1]$ such that $g_n \to 0$ while $\int g_n dm = 1$ for all $n$.

**Exercise 9.26.** Suppose $\{a_n\}_{n=-\infty}^\infty \subset \mathbb{C}$ is a summable sequence (i.e. $\sum_{n=-\infty}^\infty |a_n| < \infty$), then $f(\theta) := \sum_{n=-\infty}^\infty a_n e^{in\theta}$ is a continuous function for $\theta \in \mathbb{R}$ and

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$

**Exercise 9.27.** For any function $f \in L^1(m)$, show $x \in \mathbb{R} \to \int_{(-\infty, x]} f(t)\, dm(t)$ is continuous in $x$. Also find a finite measure, $\mu$, on $\mathcal{B}_\mathbb{R}$ such that $x \to \int_{(-\infty, x]} f(t)\, d\mu(t)$ is not continuous.

**Exercise 9.28.** Folland 2.31b and 2.31e on p. 60. (The answer in 2.13b is wrong by a factor of $-1$ and the sum is on $k = 1$ to $\infty$. In part (e), $s$ should be taken to be $a$. You may also freely use the Taylor series expansion

$$(1-z)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n n!} z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} z^n \text{ for } |z| < 1.$$

**Exercise 9.29.** Prove Lemma 9.79.

# Functional Forms of the $\pi - \lambda$ Theorem

In this chapter we will develop a very useful function analogue of the $\pi - \lambda$ theorem. The results in this section will be used often in the sequel.

## 10.1 Multiplicative System Theorems

**Notation 10.1** *Let $\Omega$ be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on $\Omega$. We say that $\mathbb{H}$ is* ***closed under bounded convergence*** *if; for every sequence, $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $|f_n(\omega)| \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f(\omega) := \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$, then $f \in \mathbb{H}$.*

*A subset, $\mathbb{M}$, of $\mathbb{H}$ is called a* ***multiplicative system*** *if $\mathbb{M}$ is closed under finite products.*

The following result may be found in Dellacherie [7, p. 14]. The style of proof given here may be found in Janson [23, Appendix A., p. 309].

**Theorem 10.2 (Dynkin's Multiplicative System Theorem).** *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from $\Omega$ to $\mathbb{R}$ which contains the constant functions and is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

**Proof.** In this proof, we may (and do) assume that $\mathbb{H}$ is the smallest subspace of bounded functions on $\Omega$ which contains the constant functions, contains $\mathbb{M}$, and is closed under bounded convergence. (As usual such a space exists by taking the intersection of all such spaces.) The remainder of the proof will be broken into four steps.

**Step 1.** ($\mathbb{H}$ is an algebra of functions.) For $f \in \mathbb{H}$, let $\mathbb{H}^f := \{g \in \mathbb{H} : gf \in \mathbb{H}\}$. The reader will now easily verify that $\mathbb{H}^f$ is a linear subspace of $\mathbb{H}$, $1 \in \mathbb{H}^f$, and $\mathbb{H}^f$ is closed under bounded convergence. Moreover if $f \in \mathbb{M}$, since $\mathbb{M}$ is a multiplicative system, $\mathbb{M} \subset \mathbb{H}^f$. Hence by the definition of $\mathbb{H}$, $\mathbb{H} = \mathbb{H}^f$, i.e. $fg \in \mathbb{H}$ for all $f \in \mathbb{M}$ and $g \in \mathbb{H}$. Having proved this it now follows for any $f \in \mathbb{H}$ that $\mathbb{M} \subset \mathbb{H}^f$ and therefore as before, $\mathbb{H}^f = \mathbb{H}$. Thus we may conclude that $fg \in \mathbb{H}$ whenever $f, g \in \mathbb{H}$, i.e. $\mathbb{H}$ is an algebra of functions.

**Step 2.** ($\mathcal{B} := \{A \subset \Omega : 1_A \in \mathbb{H}\}$ is a $\sigma$ – algebra.) Using the fact that $\mathbb{H}$ is an algebra containing constants, the reader will easily verify that $\mathcal{B}$ is closed under complementation, finite intersections, and contains $\Omega$, i.e. $\mathcal{B}$ is an algebra. Using the fact that $\mathbb{H}$ is closed under bounded convergence, it follows that $\mathcal{B}$ is closed under increasing unions and hence that $\mathcal{B}$ is $\sigma$ – algebra.

**Step 3.** ($\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.) Since $\mathbb{H}$ is a vector space and $\mathbb{H}$ contains $1_A$ for all $A \in \mathcal{B}$, $\mathbb{H}$ contains all $\mathcal{B}$ – measurable simple functions. Since every bounded $\mathcal{B}$ – measurable function may be written as a bounded limit of such simple functions (see Theorem 8.39), it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.

**Step 4.** ($\sigma(\mathbb{M}) \subset \mathcal{B}$.) Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 10.1 below) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f \in \mathbb{M}$ and $a \in \mathbb{R}$, let $F_n := \varphi_n(f - a)$ and $M := \sup_{\omega\in\Omega} |f(\omega) - a|$. By the Weierstrass approximation Theorem 5.70, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$. Since $p_l$ is a polynomial and $\mathbb{H}$ is an algebra, $p_l(f - a) \in \mathbb{H}$ for all $l$. Moreover, $p_l \circ (f - a) \to F_n$ uniformly as $l \to \infty$, from with it follows that $F_n \in \mathbb{H}$ for all $n$. Since, $F_n \uparrow 1_{\{f>a\}}$ it follows that $1_{\{f>a\}} \in \mathbb{H}$, i.e. $\{f > a\} \in \mathcal{B}$. As the sets $\{f > a\}$ with $a \in \mathbb{R}$ and $f \in \mathbb{M}$ generate $\sigma(\mathbb{M})$, it follows that $\sigma(\mathbb{M}) \subset \mathcal{B}$.

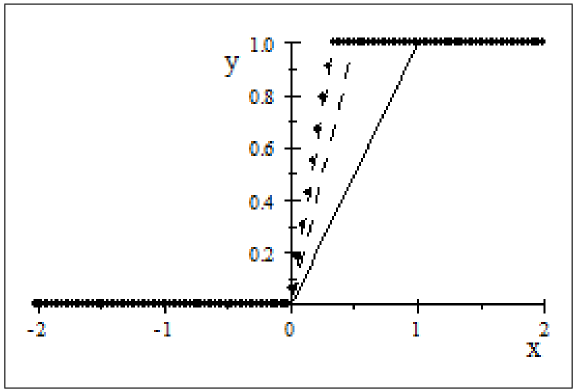



**Fig. 10.1.** Plots of $\varphi_1$, $\varphi_2$ and $\varphi_3$.

**Second proof.*** (This proof may safely be skipped.) This proof will make use of Dynkin's $\pi-\lambda$ Theorem 6.17. Let

$$\mathcal{L} := \{A \subset \Omega : 1_A \in \mathbb{H}\}.$$

We then have $\Omega \in \mathcal{L}$ since $1_\Omega = 1 \in \mathbb{H}$, if $A, B \in \mathcal{L}$ with $A \subset B$ then $B \setminus A \in \mathcal{L}$ since $1_{B\setminus A} = 1_B - 1_A \in \mathbb{H}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $A \in \mathcal{L}$ because $1_{A_n} \in \mathbb{H}$ and $1_{A_n} \uparrow 1_A \in \mathbb{H}$. Therefore $\mathcal{L}$ is $\lambda$ – system.

Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 10.1 above) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f_1, f_2, \dots, f_k \in \mathbb{M}$ and $a_1, \dots, a_k \in \mathbb{R}$, let

$$F_n := \prod_{i=1}^k \varphi_n(f_i - a_i)$$

and let

$$M := \sup_{i=1,\dots,k} \sup_\omega |f_i(\omega) - a_i|.$$

By the Weierstrass approximation Theorem 5.70, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$ .Since $p_l$ is a polynomial it is easily seen that $\prod_{i=1}^k p_l \circ (f_i - a_i) \in \mathbb{H}$. Moreover,

$$\prod_{i=1}^k p_l \circ (f_i - a_i) \to F_n \text{ uniformly as } l \to \infty,$$

from with it follows that $F_n \in \mathbb{H}$ for all $n$. Since,

$$F_n \uparrow \prod_{i=1}^k 1_{\{f_i > a_i\}} = 1_{\cap_{i=1}^k \{f_i > a_i\}}$$

it follows that $1_{\cap_{i=1}^k \{f_i > a_i\}} \in \mathbb{H}$ or equivalently that $\cap_{i=1}^k \{f_i > a_i\} \in \mathcal{L}$. Therefore $\mathcal{L}$ contains the $\pi$ – system, $\mathcal{P}$, consisting of finite intersections of sets of the form, $\{f > a\}$ with $f \in \mathbb{M}$ and $a \in \mathbb{R}$.

As a consequence of the above paragraphs and the $\pi-\lambda$ Theorem 6.17, $\mathcal{L}$ contains $\sigma(\mathcal{P}) = \sigma(\mathbb{M})$. In particular it follows that $1_A \in \mathbb{H}$ for all $A \in \sigma(\mathbb{M})$. Since any positive $\sigma(\mathbb{M})$ – measurable function may be written as a increasing limit of simple functions (see Theorem 8.39)), it follows that $\mathbb{H}$ contains all non-negative bounded $\sigma(\mathbb{M})$ – measurable functions. Finally, since any bounded $\sigma(\mathbb{M})$ – measurable functions may be written as the difference of two such non-negative simple functions, it follows that $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. ∎

**Corollary 10.3.** *Suppose $\mathbb{H}$ is a subspace of bounded real valued functions such that $1 \in \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. If $\mathcal{P} \subset 2^\Omega$ is a multiplicative class such that $1_A \in \mathbb{H}$ for all $A \in \mathcal{P}$, then $\mathbb{H}$ contains all bounded $\sigma(\mathcal{P})$ – measurable functions.*

**Proof.** Let $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$. Then $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system and the proof is completed with an application of Theorem 10.2. ∎

*Example 10.4.* Suppose $\mu$ and $\nu$ are two probability measure on $(\Omega, \mathcal{B})$ such that

$$\int_\Omega f d\mu = \int_\Omega f d\nu \tag{10.1}$$

for all $f$ in a multiplicative subset, $\mathbb{M}$, of bounded measurable functions on $\Omega$. Then $\mu = \nu$ on $\sigma(\mathbb{M})$. Indeed, apply Theorem 10.2 with $\mathbb{H}$ being the bounded measurable functions on $\Omega$ such that Eq. (10.1) holds. In particular if $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$ with $\mathcal{P}$ being a multiplicative class we learn that $\mu = \nu$ on $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

**Exercise 10.1.** Let $\Omega := \{1, 2, 3, 4\}$ and $\mathbb{M} := \{1_A, 1_B\}$ where $A := \{1, 2\}$ and $B := \{2, 3\}$.

a) Show $\sigma(\mathbb{M}) = 2^\Omega$.
b) Find two distinct probability measures, $\mu$ and $\nu$ on $2^\Omega$ such that $\mu(A) = \nu(A)$ and $\mu(B) = \nu(B)$, i.e. Eq. (10.1) holds for all $f \in \mathbb{M}$.

**Moral:** the assumption that $\mathbb{M}$ is multiplicative can not be dropped from Theorem 10.2.

Here is a complex version of Theorem 10.2.

**Theorem 10.5 (Complex Multiplicative System Theorem).** *Suppose $\mathbb{H}$ is a complex linear subspace of the bounded complex functions on $\Omega$, $1 \in \mathbb{H}$, $\mathbb{H}$ is closed under complex conjugation, and $\mathbb{H}$ is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system which is closed under conjugation, then $\mathbb{H}$ contains all bounded complex valued $\sigma(\mathbb{M})$-measurable functions.*

**Proof.** Let $\mathbb{M}_0 = \text{span}_\mathbb{C}(\mathbb{M} \cup \{1\})$ be the complex span of $\mathbb{M}$. As the reader should verify, $\mathbb{M}_0$ is an algebra, $\mathbb{M}_0 \subset \mathbb{H}$, $\mathbb{M}_0$ is closed under complex conjugation and $\sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$. Let

$$\mathbb{H}^\mathbb{R} := \{f \in \mathbb{H} : f \text{ is real valued}\} \text{ and}$$
$$\mathbb{M}_0^\mathbb{R} := \{f \in \mathbb{M}_0 : f \text{ is real valued}\}.$$

Then $\mathbb{H}^\mathbb{R}$ is a real linear space of bounded real valued functions 1 which is closed under bounded convergence and $\mathbb{M}_0^\mathbb{R} \subset \mathbb{H}^\mathbb{R}$. Moreover, $\mathbb{M}_0^\mathbb{R}$ is a multiplicative system (as the reader should check) and therefore by Theorem 10.2, $\mathbb{H}^\mathbb{R}$ contains all bounded $\sigma(\mathbb{M}_0^\mathbb{R})$ – measurable real valued functions. Since $\mathbb{H}$ and $\mathbb{M}_0$ are complex linear spaces closed under complex conjugation, for any $f \in \mathbb{H}$ or $f \in \mathbb{M}_0$, the functions $\text{Re} f = \frac{1}{2}(f + \bar{f})$ and $\text{Im} f = \frac{1}{2i}(f - \bar{f})$ are in $\mathbb{H}$ or $\mathbb{M}_0$ respectively. Therefore $\mathbb{M}_0 = \mathbb{M}_0^\mathbb{R} + i\mathbb{M}_0^\mathbb{R}$, $\sigma(\mathbb{M}_0^\mathbb{R}) = \sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$, and $\mathbb{H} = \mathbb{H}^\mathbb{R} + i\mathbb{H}^\mathbb{R}$. Hence if $f : \Omega \to \mathbb{C}$ is a bounded $\sigma(\mathbb{M})$ – measurable function, then $f = \text{Re} f + i \,\text{Im} f \in \mathbb{H}$ since $\text{Re} f$ and $\text{Im} f$ are in $\mathbb{H}^\mathbb{R}$. ∎

**Lemma 10.6.** *If* $-\infty < a < b < \infty$, *there exists* $f_n \in C_c(\mathbb{R}, [0,1])$ *such that* $\lim_{n\to\infty} f_n = 1_{(a,b]}$.

**Proof.** The reader should verify $\lim_{n\to\infty} f_n = 1_{(a,b]}$ where $f_n \in C_c(\mathbb{R}, [0,1])$ is defined (for $n$ sufficiently large) by

$$f_n(x) := \begin{cases} 0 & \text{on } (-\infty, a] \cup [b + \frac{1}{n}, \infty) \\ n(x-a) & \text{if } \quad a \le x \le a + \frac{1}{n} \\ 1 & \text{if } \quad a + \frac{1}{n} \le x \le b \\ 1 - n(b-x) & \text{if } \quad b \le x \le b + \frac{1}{n} \end{cases}.$$

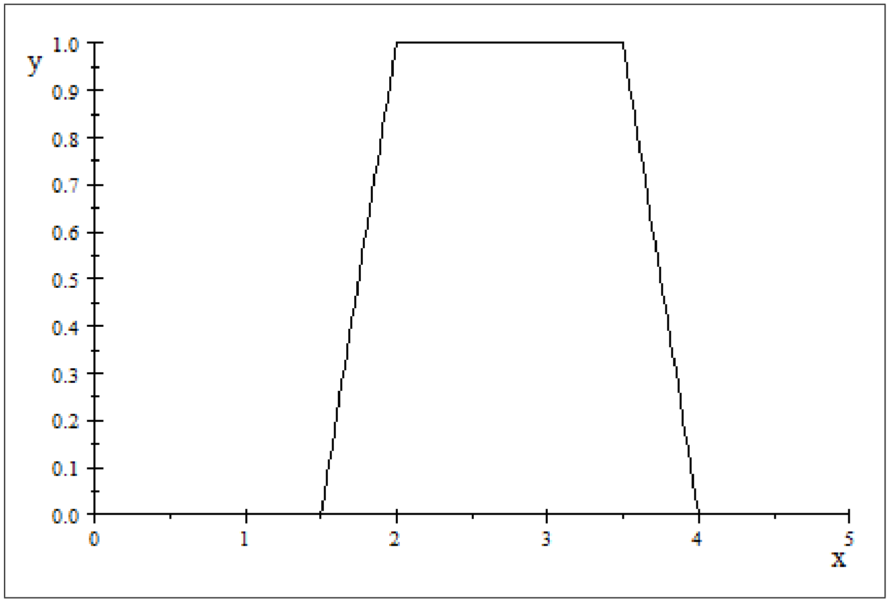



**Fig. 10.2.** Here is a plot of $f_2(x)$ when $a = 1.5$ and $b = 3.5$.

■

**Lemma 10.7.** *For each* $\lambda > 0$, *let* $e_\lambda(x) := e^{i\lambda x}$. *Then*

$$\mathcal{B}_{\mathbb{R}} = \sigma(e_\lambda : \lambda > 0) = \sigma\left(e_\lambda^{-1}(W) : \lambda > 0,\ W \in \mathcal{B}_{\mathbb{R}}\right).$$

**Proof.** Let $S^1 := \{z \in \mathbb{C} : |z| = 1\}$. For $-\pi < \alpha < \beta < \pi$ let

$$A(\alpha, \beta) := \left\{e^{i\theta} : \alpha < \theta < \beta\right\} = S^1 \cap \left\{re^{i\theta} : \alpha < \theta < \beta,\ r > 0\right\}$$

which is a measurable subset of $\mathbb{C}$ (why). Moreover we have $e_\lambda(x) \in A(\alpha, \beta)$ iff $\lambda x \in \sum_{n\in\mathbb{Z}} [(\alpha, \beta) + 2\pi n]$ and hence

$$e_\lambda^{-1}(A(\alpha, \beta)) = \sum_{n\in\mathbb{Z}} \left[\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) + 2\pi\frac{n}{\lambda}\right] \in \sigma(e_\lambda : \lambda > 0).$$

Hence if $-\infty < a < b < \infty$ and $\lambda > 0$ sufficiently small so that $-\pi < \lambda a < \lambda b < \pi$, we have

$$e_\lambda^{-1}(A(\lambda a, \lambda b)) = \sum_{n\in\mathbb{Z}} \left[(a, b) + 2\pi\frac{n}{\lambda}\right]$$

and hence

$$(a, b) = \cap_{\lambda > 0} e_\lambda^{-1}(A(\lambda a, \lambda b)) \in \sigma(e_\lambda : \lambda > 0).$$

This shows $\mathcal{B}_{\mathbb{R}} \subset \sigma(e_\lambda : \lambda > 0)$. As $e_\lambda$ is continuous and hence Borel measurable for all $\lambda > 0$ we automatically know that $\sigma(e_\lambda : \lambda > 0) \subset \mathcal{B}_{\mathbb{R}}$. ■

*Remark 10.8.* A slight modification of the above proof actually shows if $\{\lambda_n\} \subset (0, \infty)$ with $\lim_{n\to\infty} \lambda_n = 0$, then $\sigma(e_{\lambda_n} : n \in \mathbb{N}) = \mathcal{B}_{\mathbb{R}}$.

**Corollary 10.9.** *Each of the following* $\sigma$ *– algebras on* $\mathbb{R}^d$ *are equal to* $\mathcal{B}_{\mathbb{R}^d}$;

1. $\mathcal{M}_1 := \sigma\left(\cup_{i=1}^n \{x \to f(x_i) : f \in C_c(\mathbb{R})\}\right)$,
2. $\mathcal{M}_2 := \sigma(x \to f_1(x_1) \dots f_d(x_d) : f_i \in C_c(\mathbb{R}))$
3. $\mathcal{M}_3 = \sigma\left(C_c\left(\mathbb{R}^d\right)\right)$, *and*
4. $\mathcal{M}_4 := \sigma\left(\left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}\right)$.

**Proof.** As the functions defining each $\mathcal{M}_i$ are continuous and hence Borel measurable, it follows that $\mathcal{M}_i \subset \mathcal{B}_{\mathbb{R}^d}$ for each $i$. So to finish the proof it suffices to show $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_i$ for each $i$.

$\mathcal{M}_1$ case. Let $a, b \in \mathbb{R}$ with $-\infty < a < b < \infty$. By Lemma 10.6, there exists $f_n \in C_c(\mathbb{R})$ such that $\lim_{n\to\infty} f_n = 1_{(a,b]}$. Therefore it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_1$ – measurable for each $i$. Moreover if $-\infty < a_i < b_i < \infty$ for each $i$, then we may conclude that

$$x \to \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = 1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x)$$

is $\mathcal{M}_1$ – measurable as well and hence $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_1$. As such sets generate $\mathcal{B}_{\mathbb{R}^d}$ we may conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_1$.

and therefore $\mathcal{M}_1 = \mathcal{B}_{\mathbb{R}^d}$.

$\mathcal{M}_2$ case. As above, we may find $f_{i,n} \to 1_{(a_i,b_i]}$ as $n \to \infty$ for each $1 \le i \le d$ and therefore,

$$1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x) = \lim_{n\to\infty} f_{1,n}(x_1) \dots f_{d,n}(x_d) \text{ for all } x \in \mathbb{R}^d.$$

This shows that $1_{(a_1,b_1]\times\cdots\times(a_d,b_d]}$ is $\mathcal{M}_2$ – measurable and therefore $(a_1,b_1]\times\cdots\times(a_d,b_d]\in\mathcal{M}_2$.

$\mathcal{M}_3$ case. This is easy since $\mathcal{B}_{\mathbb{R}^d}=\mathcal{M}_2\subset\mathcal{M}_3\subset\mathcal{B}_{\mathbb{R}^d}$.

$\mathcal{M}_4$ case. Let $\pi_j:\mathbb{R}^d\to\mathbb{R}$ be projection onto the $j^{\text{th}}$ – factor, then for $\lambda>0$, $e_\lambda\circ\pi_j(x)=e^{i\lambda x_j}$. It then follows that

$$\begin{aligned}\sigma(e_\lambda\circ\pi_j:\lambda>0)&=\sigma\left((e_\lambda\circ\pi_j)^{-1}(W):\lambda>0,W\in\mathcal{B}_{\mathbb{C}}\right)\\&=\sigma\left(\pi_j^{-1}\left(e_\lambda^{-1}(W)\right):\lambda>0,W\in\mathcal{B}_{\mathbb{C}}\right)\\&=\pi_j^{-1}\left(\sigma\left(\left(e_\lambda^{-1}(W)\right):\lambda>0,W\in\mathcal{B}_{\mathbb{C}}\right)\right)=\pi_j^{-1}(\mathcal{B}_{\mathbb{R}})\end{aligned}$$

wherein we have used Lemma 10.7 for the last equality. Since $\sigma(e_\lambda\circ\pi_j:\lambda>0)\subset\mathcal{M}_4$ for each $j$ we must have

$$\mathcal{B}_{\mathbb{R}^d}=\overbrace{\mathcal{B}_{\mathbb{R}}\otimes\cdots\otimes\mathcal{B}_{\mathbb{R}}}^{d\text{ times}}=\sigma(\pi_j:1\le j\le d)\subset\mathcal{M}_4.$$

**Alternative proof.** By Lemma 10.15 below there exists $g_n\in$Trig$(\mathbb{R})$ such that $\lim_{n\to\infty}g_n=1_{(a,b]}$. Since $x\to g_n(x_i)$ is in the span $\left\{x\to e^{i\lambda\cdot x}:\lambda\in\mathbb{R}^d\right\}$ for each $n$, it follows that $x\to 1_{(a,b]}(x_i)$ is $\mathcal{M}_4$ – measurable for all $-\infty<a<b<\infty$. Therefore, just as in the proof of case 1., we may now conclude that $\mathcal{B}_{\mathbb{R}^d}\subset\mathcal{M}_4$. ■

**Corollary 10.10.** *Suppose that $\mathbb{H}$ is a subspace of complex valued functions on $\mathbb{R}^d$ which is closed under complex conjugation and bounded convergence. If $\mathbb{H}$ contains any one of the following collection of functions;*

1. $\mathbb{M}:=\{x\to f_1(x_1)\dots f_d(x_d):f_i\in C_c(\mathbb{R})\}$
2. $\mathbb{M}:=C_c\left(\mathbb{R}^d\right)$*, or*
3. $\mathbb{M}:=\left\{x\to e^{i\lambda\cdot x}:\lambda\in\mathbb{R}^d\right\}$

*then $\mathbb{H}$ contains all bounded complex Borel measurable functions on $\mathbb{R}^d$.*

**Proof.** Observe that if $f\in C_c(\mathbb{R})$ such that $f(x)=1$ in a neighborhood of 0, then $f_n(x):=f(x/n)\to 1$ as $n\to\infty$. Therefore in cases 1. and 2., $\mathbb{H}$ contains the constant function, 1, since

$$1=\lim_{n\to\infty}f_n(x_1)\dots f_n(x_d).$$

In case 3, $1\in\mathbb{M}\subset\mathbb{H}$ as well. The result now follows from Theorem 10.5 and Corollary 10.9. ■

**Proposition 10.11 (Change of Variables Formula).** *Suppose that $-\infty<a<b<\infty$ and $u:[a,b]\to\mathbb{R}$ is a continuously differentiable function. Let $[c,d]=u([a,b])$ where $c=\min u([a,b])$ and $d=\max u([a,b])$. (By the intermediate value theorem $u([a,b])$ is an interval.) Then for all bounded measurable functions, $f:[c,d]\to\mathbb{R}$ we have*

$$\int_{u(a)}^{u(b)}f(x)\,dx=\int_a^b f(u(t))\,\dot u(t)\,dt. \tag{10.2}$$

*Moreover, Eq. (10.2) is also valid if $f:[c,d]\to\mathbb{R}$ is measurable and*

$$\int_a^b|f(u(t))|\,|\dot u(t)|\,dt<\infty. \tag{10.3}$$

**Proof.** Let $\mathbb{H}$ denote the space of bounded measurable functions such that Eq. (10.2) holds. It is easily checked that $\mathbb{H}$ is a linear space closed under bounded convergence. Next we show that $\mathbb{M}=C([c,d],\mathbb{R})\subset\mathbb{H}$ which coupled with Corollary 10.10 will show that $\mathbb{H}$ contains all bounded measurable functions from $[c,d]$ to $\mathbb{R}$.

If $f:[c,d]\to\mathbb{R}$ is a continuous function and let $F$ be an anti-derivative of $f$. Then by the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(u(t))\,\dot u(t)\,dt&=\int_a^b F'(u(t))\,\dot u(t)\,dt\\&=\int_a^b\frac{d}{dt}F(u(t))\,dt=F(u(t))\,|_a^b\\&=F(u(b))-F(u(a))=\int_{u(a)}^{u(b)}F'(x)\,dx=\int_{u(a)}^{u(b)}f(x)\,dx.\end{aligned}$$

Thus $\mathbb{M}\subset\mathbb{H}$ and the first assertion of the proposition is proved.

Now suppose that $f:[c,d]\to\mathbb{R}$ is measurable and Eq. (10.3) holds. For $M<\infty$, let $f_M(x)=f(x)\cdot 1_{|f(x)|\le M}$ – a bounded measurable function. Therefore applying Eq. (10.2) with $f$ replaced by $|f_M|$ shows,

$$\left|\int_{u(a)}^{u(b)}|f_M(x)|\,dx\right|=\left|\int_a^b|f_M(u(t))|\,\dot u(t)\,dt\right|\le\int_a^b|f_M(u(t))|\,|\dot u(t)|\,dt.$$

Using the MCT, we may let $M\uparrow\infty$ in the previous inequality to learn

$$\left|\int_{u(a)}^{u(b)}|f(x)|\,dx\right|\le\int_a^b|f(u(t))|\,|\dot u(t)|\,dt<\infty.$$

Now apply Eq. (10.2) with $f$ replaced by $f_M$ to learn

$$\int_{u(a)}^{u(b)}f_M(x)\,dx=\int_a^b f_M(u(t))\,\dot u(t)\,dt.$$

Using the DCT we may now let $M \to \infty$ in this equation to show that Eq. (10.2) remains valid. ■

**Exercise 10.2.** Suppose that $u : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that $\dot{u}(t) \geq 0$ for all $t$ and $\lim_{t\to\pm\infty} u(t) = \pm\infty$. Show that

$$\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} f(u(t))\,\dot{u}(t)\,dt \tag{10.4}$$

for all measurable functions $f : \mathbb{R} \to [0,\infty]$. In particular applying this result to $u(t) = at + b$ where $a > 0$ implies,

$$\int_{\mathbb{R}} f(x)\,dx = a \int_{\mathbb{R}} f(at+b)\,dt.$$

**Definition 10.12.** *The* ***Fourier transform*** *or* ***characteristic function*** *of a finite measure, $\mu$, on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, is the function, $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^d} e^{i\lambda\cdot x} d\mu(x) \text{ for all } \lambda \in \mathbb{R}^d$$

**Corollary 10.13.** *Suppose that $\mu$ and $\nu$ are two probability measures on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Then any one of the next three conditions implies that $\mu = \nu$;*

1. $\int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\nu(x) = \int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\mu(x)$ *for all* $f_i \in C_c(\mathbb{R})$.
2. $\int_{\mathbb{R}^d} f(x)\,d\nu(x) = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$ *for all* $f \in C_c(\mathbb{R}^d)$.
3. $\hat{\nu} = \hat{\mu}$.

*Item 3. asserts that the Fourier transform is injective.*

**Proof.** Let $\mathbb{H}$ be the collection of bounded complex measurable functions from $\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu. \tag{10.5}$$

It is easily seen that $\mathbb{H}$ is a linear space closed under complex conjugation and bounded convergence (by the DCT). Since $\mathbb{H}$ contains one of the multiplicative systems appearing in Corollary 10.10, it contains all bounded Borel measurable functions form $\mathbb{R}^d \to \mathbb{C}$. Thus we may take $f = 1_A$ with $A \in \mathcal{B}_{\mathbb{R}^d}$ in Eq. (10.5) to learn, $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$. ■

In many cases we can replace the condition in item 3. of Corollary 10.13 by;

$$\int_{\mathbb{R}^d} e^{\lambda\cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{\lambda\cdot x} d\nu(x) \text{ for all } \lambda \in U, \tag{10.6}$$

where $U$ is a neighborhood of $0 \in \mathbb{R}^d$. In order to do this, one must assume at least assume that the integrals involved are finite for all $\lambda \in U$. The idea is to show that Condition 10.6 implies $\hat{\nu} = \hat{\mu}$. You are asked to carry out this argument in Exercise 10.3 making use of the following lemma.

**Lemma 10.14 (Analytic Continuation).** *Let $\varepsilon > 0$ and $S_\varepsilon := \{x + iy \in \mathbb{C} : |x| < \varepsilon\}$ be an $\varepsilon$ strip in $\mathbb{C}$ about the imaginary axis. Suppose that $h : S_\varepsilon \to \mathbb{C}$ is a function such that for each $b \in \mathbb{R}$, there exists $\{c_n(b)\}_{n=0}^\infty \subset \mathbb{C}$ such that*

$$h(z + ib) = \sum_{n=0}^{\infty} c_n(b) z^n \text{ for all } |z| < \varepsilon. \tag{10.7}$$

*If $c_n(0) = 0$ for all $n \in \mathbb{N}_0$, then $h \equiv 0$.*

**Proof.** It suffices to prove the following assertion; if for some $b \in \mathbb{R}$ we know that $c_n(b) = 0$ for all $n$, then $c_n(y) = 0$ for all $n$ and $y \in (b-\varepsilon, b+\varepsilon)$. We now prove this assertion.

Let us assume that $b \in \mathbb{R}$ and $c_n(b) = 0$ for all $n \in \mathbb{N}_0$. It then follows from Eq. (10.7) that $h(z+ib) = 0$ for all $|z| < \varepsilon$. Thus if $|y - b| < \varepsilon$, we may conclude that $h(x+iy) = 0$ for $x$ in a (possibly very small) neighborhood $(-\delta, \delta)$ of 0. Since

$$\sum_{n=0}^{\infty} c_n(y) x^n = h(x+iy) = 0 \text{ for all } |x| < \delta,$$

it follows that

$$0 = \frac{1}{n!}\frac{d^n}{dx^n} h(x+iy)|_{x=0} = c_n(y)$$

and the proof is complete. ■

**Lemma 10.15 (This may be omitted.).** *Suppose that $-\infty < a < b < \infty$ and let Trig($\mathbb{R}$) $\subset C(\mathbb{R},\mathbb{C})$ be the complex linear span of $\{x \to e^{i\lambda x} : \lambda \in \mathbb{R}\}$. Then there exists $f_n \in C_c(\mathbb{R},[0,1])$ and $g_n \in$ Trig($\mathbb{R}$) such that $\lim_{n\to\infty} f_n(x) = 1_{(a,b]}(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$.*

**Proof.** The assertion involving $f_n \in C_c(\mathbb{R},[0,1])$ was the content of one of your homework assignments. For the assertion involving $g_n \in$Trig($\mathbb{R}$), it will suffice to show that any $f \in C_c(\mathbb{R})$ may be written as $f(x) = \lim_{n\to\infty} g_n(x)$ for some $\{g_n\} \subset$Trig($\mathbb{R}$) where the limit is uniform for $x$ in compact subsets of $\mathbb{R}$.

So suppose that $f \in C_c(\mathbb{R})$ and $L > 0$ such that $f(x) = 0$ if $|x| \geq L/4$. Then

$$f_L(x) := \sum_{n=-\infty}^{\infty} f(x+nL)$$

is a continuous $L$ – periodic function on $\mathbb{R}$, see Figure 10.3. If $\varepsilon > 0$ is given, we may apply Theorem 5.58 to find $\Lambda \subset_f \mathbb{Z}$ such that

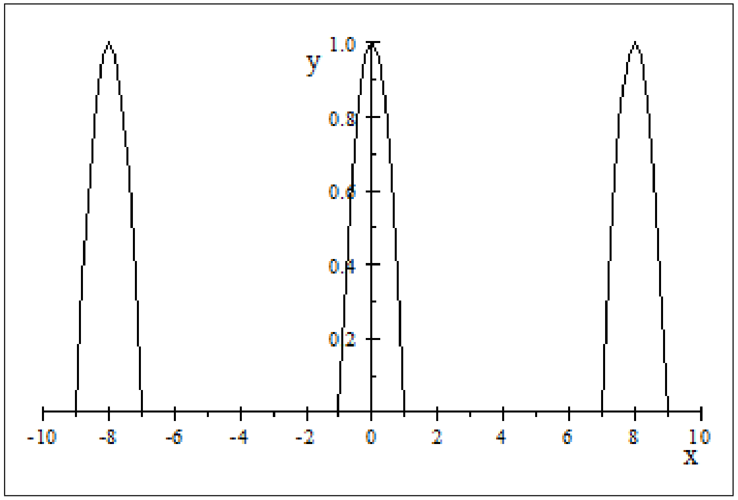



**Fig. 10.3.** This is plot of $f_8(x)$ where $f(x) = \left(1 - x^2\right) 1_{|x| \leq 1}$. The center hump by itself would be the plot of $f(x)$.

$$\left| f_L\left(\frac{L}{2\pi} x\right) - \sum_{\alpha \in \Lambda} a_\lambda e^{i\alpha x} \right| \leq \varepsilon \text{ for all } x \in \mathbb{R},$$

wherein we have use the fact that $x \to f_L\left(\frac{L}{2\pi} x\right)$ is a $2\pi$ – periodic function of $x$. Equivalently we have,

$$\max_x \left| f_L(x) - \sum_{\alpha \in \Lambda} a_\lambda e^{i\frac{2\pi\alpha}{L} x} \right| \leq \varepsilon.$$

In particular it follows that $f_L(x)$ is a uniform limit of functions from Trig($\mathbb{R}$). Since $\lim_{L\to\infty} f_L(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$, it is easy to conclude there exists $g_n \in$Trig($\mathbb{R}$) such that $\lim_{n\to\infty} g_n(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$. ■

## 10.2 Exercises

**Exercise 10.3.** Suppose $\varepsilon > 0$ and $X$ and $Y$ are two random variables such that $\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{tY}\right] < \infty$ for all $|t| \leq \varepsilon$. Show;

1. $\mathbb{E}\left[e^{\varepsilon|X|}\right]$ and $\mathbb{E}\left[e^{\varepsilon|Y|}\right]$ are finite.
2. $\mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{itY}\right]$ for all $t \in \mathbb{R}$. **Hint:** Consider $h(z) := \mathbb{E}\left[e^{zX}\right] - \mathbb{E}\left[e^{zY}\right]$ for $z \in S_\varepsilon$. Now show for $|z| \leq \varepsilon$ and $b \in \mathbb{R}$, that

$$h(z + ib) = \mathbb{E}\left[e^{ibX} e^{zX}\right] - \mathbb{E}\left[e^{ibY} e^{zY}\right] = \sum_{n=0}^{\infty} c_n(b) z^n \tag{10.8}$$

where

$$c_n(b) := \frac{1}{n!}\left(\mathbb{E}\left[e^{ibX} X^n\right] - \mathbb{E}\left[e^{ibY} Y^n\right]\right). \tag{10.9}$$

3. Conclude from item 2. that $X \overset{d}{=} Y$, i.e. that $\mathrm{Law}_P(X) = \mathrm{Law}_P(Y)$.

**Exercise 10.4.** Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be a pair of random variables such that

$$\mathbb{E}[f(X) g(Y)] = \mathbb{E}[f(X) g(X)]$$

for every pair of bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$. Show $P(X = Y) = 1$. **Hint:** Let $\mathbb{H}$ denote the bounded Borel measurable functions, $h : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbb{E}[h(X, Y)] = \mathbb{E}[h(X, X)].$$

Use Theorem 10.2 to show $\mathbb{H}$ is the vector space of all bounded Borel measurable functions. Then take $h(x, y) = 1_{\{x=y\}}$.

**Exercise 10.5 (Density of $\mathcal{A}$ – simple functions).** Let $(\Omega, \mathcal{B}, P)$ be a probability space and assume that $\mathcal{A}$ is a sub-algebra of $\mathcal{B}$ such that $\mathcal{B} = \sigma(\mathcal{A})$. Let $\mathbb{H}$ denote the bounded measurable functions $f : \Omega \to \mathbb{R}$ such that for every $\varepsilon > 0$ there exists an an $\mathcal{A}$ – simple function, $\varphi : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f - \varphi| < \varepsilon$. Show $\mathbb{H}$ consists of all bounded measurable functions, $f : \Omega \to \mathbb{R}$. **Hint:** let $\mathbb{M}$ denote the collection of $\mathcal{A}$ – simple functions.

**Corollary 10.16.** *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $\{X_n\}_{n=1}^{\infty}$ is a collection of random variables on $\Omega$, and $\mathcal{B}_\infty := \sigma(X_1, X_2, X_3, \dots)$. Then for all $\varepsilon > 0$ and all bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, there exists an $n \in \mathbb{N}$ and a bounded $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}|f - G(X_1, \dots, X_n)| < \varepsilon$. Moreover we may assume that $\sup_{x\in\mathbb{R}^n} |G(x)| \leq M := \sup_{\omega\in\Omega} |f(\omega)|$.*

**Proof.** Apply Exercise 10.5 with $\mathcal{A} := \cup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$ in order to find an $\mathcal{A}$ – measurable simple function, $\varphi$, such that $\mathbb{E}|f - \varphi| < \varepsilon$. By the definition of $\mathcal{A}$ we know that $\varphi$ is $\sigma(X_1, \dots, X_n)$ – measurable for some $n \in \mathbb{N}$. It now follows by the factorization Lemma 8.40 that $\varphi = G(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$. If necessary, replace $G$ by $[G \wedge M] \vee (-M)$ in order to insure $\sup_{x\in\mathbb{R}^n} |G(x)| \leq M$. ■

**Exercise 10.6 (Density of $\mathcal{A}$ in $\mathcal{B} = \sigma(\mathcal{A})$).** Keeping the same notation as in Exercise 10.5 but now take $f = 1_B$ for some $B \in \mathcal{B}$ and given $\varepsilon > 0$, write

$\varphi = \sum_{i=0}^{n} \lambda_i 1_{A_i}$ where $\lambda_0 = 0$, $\{\lambda_i\}_{i=1}^n$ is an enumeration of $\varphi(\Omega) \setminus \{0\}$, and $A_i := \{\varphi = \lambda_i\}$. Show; 1.

$$\mathbb{E}|1_B - \varphi| = P(A_0 \cap B) + \sum_{i=1}^{n} [|1-\lambda_i| P(B \cap A_i) + |\lambda_i| P(A_i \setminus B)] \tag{10.10}$$

$$\geq P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\}. \tag{10.11}$$

2. Now let $\psi = \sum_{i=0}^{n} \alpha_i 1_{A_i}$ with

$$\alpha_i = \begin{cases} 1 \text{ if } P(A_i \setminus B) \leq P(B \cap A_i) \\ 0 \text{ if } P(A_i \setminus B) > P(B \cap A_i) \end{cases}.$$

Then show that

$$\mathbb{E}|1_B - \psi| = P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\} \leq \mathbb{E}|1_B - \varphi|.$$

Observe that $\psi = 1_D$ where $D = \cup_{i:\alpha_i=1} A_i \in \mathcal{A}$ and so you have shown; for every $\varepsilon > 0$ there exists a $D \in \mathcal{A}$ such that

$$P(B \Delta D) = \mathbb{E}|1_B - 1_D| < \varepsilon.$$

**Exercise 10.7.** Suppose that $\{(X_i, \mathcal{B}_i)\}_{i=1}^n$ are measurable spaces and for each $i$, $\mathbb{M}_i$ is a multiplicative system of real bounded measurable functions on $X_i$ such that $\sigma(\mathbb{M}_i) = \mathcal{B}_i$ and there exist $\chi_n \in \mathbb{M}_i$ such that $\chi_n \to 1$ boundedly as $n \to \infty$. Given $f_i : X_i \to \mathbb{R}$ let $f_1 \otimes \cdots \otimes f_n : X_1 \times \cdots \times X_n \to \mathbb{R}$ be defined by

$$(f_1 \otimes \cdots \otimes f_n)(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

Show

$$\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n := \{f_1 \otimes \cdots \otimes f_n : f_i \in \mathbb{M}_i \text{ for } 1 \leq i \leq n\}$$

is a multiplicative system of bounded measurable functions on $(X := X_1 \times \cdots \times X_n, \mathcal{B} := \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$ such that $\sigma(\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n) = \mathcal{B}$.

### 10.2.1 Obsolete stuff follows.

**Notation 10.17** *Suppose $\mathbb{M}$ is a subset of $\ell^\infty(X, \mathbb{R})$.*

1. *Let $\mathcal{H}(\mathbb{M})$ denote the smallest subspace of $\ell^\infty(X, \mathbb{R})$ which contains $\mathbb{M}$, the constant functions, and is closed under bounded convergence.*
2. *Let $\mathcal{H}_\sigma(\mathbb{M})$ denote the smallest $\sigma$ – function algebra containing $\mathbb{M}$.*

**Exercise 10.8.** Let $X = \{1, 2, 3, 4\}$, $A = \{1, 2\}$, $B = \{2, 3\}$ and $\mathbb{M} := \{1_A, 1_B\}$. Show $\mathcal{H}_\sigma(\mathbb{M}) \neq \mathcal{H}(\mathbb{M})$ in this case.

## 10.3 A Strengthening of the Multiplicative System Theorem*

**Notation 10.18** *We say that $\mathbb{H} \subset \ell^\infty(\Omega, \mathbb{R})$ is **closed under monotone convergence** if; for every sequence, $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $0 \leq f_n(\omega) \leq M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f_n(\omega)$ is increasing in $n$ for all $\omega \in \Omega$, then $f := \lim_{n\to\infty} f_n \in \mathbb{H}$.*

Clearly if $\mathbb{H}$ is closed under bounded convergence then it is also closed under monotone convergence. I learned the proof of the converse from Pat Fitzsimmons but this result appears in Sharpe [41, p. 365].

**Proposition 10.19.** **Let $\Omega$ be a set. Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions from $\Omega$ to $\mathbb{R}$ which is closed under monotone convergence. Then $\mathbb{H}$ is closed under uniform convergence as well, i.e. $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ with $\sup_{n\in\mathbb{N}} \sup_{\omega\in\Omega} |f_n(\omega)| < \infty$ and $f_n \to f$, then $f \in \mathbb{H}$.*

**Proof.** Let us first assume that $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ such that $f_n$ converges uniformly to a bounded function, $f : \Omega \to \mathbb{R}$. Let $\|f\|_\infty := \sup_{\omega\in\Omega} |f(\omega)|$. Let $\varepsilon > 0$ be given. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_\infty \leq \varepsilon 2^{-(n+1)}$. Let

$$g_n := f_n - \delta_n + M$$

with $\delta_n$ and $M$ constants to be determined shortly. We then have

$$g_{n+1} - g_n = f_{n+1} - f_n + \delta_n - \delta_{n+1} \geq -\varepsilon 2^{-(n+1)} + \delta_n - \delta_{n+1}.$$

Taking $\delta_n := \varepsilon 2^{-n}$, then $\delta_n - \delta_{n+1} = \varepsilon 2^{-n}(1 - 1/2) = \varepsilon 2^{-(n+1)}$ in which case $g_{n+1} - g_n \geq 0$ for all $n$. By choosing $M$ sufficiently large, we will also have $g_n \geq 0$ for all $n$. Since $\mathbb{H}$ is a vector space containing the constant functions, $g_n \in \mathbb{H}$ and since $g_n \uparrow f + M$, it follows that $f = f + M - M \in \mathbb{H}$. So we have shown that $\mathbb{H}$ is closed under uniform convergence. ■

This proposition immediately leads to the following strengthening of Theorem 10.2.

**Theorem 10.20.** **Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions on $\Omega$ which contains the constant functions and is closed under monotone convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

**Proof.** Proposition 10.19 reduces this theorem to Theorem 10.2. ■

## 10.4 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

**Notation 10.21** *Given a collection of bounded functions, $\mathbb{M}$, from a set, $\Omega$, to $\mathbb{R}$, let $\mathbb{M}_\uparrow$ ($\mathbb{M}_\downarrow$) denote the the bounded monotone increasing (decreasing) limits of functions from $\mathbb{M}$. More explicitly a bounded function, $f : \Omega \to \mathbb{R}$ is in $\mathbb{M}_\uparrow$ respectively $\mathbb{M}_\downarrow$ iff there exists $f_n \in \mathbb{M}$ such that $f_n \uparrow f$ respectively $f_n \downarrow f$.*

**Theorem 10.22 (Bounded Approximation Theorem*).** *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that:*

1. $\sigma(\mathbb{M}) = \mathcal{B}$,
2. $1 \in \mathbb{M}$, *and*
3. $|f| \in \mathbb{M}$ *for all* $f \in \mathbb{M}$.

*Then for every bounded $\sigma(\mathbb{M})$ measurable function, $g : \Omega \to \mathbb{R}$, and every $\varepsilon > 0$, there exists $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$.*[1]

**Proof.** Let us begin with a few simple observations.

1. $\mathbb{M}$ is a "lattice" – if $f, g \in \mathbb{M}$ then
$$f \vee g = \frac{1}{2}(f + g + |f - g|) \in \mathbb{M}$$
and
$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \in \mathbb{M}.$$
2. If $f, g \in \mathbb{M}_\uparrow$ or $f, g \in \mathbb{M}_\downarrow$ then $f + g \in \mathbb{M}_\uparrow$ or $f + g \in \mathbb{M}_\downarrow$ respectively.
3. If $\lambda \ge 0$ and $f \in \mathbb{M}_\uparrow$ ($f \in \mathbb{M}_\downarrow$), then $\lambda f \in \mathbb{M}_\uparrow$ ($\lambda f \in \mathbb{M}_\downarrow$).
4. If $f \in \mathbb{M}_\uparrow$ then $-f \in \mathbb{M}_\downarrow$ and visa versa.
5. If $f_n \in \mathbb{M}_\uparrow$ and $f_n \uparrow f$ where $f : \Omega \to \mathbb{R}$ is a bounded function, then $f \in \mathbb{M}_\uparrow$. Indeed, by assumption there exists $f_{n,i} \in \mathbb{M}$ such that $f_{n,i} \uparrow f_n$ as $i \to \infty$. By observation (1), $g_n := \max\{f_{ij} : i, j \le n\} \in \mathbb{M}$. Moreover it is clear that $g_n \le \max\{f_k : k \le n\} = f_n \le f$ and hence $g_n \uparrow g := \lim_{n\to\infty} g_n \le f$. Since $f_{ij} \le g$ for all $i, j$, it follows that $f_n = \lim_{j\to\infty} f_{nj} \le g$ and consequently that $f = \lim_{n\to\infty} f_n \le g \le f$. So we have shown that $g_n \uparrow f \in \mathbb{M}_\uparrow$.

Now let $\mathbb{H}$ denote the collection of bounded measurable functions which satisfy the assertion of the theorem. Clearly, $\mathbb{M} \subset \mathbb{H}$ and in fact it is also easy to see that $\mathbb{M}_\uparrow$ and $\mathbb{M}_\downarrow$ are contained in $\mathbb{H}$ as well. For example, if $f \in \mathbb{M}_\uparrow$, by definition, there exists $f_n \in \mathbb{M} \subset \mathbb{M}_\downarrow$ such that $f_n \uparrow f$. Since $\mathbb{M}_\downarrow \ni f_n \le f \le f \in \mathbb{M}_\uparrow$ and $\mu(f - f_n) \to 0$ by the dominated convergence theorem, it follows that $f \in \mathbb{H}$. As similar argument shows $\mathbb{M}_\downarrow \subset \mathbb{H}$. We will now show $\mathbb{H}$ is a vector sub-space of the bounded $\mathcal{B} = \sigma(\mathbb{M})$ – measurable functions.

$\mathbb{H}$ **is closed under addition.** If $g_i \in \mathbb{H}$ for $i = 1, 2$, and $\varepsilon > 0$ is given, we may find $f_i \in \mathbb{M}_\downarrow$ and $h_i \in \mathbb{M}_\uparrow$ such that $f_i \le g_i \le h_i$ and $\mu(h_i - f_i) < \varepsilon/2$ for $i = 1, 2$. Since $h = h_1 + h_2 \in \mathbb{M}_\uparrow$, $f := f_1 + f_2 \in \mathbb{M}_\downarrow$, $f \le g_1 + g_2 \le h$, and

$$\mu(h - f) = \mu(h_1 - f_1) + \mu(h_2 - f_2) < \varepsilon,$$

it follows that $g_1 + g_2 \in \mathbb{H}$.

$\mathbb{H}$ **is closed under scalar multiplication.** If $g \in \mathbb{H}$ then $\lambda g \in \mathbb{H}$ for all $\lambda \in \mathbb{R}$. Indeed suppose that $\varepsilon > 0$ is given and $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$. Then for $\lambda \ge 0$, $\mathbb{M}_\downarrow \ni \lambda f \le \lambda g \le \lambda h \in \mathbb{M}_\uparrow$ and

$$\mu(\lambda h - \lambda f) = \lambda \mu(h - f) < \lambda \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, if follows that $\lambda g \in \mathbb{H}$ for $\lambda \ge 0$. Similarly, $\mathbb{M}_\downarrow \ni -h \le -g \le -f \in \mathbb{M}_\uparrow$ and

$$\mu(-f - (-h)) = \mu(h - f) < \varepsilon.$$

which shows $-g \in \mathbb{H}$ as well.

Because of Theorem 10.20, to complete this proof, it suffices to show $\mathbb{H}$ is closed under monotone convergence. So suppose that $g_n \in \mathbb{H}$ and $g_n \uparrow g$, where $g : \Omega \to \mathbb{R}$ is a bounded function. Since $\mathbb{H}$ is a vector space, it follows that $0 \le \delta_n := g_{n+1} - g_n \in \mathbb{H}$ for all $n \in \mathbb{N}$. So if $\varepsilon > 0$ is given, we can find, $\mathbb{M}_\downarrow \ni u_n \le \delta_n \le v_n \in \mathbb{M}_\uparrow$ such that $\mu(v_n - u_n) \le 2^{-n}\varepsilon$ for all $n$. By replacing $u_n$ by $u_n \vee 0 \in \mathbb{M}_\downarrow$ (by observation 1.), we may further assume that $u_n \ge 0$. Let

$$v := \sum_{n=1}^\infty v_n =\uparrow \lim_{N\to\infty} \sum_{n=1}^N v_n \in \mathbb{M}_\uparrow \text{ (using observations 2. and 5.)}$$

and for $N \in \mathbb{N}$, let

$$u^N := \sum_{n=1}^N u_n \in \mathbb{M}_\downarrow \text{ (using observation 2).}$$

Then

$$\sum_{n=1}^\infty \delta_n = \lim_{N\to\infty} \sum_{n=1}^N \delta_n = \lim_{N\to\infty} (g_{N+1} - g_1) = g - g_1$$

and $u^N \le g - g_1 \le v$. Moreover,

[1] Bruce: rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

$$\begin{aligned}\mu\left(v-u^{N}\right) &= \sum_{n=1}^{N}\mu\left(v_{n}-u_{n}\right)+\sum_{n=N+1}^{\infty}\mu\left(v_{n}\right)\leq\sum_{n=1}^{N}\varepsilon2^{-n}+\sum_{n=N+1}^{\infty}\mu\left(v_{n}\right)\\ &\leq\varepsilon+\sum_{n=N+1}^{\infty}\mu\left(v_{n}\right).\end{aligned}$$

However, since

$$\begin{aligned}\sum_{n=1}^{\infty}\mu\left(v_{n}\right) &\leq\sum_{n=1}^{\infty}\mu\left(\delta_{n}+\varepsilon2^{-n}\right)=\sum_{n=1}^{\infty}\mu\left(\delta_{n}\right)+\varepsilon\mu\left(\varOmega\right)\\ &=\sum_{n=1}^{\infty}\mu\left(g-g_{1}\right)+\varepsilon\mu\left(\varOmega\right)<\infty,\end{aligned}$$

it follows that for $N\in\mathbb{N}$ sufficiently large that $\sum_{n=N+1}^{\infty}\mu\left(v_{n}\right)<\varepsilon$. Therefore, for this $N$, we have $\mu\left(v-u^{N}\right)<2\varepsilon$ and since $\varepsilon>0$ is arbitrary, if follows that $g-g_{1}\in\mathbb{H}$. Since $g_{1}\in\mathbb{H}$ and $\mathbb{H}$ is a vector space, we may conclude that $g=\left(g-g_{1}\right)+g_{1}\in\mathbb{H}$. ∎

# 11

# Multiple and Iterated Integrals

## 11.1 Iterated Integrals

**Notation 11.1 (Iterated Integrals)** *If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are two measure spaces and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function, the **iterated integrals** of $f$ (when they make sense) are:*

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) := \int_X \left[ \int_Y f(x,y) d\nu(y) \right] d\mu(x)$$

*and*

$$\int_Y d\nu(y) \int_X d\mu(x) f(x,y) := \int_Y \left[ \int_X f(x,y) d\mu(x) \right] d\nu(y).$$

**Notation 11.2** *Suppose that $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ are functions, let $f \otimes g$ denote the function on $X \times Y$ given by*

$$f \otimes g(x,y) = f(x)g(y).$$

Notice that if $f, g$ are measurable, then $f \otimes g$ is $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{C}})$ – measurable. To prove this let $F(x,y) = f(x)$ and $G(x,y) = g(y)$ so that $f \otimes g = F \cdot G$ will be measurable provided that $F$ and $G$ are measurable. Now $F = f \circ \pi_1$ where $\pi_1 : X \times Y \to X$ is the projection map. This shows that $F$ is the composition of measurable functions and hence measurable. Similarly one shows that $G$ is measurable.

## 11.2 Tonelli's Theorem and Product Measure

**Theorem 11.3.** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are $\sigma$-finite measure spaces and $f$ is a nonnegative $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable function, then for each $y \in Y$,*

$$x \to f(x,y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{11.1}$$

*for each $x \in X$,*

$$y \to f(x,y) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{11.2}$$

$$x \to \int_Y f(x,y) d\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{11.3}$$

$$y \to \int_X f(x,y) d\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{11.4}$$

*and*

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \int_Y d\nu(y) \int_X d\mu(x) f(x,y). \tag{11.5}$$

**Proof.** Suppose that $E = A \times B \in \mathcal{E} := \mathcal{M} \times \mathcal{N}$ and $f = 1_E$. Then

$$f(x,y) = 1_{A \times B}(x,y) = 1_A(x) 1_B(y)$$

and one sees that Eqs. (11.1) and (11.2) hold. Moreover

$$\int_Y f(x,y) d\nu(y) = \int_Y 1_A(x) 1_B(y) d\nu(y) = 1_A(x) \nu(B),$$

so that Eq. (11.3) holds and we have

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \nu(B)\mu(A). \tag{11.6}$$

Similarly,

$$\int_X f(x,y) d\mu(x) = \mu(A) 1_B(y) \text{ and}$$
$$\int_Y d\nu(y) \int_X d\mu(x) f(x,y) = \nu(B)\mu(A)$$

from which it follows that Eqs. (11.4) and (11.5) hold in this case as well.

For the moment let us now further assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ and let $\mathbb{H}$ be the collection of all bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable functions on $X \times Y$ such that Eqs. (11.1) – (11.5) hold. Using the fact that measurable functions are closed under pointwise limits and the dominated convergence theorem (the dominating function always being a constant), one easily shows that $\mathbb{H}$ closed under bounded convergence. Since we have just verified that $1_E \in \mathbb{H}$ for all $E$ in the $\pi$ – class, $\mathcal{E}$, it follows by Corollary 10.3 that $\mathbb{H}$ is the space of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$. Moreover, if $f:X\times Y\to[0,\infty]$ is a $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, let $f_M=M\wedge f$ so that $f_M\uparrow f$ as $M\to\infty$. Then Eqs. (11.1) – (11.5) hold with $f$ replaced by $f_M$ for all $M\in\mathbb{N}$. Repeated use of the monotone convergence theorem allows us to pass to the limit $M\to\infty$ in these equations to deduce the theorem in the case $\mu$ and $\nu$ are finite measures.

For the $\sigma$ – finite case, choose $X_n\in\mathcal{M}$, $Y_n\in\mathcal{N}$ such that $X_n\uparrow X$, $Y_n\uparrow Y$, $\mu(X_n)<\infty$ and $\nu(Y_n)<\infty$ for all $m,n\in\mathbb{N}$. Then define $\mu_m(A)=\mu(X_m\cap A)$ and $\nu_n(B)=\nu(Y_n\cap B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ or equivalently $d\mu_m=1_{X_m}d\mu$ and $d\nu_n=1_{Y_n}d\nu$. By what we have just proved Eqs. (11.1) – (11.5) with $\mu$ replaced by $\mu_m$ and $\nu$ by $\nu_n$ for all $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable functions, $f:X\times Y\to[0,\infty]$. The validity of Eqs. (11.1) – (11.5) then follows by passing to the limits $m\to\infty$ and then $n\to\infty$ making use of the monotone convergence theorem in the following context. For all $u\in L^+(X,\mathcal{M})$,

$$\int_X ud\mu_m=\int_X u1_{X_m}d\mu\uparrow\int_X ud\mu \text{ as } m\to\infty,$$

and for all and $v\in L^+(Y,\mathcal{N})$,

$$\int_Y vd\mu_n=\int_Y v1_{Y_n}d\mu\uparrow\int_Y vd\mu \text{ as } n\to\infty.$$

■

**Corollary 11.4.** *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are $\sigma$ – finite measure spaces. Then there exists a unique measure $\pi$ on $\mathcal{M}\otimes\mathcal{N}$ such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$. Moreover $\pi$ is given by*

$$\pi(E)=\int_X d\mu(x)\int_Y d\nu(y)1_E(x,y)=\int_Y d\nu(y)\int_X d\mu(x)1_E(x,y) \tag{11.7}$$

*for all $E\in\mathcal{M}\otimes\mathcal{N}$ and $\pi$ is $\sigma$ – finite.*

**Proof.** Notice that any measure $\pi$ such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ is necessarily $\sigma$ – finite. Indeed, let $X_n\in\mathcal{M}$ and $Y_n\in\mathcal{N}$ be chosen so that $\mu(X_n)<\infty$, $\nu(Y_n)<\infty$, $X_n\uparrow X$ and $Y_n\uparrow Y$, then $X_n\times Y_n\in\mathcal{M}\otimes\mathcal{N}$, $X_n\times Y_n\uparrow X\times Y$ and $\pi(X_n\times Y_n)<\infty$ for all $n$. The uniqueness assertion is a consequence of the combination of Exercises 4.12 and 6.3 Proposition 4.27 with $\mathcal{E}=\mathcal{M}\times\mathcal{N}$. For the existence, it suffices to observe, using the monotone convergence theorem, that $\pi$ defined in Eq. (11.7) is a measure on $\mathcal{M}\otimes\mathcal{N}$. Moreover this measure satisfies $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ from Eq. (11.6). ■

**Notation 11.5** *The measure $\pi$ is called the product measure of $\mu$ and $\nu$ and will be denoted by $\mu\otimes\nu$.*

**Theorem 11.6 (Tonelli's Theorem).** *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are $\sigma$ – finite measure spaces and $\pi=\mu\otimes\nu$ is the product measure on $\mathcal{M}\otimes\mathcal{N}$. If $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$, then $f(\cdot,y)\in L^+(X,\mathcal{M})$ for all $y\in Y$, $f(x,\cdot)\in L^+(Y,\mathcal{N})$ for all $x\in X$,*

$$\int_Y f(\cdot,y)d\nu(y)\in L^+(X,\mathcal{M}),\ \int_X f(x,\cdot)d\mu(x)\in L^+(Y,\mathcal{N})$$

*and*

$$\int_{X\times Y} f\,d\pi=\int_X d\mu(x)\int_Y d\nu(y)f(x,y) \tag{11.8}$$

$$=\int_Y d\nu(y)\int_X d\mu(x)f(x,y). \tag{11.9}$$

**Proof.** By Theorem 11.3 and Corollary 11.4, the theorem holds when $f=1_E$ with $E\in\mathcal{M}\otimes\mathcal{N}$. Using the linearity of all of the statements, the theorem is also true for non-negative simple functions. Then using the monotone convergence theorem repeatedly along with the approximation Theorem 8.39, one deduces the theorem for general $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$. ■

*Example 11.7.* In this example we are going to show, $I:=\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)=\sqrt{2\pi}$. To this end we observe, using Tonelli's theorem, that

$$\begin{aligned}I^2&=\left[\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)\right]^2=\int_{\mathbb{R}}e^{-y^2/2}\left[\int_{\mathbb{R}}e^{-x^2/2}dm\,(x)\right]dm\,(y)\\&=\int_{\mathbb{R}^2}e^{-\left(x^2+y^2\right)/2}dm^2\,(x,y)\end{aligned}$$

where $m^2=m\otimes m$ is "Lebesgue measure" on $\left(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\right)$. From the monotone convergence theorem,

$$I^2=\lim_{R\to\infty}\int_{D_R}e^{-\left(x^2+y^2\right)/2}dm^2\,(x,y)$$

where $D_R=\left\{(x,y):x^2+y^2<R^2\right\}$. Using the change of variables theorem described in Section 11.5 below,[1] we find

$$\begin{aligned}\int_{D_R}e^{-\left(x^2+y^2\right)/2}d\pi\,(x,y)&=\int_{(0,R)\times(0,2\pi)}e^{-r^2/2}rdrd\theta\\&=2\pi\int_0^R e^{-r^2/2}rdr=2\pi\left(1-e^{-R^2/2}\right).\end{aligned}$$

[1] Alternatively, you can easily show that the integral $\int_{D_R}fdm^2$ agrees with the multiple integral in undergraduate analysis when $f$ is continuous. Then use the change of variables theorem from undergraduate analysis.

From this we learn that

$$I^2 = \lim_{R\to\infty} 2\pi \left(1 - e^{-R^2/2}\right) = 2\pi$$

as desired.

## 11.3 Fubini's Theorem

**Notation 11.8** *If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is any measurable function, let*

$$\bar{\int}_X f d\mu := \begin{cases} \int_X f d\mu & \text{if} \quad \int_X |f|\, d\mu < \infty \\ 0 & \text{otherwise.} \end{cases}$$

**Theorem 11.9 (Fubini's Theorem).** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are $\sigma$ – finite measure spaces, $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$ and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Then the following three conditions are equivalent:*

$$\int_{X\times Y} |f|\, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{11.10}$$

$$\int_X \left(\int_Y |f(x, y)|\, d\nu(y)\right) d\mu(x) < \infty \text{ and} \tag{11.11}$$

$$\int_Y \left(\int_X |f(x, y)|\, d\mu(x)\right) d\nu(y) < \infty. \tag{11.12}$$

*If any one (and hence all) of these condition hold, then $f(x, \cdot) \in L^1(\nu)$ for $\mu$-a.e. $x$, $f(\cdot, y) \in L^1(\mu)$ for $\nu$-a.e. $y$, $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$, $\bar{\int}_X f(x, \cdot)d\mu(x) \in L^1(\nu)$ and Eqs. (11.8) and (11.9) are still valid after putting a bar over the integral symbols.*

**Proof.** The equivalence of Eqs. (11.10) – (11.12) is a direct consequence of Tonelli's Theorem 11.6. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{x \in X : \int_Y |f(x, y)|\, d\nu(y) = \infty\right\}. \tag{11.13}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x, y)|\, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

$$\int_X \left[\int_Y |f(x, y)|\, d\nu(y)\right] d\mu(x) = \int_{X\times Y} |f|\, d\pi < \infty$$

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of $f$, then

$$\begin{aligned} \bar{\int}_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) [f_+(x, y) - f_-(x, y)]\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x, y)\, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x, y)\, d\nu(y). \end{aligned} \tag{11.14}$$

Noting that $1_{E^c}(x) f_\pm(x, y) = (1_{E^c} \otimes 1_Y \cdot f_\pm)(x, y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \bar{\int}_Y f(x, y)\, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left|\bar{\int}_Y f(x, y)\, d\nu(y)\right| d\mu(x) \le \int_X \left[\int_Y |f(x, y)|\, d\nu(y)\right] d\mu(x) < \infty,$$

which shows $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$. Integrating Eq. (11.14) on $x$ and using Tonelli's theorem repeatedly implies,

$$\begin{aligned} &\int_X \left[\bar{\int}_Y f(x, y)\, d\nu(y)\right] d\mu(x) \\ &= \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_+(x, y) - \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) f_-(x, y) \\ &= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-)\, d\pi = \int_{X\times Y} f d\pi \end{aligned} \tag{11.15}$$

which proves Eq. (11.8) holds.

Now suppose that $f = u + iv$ is complex valued and again let $E$ be as in Eq. (11.13). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) = 0$ and

$$\begin{aligned} \bar{\int}_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) = \int_Y 1_{E^c}(x) [u(x, y) + iv(x, y)]\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) u(x, y)\, d\nu(y) + i \int_Y 1_{E^c}(x) v(x, y)\, d\nu(y). \end{aligned}$$

The last line is a measurable in $x$ as we have just proved. Similarly one shows $\bar{\int}_Y f(\cdot, y)\, d\nu(y) \in L^1(\mu)$ and Eq. (11.8) still holds by a computation similar to that done in Eq. (11.15). The assertions pertaining to Eq. (11.9) may be proved in the same way. ■

The previous theorems generalize to products of any finite number of $\sigma$ – finite measure spaces.

**Theorem 11.10.** *Suppose $\{(X_i, \mathcal{M}_i, \mu_i)\}_{i=1}^n$ are $\sigma$ – finite measure spaces and $X := X_1 \times \cdots \times X_n$. Then there exists a unique measure $(\pi)$ on $(X, \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n)$ such that*

$$\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1) \ldots \mu_n(A_n) \text{ for all } A_i \in \mathcal{M}_i. \tag{11.16}$$

*(This measure and its completion will be denoted by $\mu_1 \otimes \cdots \otimes \mu_n$.) If $f : X \to [0, \infty]$ is a $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ – measurable function then*

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \cdots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1, \ldots, x_n) \tag{11.17}$$

*where $\sigma$ is any permutation of $\{1, 2, \ldots, n\}$. In particular $f \in L^1(\pi)$, iff*

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \cdots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ |f(x_1, \ldots, x_n)| < \infty$$

*for some (and hence all) permutations, $\sigma$. Furthermore, if $f \in L^1(\pi)$, then*

$$\int_X f d\pi = \bar{\int}_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \bar{\int}_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1, \ldots, x_n) \tag{11.18}$$

*for all permutations $\sigma$.*

**Proof.** (* I would consider skipping this tedious proof.) The proof will be by induction on $n$ with the case $n = 2$ being covered in Theorems 11.6 and 11.9. So let $n \geq 3$ and assume the theorem is valid for $n - 1$ factors or less. To simplify notation, for $1 \leq i \leq n$, let $X^i = \prod_{j \neq i} X_j$, $\mathcal{M}^i := \otimes_{j \neq i} \mathcal{M}_i$, and $\mu^i := \otimes_{j \neq i} \mu_j$ be the product measure on $(X^i, \mathcal{M}^i)$ which is assumed to exist by the induction hypothesis. Also let $\mathcal{M} := \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ and for $x = (x_1, \ldots, x_i, \ldots, x_n) \in X$ let

$$x^i := (x_1, \ldots, \hat{x}_i, \ldots, x_n) := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Here is an outline of the argument with some details being left to the reader.

1. If $f : X \to [0, \infty]$ is $\mathcal{M}$ -measurable, then
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ -measurable. Thus by the induction hypothesis, the right side of Eq. (11.17) is well defined.
2. If $\sigma \in S_n$ (the permutations of $\{1, 2, \ldots, n\}$) we may define a measure $\pi$ on $(X, \mathcal{M})$ by;
$$\pi(A) := \int_{X_{\sigma 1}} d\mu_{\sigma 1}(x_{\sigma 1}) \cdots \int_{X_{\sigma n}} d\mu_{\sigma n}(x_{\sigma n})\, 1_A(x_1, \ldots, x_n). \tag{11.19}$$
It is easy to check that $\pi$ is a measure which satisfies Eq. (11.16). Using the $\sigma$ – finiteness assumptions and the fact that
$$\mathcal{P} := \{A_1 \times \cdots \times A_n : A_i \in \mathcal{M}_i \text{ for } 1 \leq i \leq n\}$$
is a $\pi$ – system such that $\sigma(\mathcal{P}) = \mathcal{M}$, it follows from Exercise 6.1 that there is only one such measure satisfying Eq. (11.16). Thus the formula for $\pi$ in Eq. (11.19) is independent of $\sigma \in S_n$.
3. From Eq. (11.19) and the usual simple function approximation arguments we may conclude that Eq. (11.17) is valid.
Now suppose that $f \in L^1(X, \mathcal{M}, \pi)$.
4. Using step 1 it is easy to check that
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \bar{\int}_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable. Indeed,
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable and therefore
$$E := \left\{(x_1, \ldots, \hat{x}_i, \ldots, x_n) : \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i) < \infty\right\} \in \mathcal{M}^i.$$
Now let $u := \operatorname{Re} f$ and $v := \operatorname{Im} f$ and $u_\pm$ and $v_\pm$ are the positive and negative parts of $u$ and $v$ respectively, then
$$\begin{aligned}\bar{\int}_{X_i} f(x)\, d\mu_i(x_i) &= \int_{X_i} 1_E(x^i)\, f(x)\, d\mu_i(x_i) \\ &= \int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) + i \int_{X_i} 1_E(x^i)\, v(x)\, d\mu_i(x_i).\end{aligned}$$
Both of these later terms are $\mathcal{M}^i$ – measurable since, for example,
$$\int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) = \int_{X_i} 1_E(x^i)\, u_+(x)\, d\mu_i(x_i) - \int_{X_i} 1_E(x^i)\, u_-(x)\, d\mu_i(x_i)$$
which is $\mathcal{M}^i$ – measurable by step 1.

5. It now follows by induction that the right side of Eq. (11.18) is well defined.
6. Let $i := \sigma n$ and $T : X \to X_i \times X^i$ be the obvious identification;
$$T\left(x_i, (x_1, \dots, \hat{x}_i, \dots, x_n)\right) = (x_1, \dots, x_n).$$
One easily verifies $T$ is $\mathcal{M}/\mathcal{M}_i \otimes \mathcal{M}^i$ – measurable (use Corollary 8.19 repeatedly) and that $\pi \circ T^{-1} = \mu_i \otimes \mu^i$ (see Exercise 6.1).
7. Let $f \in L^1(\pi)$. Combining step 6. with the abstract change of variables Theorem (Exercise 9.11) implies
$$\int_X f d\pi = \int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right). \tag{11.20}$$
By Theorem 11.9, we also have
$$\begin{aligned}\int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right) &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f \circ T(x_i, x^i) \\ &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f(x_1, \dots, x_n).\end{aligned} \tag{11.21}$$
Then by the induction hypothesis,
$$\int_{X^i} d\mu^i\left(x_i\right) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) = \prod_{j \neq i} \bar{\int}_{X_j} d\mu_j\left(x_j\right) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) \tag{11.22}$$
where the ordering the integrals in the last product are inconsequential. Combining Eqs. (11.20) – (11.22) completes the proof.

■

**Convention:** We are now going to drop the bar above the integral sign with the understanding that $\int_X f d\mu = 0$ whenever $f : X \to \mathbb{C}$ is a measurable function such that $\int_X |f|\, d\mu = \infty$. However if $f$ is a non-negative function (i.e. $f : X \to [0, \infty]$) non-integrable function we will interpret $\int_X f d\mu$ to be infinite.

*Example 11.11.* In this example we will show
$$\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} dx = \pi/2. \tag{11.23}$$
To see this write $\frac{1}{x} = \int_0^\infty e^{-tx} dt$ and use Fubini-Tonelli to conclude that
$$\begin{aligned}\int_0^M \frac{\sin x}{x} dx &= \int_0^M \left[\int_0^\infty e^{-tx} \sin x\ dt\right] dx \\ &= \int_0^\infty \left[\int_0^M e^{-tx} \sin x\ dx\right] dt \\ &= \int_0^\infty \frac{1}{1+t^2}\left(1 - te^{-Mt} \sin M - e^{-Mt} \cos M\right) dt \\ &\to \int_0^\infty \frac{1}{1+t^2} dt = \frac{\pi}{2} \text{ as } M \to \infty,\end{aligned}$$
wherein we have used the dominated convergence theorem (for instance, take $g(t) := \frac{1}{1+t^2}\left(1 + te^{-t} + e^{-t}\right))$ to pass to the limit.

The next example is a refinement of this result.

*Example 11.12.* We have
$$\int_0^\infty \frac{\sin x}{x} e^{-\Lambda x} dx = \frac{1}{2}\pi - \arctan \Lambda \text{ for all } \Lambda > 0 \tag{11.24}$$
and for$\Lambda, M \in [0, \infty)$,
$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C \frac{e^{-M\Lambda}}{M} \tag{11.25}$$
where $C = \max_{x \ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (11.23) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,
$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1 dy\right| = |x|$$
so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x \neq 0$. Making use of the identity
$$\int_0^\infty e^{-tx} dt = 1/x$$
and Fubini's theorem,

$$\begin{aligned}\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x\, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\ &= \int_0^\infty dt \int_0^M dx \sin x\, e^{-(\Lambda+t)x} \\ &= \int_0^\infty \frac{1 - (\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2+1} dt \\ &= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\ &= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M,\Lambda) \qquad (11.26)\end{aligned}$$

where

$$\varepsilon(M,\Lambda) = \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt.$$

Since

$$\left|\frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1}\right| \le \frac{1+(\Lambda+t)}{(\Lambda+t)^2+1} \le C,$$

$$|\varepsilon(M,\Lambda)| \le \int_0^\infty e^{-M(\Lambda+t)} dt = C\frac{e^{-M\Lambda}}{M}.$$

This estimate along with Eq. (11.26) proves Eq. (11.25) from which Eq. (11.23) follows by taking $\Lambda \to \infty$ and Eq. (11.24) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

**Lemma 11.13.** *Suppose that $X$ is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is a $C^1$ – functions such that $\lim_{x\to-\infty} \varphi(x) = 0$ and either $\varphi'(x) \ge 0$ for all $x$ or $\int_{\mathbb{R}} |\varphi'(x)|\, dx < \infty$. Then*

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^\infty \varphi'(y)\, P(X > y)\, dy.$$

*Similarly if $X \ge 0$ and $\varphi : [0,\infty) \to \mathbb{R}$ is a $C^1$ – function such that $\varphi(0) = 0$ and either $\varphi' \ge 0$ or $\int_0^\infty |\varphi'(x)|\, dx < \infty$, then*

$$\mathbb{E}[\varphi(X)] = \int_0^\infty \varphi'(y)\, P(X > y)\, dy.$$

**Proof.** By the fundamental theorem of calculus for all $M < \infty$ and $x \in \mathbb{R}$,

$$\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y)\, dy. \qquad (11.27)$$

Under the stated assumptions on $\varphi$, we may use either the monotone or the dominated convergence theorem to let $M \to \infty$ in Eq. (11.27) to find,

$$\varphi(x) = \int_{-\infty}^x \varphi'(y)\, dy = \int_{\mathbb{R}} 1_{y<x} \varphi'(y)\, dy \text{ for all } x \in \mathbb{R}.$$

Therefore,

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left[\int_{\mathbb{R}} 1_{y<X}\varphi'(y)\, dy\right] = \int_{\mathbb{R}} \mathbb{E}[1_{y<X}]\, \varphi'(y)\, dy = \int_{-\infty}^\infty \varphi'(y)\, P(X > y)\, dy,$$

where we applied Fubini's theorem for the second equality. The proof of the second assertion is similar and will be left to the reader. ■

*Example 11.14.* Here are a couple of examples involving Lemma 11.13.

1. Suppose $X$ is a random variable, then

$$\mathbb{E}\left[e^X\right] = \int_{-\infty}^\infty P(X > y)\, e^y dy = \int_0^\infty P(X > \ln u)\, du, \qquad (11.28)$$

   where we made the change of variables, $u = e^y$, to get the second equality.
2. If $X \ge 0$ and $p \ge 1$, then

$$\mathbb{E}X^p = p\int_0^\infty y^{p-1} P(X > y)\, dy. \qquad (11.29)$$

## 11.4 Fubini's Theorem and Completions*

**Notation 11.15** *Given $E \subset X \times Y$ and $x \in X$, let*

$${}_xE := \{y \in Y : (x,y) \in E\}.$$

*Similarly if $y \in Y$ is given let*

$$E_y := \{x \in X : (x,y) \in E\}.$$

*If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.*

**Theorem 11.16.** *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are **complete** $\sigma$ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If $f$ is $\mathcal{L}$ – measurable and (a) $f \ge 0$ or (b) $f \in L^1(\lambda)$ then $f_x$ is $\mathcal{N}$ – measurable for $\mu$ a.e. $x$ and $f^y$ is $\mathcal{M}$ – measurable for $\nu$ a.e. $y$ and in case (b) $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$ for $\mu$ a.e. $x$ and $\nu$ a.e. $y$ respectively. Moreover,*

$$\left(x \to \int_Y f_x d\nu\right) \in L^1(\mu) \;\; and \;\; \left(y \to \int_X f^y d\mu\right) \in L^1(\nu)$$

*and*

$$\int_{X\times Y} f d\lambda = \int_Y d\nu \int_X d\mu\, f = \int_X d\mu \int_Y d\nu\, f.$$

**Proof.** If $E \in \mathcal{M}\otimes\mathcal{N}$ is a $\mu\otimes\nu$ null set (i.e. $(\mu\otimes\nu)(E) = 0$), then

$$0 = (\mu\otimes\nu)(E) = \int_X \nu({}_xE)d\mu(x) = \int_X \mu(E_y)d\nu(y).$$

This shows that

$$\mu(\{x : \nu({}_xE) \neq 0\}) = 0 \text{ and } \nu(\{y : \mu(E_y) \neq 0\}) = 0,$$

i.e. $\nu({}_xE) = 0$ for $\mu$ a.e. $x$ and $\mu(E_y) = 0$ for $\nu$ a.e. $y$. If $h$ is $\mathcal{L}$ measurable and $h = 0$ for $\lambda$ – a.e., then there exists $E \in \mathcal{M}\otimes\mathcal{N}$ such that $\{(x,y) : h(x,y) \neq 0\} \subset E$ and $(\mu\otimes\nu)(E) = 0$. Therefore $|h(x,y)| \le 1_E(x,y)$ and $(\mu\otimes\nu)(E) = 0$. Since

$$\{h_x \neq 0\} = \{y \in Y : h(x,y) \neq 0\} \subset {}_xE \text{ and}$$
$$\{h_y \neq 0\} = \{x \in X : h(x,y) \neq 0\} \subset E_y$$

we learn that for $\mu$ a.e. $x$ and $\nu$ a.e. $y$ that $\{h_x \neq 0\} \in \mathcal{M}$, $\{h_y \neq 0\} \in \mathcal{N}$, $\nu(\{h_x \neq 0\}) = 0$ and a.e. and $\mu(\{h_y \neq 0\}) = 0$. This implies $\int_Y h(x,y)d\nu(y)$ exists and equals 0 for $\mu$ a.e. $x$ and similarly that $\int_X h(x,y)d\mu(x)$ exists and equals 0 for $\nu$ a.e. $y$. Therefore

$$0 = \int_{X\times Y} h d\lambda = \int_Y \left(\int_X h d\mu\right) d\nu = \int_X \left(\int_Y h d\nu\right) d\mu.$$

For general $f \in L^1(\lambda)$, we may choose $g \in L^1(\mathcal{M}\otimes\mathcal{N}, \mu\otimes\nu)$ such that $f(x,y) = g(x,y)$ for $\lambda-$ a.e. $(x,y)$. Define $h := f - g$. Then $h = 0$, $\lambda-$ a.e. Hence by what we have just proved and Theorem 11.6 $f = g + h$ has the following properties:

1. For $\mu$ a.e. $x$, $y \to f(x,y) = g(x,y) + h(x,y)$ is in $L^1(\nu)$ and
$$\int_Y f(x,y)d\nu(y) = \int_Y g(x,y)d\nu(y).$$
2. For $\nu$ a.e. $y$, $x \to f(x,y) = g(x,y) + h(x,y)$ is in $L^1(\mu)$ and
$$\int_X f(x,y)d\mu(x) = \int_X g(x,y)d\mu(x).$$



From these assertions and Theorem 11.6, it follows that

$$\begin{aligned}\int_X d\mu(x)\int_Y d\nu(y) f(x,y) &= \int_X d\mu(x)\int_Y d\nu(y) g(x,y)\\ &= \int_Y d\nu(y)\int_Y d\nu(x) g(x,y)\\ &= \int_{X\times Y} g(x,y) d(\mu\otimes\nu)(x,y)\\ &= \int_{X\times Y} f(x,y) d\lambda(x,y).\end{aligned}$$

Similarly it is shown that

$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) = \int_{X\times Y} f(x,y)d\lambda(x,y).$$

■

## 11.5 Lebesgue Measure on $\mathbb{R}^d$ and the Change of Variables Theorem

[There are a number of proof in the Math 140 notes which might be better to uses here.]

**Notation 11.17** *Let*

$$m^d := \overbrace{m\otimes\cdots\otimes m}^{d \text{ times}} \text{ on } \mathcal{B}_{\mathbb{R}^d} = \overbrace{\mathcal{B}_{\mathbb{R}}\otimes\cdots\otimes\mathcal{B}_{\mathbb{R}}}^{d \text{ times}}$$

*be the $d$ – fold product of Lebesgue measure $m$ on $\mathcal{B}_{\mathbb{R}}$. We will also use $m^d$ to denote its completion and let $\mathcal{L}_d$ be the completion of $\mathcal{B}_{\mathbb{R}^d}$ relative to $m^d$. A subset $A \in \mathcal{L}_d$ is called a Lebesgue measurable set and $m^d$ is called $d$ – dimensional Lebesgue measure, or just Lebesgue measure for short.*

**Definition 11.18.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is* ***Lebesgue measurable*** *if $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{L}_d$.*

**Notation 11.19** *I will often be sloppy in the sequel and write $m$ for $m^d$ and $dx$ for $dm(x) = dm^d(x)$, i.e.*

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d} f dm^d.$$

*Hopefully the reader will understand the meaning from the context.*

**Theorem 11.20.** *Lebesgue measure $m^d$ is translation invariant. Moreover $m^d$ is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$.*

**Proof.** Let $A = J_1 \times \cdots \times J_d$ with $J_i \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}^d$. Then

$$x + A = (x_1 + J_1) \times (x_2 + J_2) \times \cdots \times (x_d + J_d)$$

and therefore by translation invariance of $m$ on $\mathcal{B}_{\mathbb{R}}$ we find that

$$m^d(x + A) = m(x_1 + J_1) \ldots m(x_d + J_d) = m(J_1) \ldots m(J_d) = m^d(A)$$

and hence $m^d(x + A) = m^d(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ since it holds for $A$ in a multiplicative system which generates $\mathcal{B}_{\mathbb{R}^d}$. From this fact we see that the measure $m^d(x + \cdot)$ and $m^d(\cdot)$ have the same null sets. Using this it is easily seen that $m(x + A) = m(A)$ for all $A \in \mathcal{L}_d$. The proof of the second assertion is Exercise 11.14. ∎

**Exercise 11.1.** In this problem you are asked to show there is no reasonable notion of Lebesgue measure on an infinite dimensional Hilbert space. To be more precise, suppose $H$ is an infinite dimensional Hilbert space and $m$ is a **countably additive** measure on $\mathcal{B}_H$ which is invariant under translations and satisfies, $m(B_0(\varepsilon)) > 0$ for all $\varepsilon > 0$. Show $m(V) = \infty$ for all non-empty open subsets $V \subset H$.

**Theorem 11.21 (Change of Variables Theorem).** *Let $\Omega \subset_o \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ be a $C^1$ – diffeomorphism,[2] see Figure 11.1. Then for any Borel measurable function, $f : T(\Omega) \to [0, \infty]$,*

$$\int_\Omega f(T(x)) |\det T'(x)| dx = \int_{T(\Omega)} f(y)\, dy, \tag{11.30}$$

*where $T'(x)$ is the linear transformation on $\mathbb{R}^d$ defined by $T'(x)v := \frac{d}{dt}|_0 T(x + tv)$. More explicitly, viewing vectors in $\mathbb{R}^d$ as columns, $T'(x)$ may be represented by the matrix*

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \ldots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \ldots & \partial_d T_d(x) \end{bmatrix}, \tag{11.31}$$

*i.e. the $i$ - $j$ – matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \ldots, T_d(x))^{\mathrm{tr}}$ and $\partial_i = \partial/\partial x_i$.*

[2] That is $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ is a continuously differentiable bijection and the inverse map $T^{-1} : T(\Omega) \to \Omega$ is also continuously differentiable.

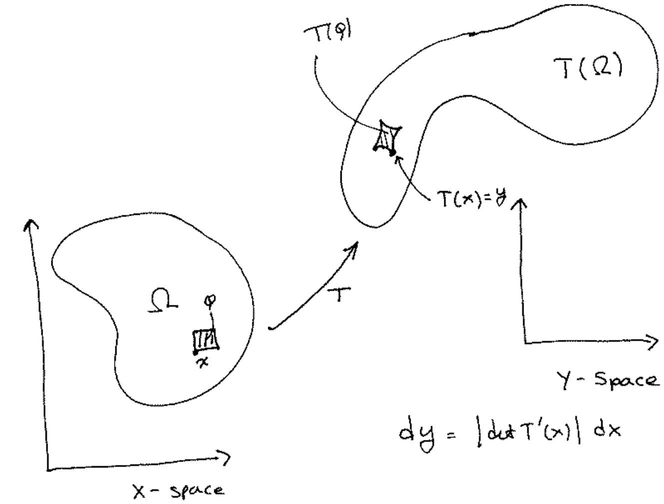



**Fig. 11.1.** The geometric setup of Theorem 11.21.

*Remark 11.22.* Theorem 11.21 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = |\det T'(x)| dx$. As usual, you must also change the limits of integration appropriately, i.e. if $x$ ranges through $\Omega$ then $y$ must range through $T(\Omega)$.

## Note: you may skip the rest of this section!

**Proof.** The proof will be by induction on $d$. The case $d = 1$ was essentially done in Exercise 9.12. Nevertheless, for the sake of completeness let us give a proof here. Suppose $d = 1$, $a < \alpha < \beta < b$ such that $[a, b]$ is a compact subinterval of $\Omega$. Then $|\det T'| = |T'|$ and

$$\int_{[a,b]} 1_{T((\alpha,\beta])}(T(x)) |T'(x)| dx = \int_{[a,b]} 1_{(\alpha,\beta]}(x) |T'(x)| dx = \int_\alpha^\beta |T'(x)| dx.$$

If $T'(x) > 0$ on $[a, b]$, then

$$\begin{aligned} \int_\alpha^\beta |T'(x)| dx &= \int_\alpha^\beta T'(x)\, dx = T(\beta) - T(\alpha) \\ &= m(T((\alpha, \beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y)\, dy \end{aligned}$$

while if $T'(x) < 0$ on $[a, b]$, then

$$\int_\alpha^\beta |T'(x)|\, dx = -\int_\alpha^\beta T'(x)\, dx = T(\alpha) - T(\beta)$$
$$= m(T((\alpha, \beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y)\, dy.$$

Combining the previous three equations shows

$$\int_{[a,b]} f(T(x))\, |T'(x)|\, dx = \int_{T([a,b])} f(y)\, dy \tag{11.32}$$

whenever $f$ is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 10.20 then implies that Eq. (11.32) holds for every bounded measurable function $f : T([a,b]) \to \mathbb{R}$. (Observe that $|T'(x)|$ is continuous and hence bounded for $x$ in the compact interval, $[a,b]$.) Recall that $\Omega = \sum_{n=1}^N (a_n, b_n)$ where $a_n, b_n \in \mathbb{R}\cup\{\pm\infty\}$ for $n = 1, 2, \dots < N$ with $N = \infty$ possible. Hence if $f : T(\Omega) \to \mathbb{R}_+$ is a Borel measurable function and $a_n < \alpha_k < \beta_k < b_n$ with $\alpha_k \downarrow a_n$ and $\beta_k \uparrow b_n$, then by what we have already proved and the monotone convergence theorem

$$\int_\Omega 1_{(a_n,b_n)} \cdot (f \circ T) \cdot |T'| dm = \int_\Omega \left(1_{T((a_n,b_n))} \cdot f\right) \circ T \cdot |T'| dm$$
$$= \lim_{k\to\infty} \int_\Omega \left(1_{T([\alpha_k,\beta_k])} \cdot f\right) \circ T \cdot |T'|\, dm$$
$$= \lim_{k\to\infty} \int_{T(\Omega)} 1_{T([\alpha_k,\beta_k])} \cdot f\, dm$$
$$= \int_{T(\Omega)} 1_{T((a_n,b_n))} \cdot f\, dm.$$

Summing this equality on $n$, then shows Eq. (11.30) holds.

To carry out the induction step, we now suppose $d > 1$ and suppose the theorem is valid with $d$ being replaced by $d-1$. For notational compactness, let us write vectors in $\mathbb{R}^d$ as row vectors rather than column vectors. Nevertheless, the matrix associated to the differential, $T'(x)$, will always be taken to be given as in Eq. (11.31).

**Case 1.** Suppose $T(x)$ has the form

$$T(x) = (x_i, T_2(x), \dots, T_d(x)) \tag{11.33}$$

or

$$T(x) = (T_1(x), \dots, T_{d-1}(x), x_i) \tag{11.34}$$



for some $i \in \{1, \dots, d\}$. For definiteness we will assume $T$ is as in Eq. (11.33), the case of $T$ in Eq. (11.34) may be handled similarly. For $t \in \mathbb{R}$, let $i_t : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the inclusion map defined by

$$i_t(w) := w_t := (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}),$$

$\Omega_t$ be the (possibly empty) open subset of $\mathbb{R}^{d-1}$ defined by

$$\Omega_t := \left\{w \in \mathbb{R}^{d-1} : (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}) \in \Omega\right\}$$

and $T_t : \Omega_t \to \mathbb{R}^{d-1}$ be defined by

$$T_t(w) = (T_2(w_t), \dots, T_d(w_t)),$$

see Figure 11.2. Expanding $\det T'(w_t)$ along the first row of the matrix $T'(w_t)$

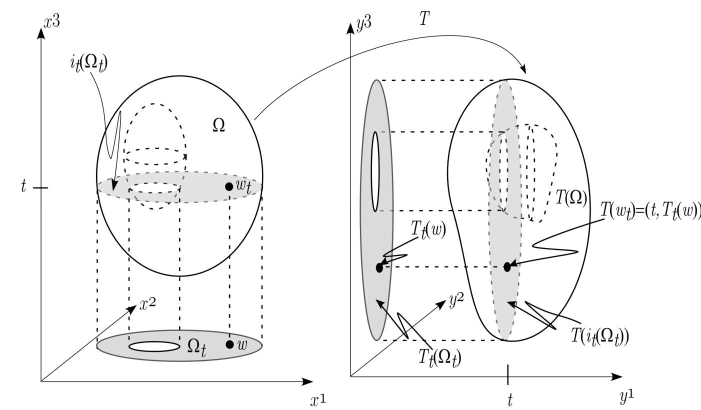



**Fig. 11.2.** In this picture $d = i = 3$ and $\Omega$ is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map $T$ and slicing the set $\Omega$ along planes where $x_3 = t$.

shows

$$|\det T'(w_t)| = |\det T_t'(w)|.$$

Now by the Fubini-Tonelli Theorem and the induction hypothesis,

$$\begin{aligned}\int_{\Omega} f\circ T|\det T'|dm &= \int_{\mathbb{R}^d} 1_\Omega \cdot f\circ T|\det T'|dm \\ &= \int_{\mathbb{R}^d} 1_\Omega (w_t)(f\circ T)(w_t)|\det T'(w_t)|dwdt \\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t}(f\circ T)(w_t)|\det T'(w_t)|dw\right]dt \\ &= \int_{\mathbb{R}}\left[\int_{\Omega_t} f(t,T_t(w))|\det T_t'(w)|dw\right]dt \\ &= \int_{\mathbb{R}}\left[\int_{T_t(\Omega_t)} f(t,z)\,dz\right]dt = \int_{\mathbb{R}}\left[\int_{\mathbb{R}^{d-1}} 1_{T(\Omega)}(t,z)\,f(t,z)\,dz\right]dt \\ &= \int_{T(\Omega)} f(y)\,dy\end{aligned}$$

wherein the last two equalities we have used Fubini-Tonelli along with the identity;

$$T(\Omega) = \sum_{t\in\mathbb{R}} T(i_t(\Omega)) = \sum_{t\in\mathbb{R}} \{(t,z) : z\in T_t(\Omega_t)\}.$$

**Case 2.** (Eq. (11.30) is true locally.) Suppose that $T:\Omega\to\mathbb{R}^d$ is a general map as in the statement of the theorem and $x_0\in\Omega$ is an arbitrary point. We will now show there exists an open neighborhood $W\subset\Omega$ of $x_0$ such that

$$\int_W f\circ T|\det T'|dm = \int_{T(W)} fdm$$

holds for all Borel measurable function, $f:T(W)\to[0,\infty]$. Let $M_i$ be the 1-$i$ minor of $T'(x_0)$, i.e. the determinant of $T'(x_0)$ with the first row and $i^{\text{th}}$ – column removed. Since

$$0\neq \det T'(x_0) = \sum_{i=1}^d (-1)^{i+1}\partial_i T_j(x_0)\cdot M_i,$$

there must be some $i$ such that $M_i\neq 0$. Fix an $i$ such that $M_i\neq 0$ and let,

$$S(x) := (x_i, T_2(x),\dots,T_d(x)). \tag{11.35}$$

Observe that $|\det S'(x_0)| = |M_i| \neq 0$. Hence by the inverse function Theorem, there exist an open neighborhood $W$ of $x_0$ such that $W\subset_o\Omega$ and $S(W)\subset_o\mathbb{R}^d$ and $S:W\to S(W)$ is a $C^1$ – diffeomorphism. Let $R:S(W)\to T(W)\subset_o\mathbb{R}^d$ to be the $C^1$ – diffeomorphism defined by

$$R(z) := T\circ S^{-1}(z) \text{ for all } z\in S(W).$$

Because

$$(T_1(x),\dots,T_d(x)) = T(x) = R(S(x)) = R((x_i,T_2(x),\dots,T_d(x)))$$

for all $x\in W$, if

$$(z_1,z_2,\dots,z_d) = S(x) = (x_i,T_2(x),\dots,T_d(x))$$

then

$$R(z) = \left(T_1\left(S^{-1}(z)\right), z_2,\dots,z_d\right). \tag{11.36}$$

Observe that $S$ is a map of the form in Eq. (11.33), $R$ is a map of the form in Eq. (11.34), $T'(x) = R'(S(x))S'(x)$ (by the chain rule) and (by the multiplicative property of the determinant)

$$|\det T'(x)| = |\det R'(S(x))|\,|\det S'(x)| \ \forall\, x\in W.$$

So if $f:T(W)\to[0,\infty]$ is a Borel measurable function, two applications of the results in Case 1. shows,

$$\begin{aligned}\int_W f\circ T\cdot|\det T'|dm &= \int_W (f\circ R\cdot|\det R'|)\circ S\cdot|\det S'|\ dm \\ &= \int_{S(W)} f\circ R\cdot|\det R'|dm = \int_{R(S(W))} fdm \\ &= \int_{T(W)} fdm\end{aligned}$$

and Case 2. is proved.

**Case 3.** (General Case.) Let $f:\Omega\to[0,\infty]$ be a general non-negative Borel measurable function and let

$$K_n := \{x\in\Omega : \operatorname{dist}(x,\Omega^c)\geq 1/n \text{ and } |x|\leq n\}.$$

Then each $K_n$ is a compact subset of $\Omega$ and $K_n\uparrow\Omega$ as $n\to\infty$. Using the compactness of $K_n$ and case 2, for each $n\in\mathbb{N}$, there is a finite open cover $\mathcal{W}_n$ of $K_n$ such that $W\subset\Omega$ and Eq. (11.30) holds with $\Omega$ replaced by $W$ for each $W\in\mathcal{W}_n$. Let $\{W_i\}_{i=1}^\infty$ be an enumeration of $\cup_{n=1}^\infty\mathcal{W}_n$ and set $\tilde{W}_1 = W_1$ and $\tilde{W}_i := W_i\setminus(W_1\cup\dots\cup W_{i-1})$ for all $i\geq 2$. Then $\Omega = \sum_{i=1}^\infty \tilde{W}_i$ and by repeated use of case 2.,

$$\int_\Omega f\circ T|\det T'|dm = \sum_{i=1}^\infty \int_\Omega 1_{\tilde W_i}\cdot (f\circ T)\cdot |\det T'|dm$$
$$= \sum_{i=1}^\infty \int_{W_i} \left[\left(1_{T(\tilde W_i)} f\right)\circ T\right]\cdot |\det T'|dm$$
$$= \sum_{i=1}^\infty \int_{T(W_i)} 1_{T(\tilde W_i)}\cdot f\ dm = \sum_{i=1}^n \int_{T(\Omega)} 1_{T(\tilde W_i)}\cdot f\ dm$$
$$= \int_{T(\Omega)} f dm.$$

■

*Remark 11.23.* When $d=1$, one often learns the change of variables formula as

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_{T(a)}^{T(b)} f(y)\,dy \tag{11.37}$$

where $f:[a,b]\to\mathbb{R}$ is a continuous function and $T$ is $C^1$ – function defined in a neighborhood of $[a,b]$. If $T'>0$ on $(a,b)$ then $T((a,b)) = (T(a),T(b))$ and Eq. (11.37) is implies Eq. (11.30) with $\Omega=(a,b)$. On the other hand if $T'<0$ on $(a,b)$ then $T((a,b)) = (T(b),T(a))$ and Eq. (11.37) is equivalent to

$$\int_{(a,b)} f(T(x))\,(-|T'(x)|)\,dx = -\int_{T(b)}^{T(a)} f(y)\,dy = -\int_{T((a,b))} f(y)\,dy$$

which is again implies Eq. (11.30). On the other hand Eq. (11.37) is more general than Eq. (11.30) since it does not require $T$ to be injective. The standard proof of Eq. (11.37) is as follows. For $z\in T([a,b])$, let

$$F(z) := \int_{T(a)}^z f(y)\,dy.$$

Then by the chain rule and the fundamental theorem of calculus,

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_a^b F'(T(x))\,T'(x)\,dx = \int_a^b \frac{d}{dx}\left[F(T(x))\right]dx$$
$$= F(T(x))\,|_a^b = \int_{T(a)}^{T(b)} f(y)\,dy.$$

An application of Dynkin's multiplicative systems theorem now shows that Eq. (11.37) holds for all bounded measurable functions $f$ on $(a,b)$. Then by the usual truncation argument, it also holds for all positive measurable functions on $(a,b)$.



**Exercise 11.2.** Continuing the setup in Theorem 11.21, show that $f\in L^1\left(T(\Omega),m^d\right)$ iff

$$\int_\Omega |f\circ T|\,|\det T'|dm < \infty$$

and if $f\in L^1\left(T(\Omega),m^d\right)$, then Eq. (11.30) holds.

*Example 11.24.* Continuing the setup in Theorem 11.21, if $A\in\mathcal{B}_\Omega$, then

$$m(T(A)) = \int_{\mathbb{R}^d} 1_{T(A)}(y)\,dy = \int_{\mathbb{R}^d} 1_{T(A)}(Tx)\,|\det T'(x)|\,dx$$
$$= \int_{\mathbb{R}^d} 1_A(x)\,|\det T'(x)|\,dx$$

wherein the second equality we have made the change of variables, $y=T(x)$. Hence we have shown

$$d(m\circ T) = |\det T'(\cdot)|\,dm.$$

Taking $T\in GL(d,\mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d\times d$ invertible matrices in the previous example implies $m\circ T = |\det T|\,m$, i.e.

$$m(T(A)) = |\det T|\,m(A) \text{ for all } A\in\mathcal{B}_{\mathbb{R}^d}. \tag{11.38}$$

This equation also shows that $m\circ T$ and $m$ have the same null sets and hence the equality in Eq. (11.38) is valid for any $A\in\mathcal{L}_d$. In particular we may conclude that $m$ is invariant under those $T\in GL(d,\mathbb{R})$ with $|\det(T)|=1$. For example if $T$ is a rotation (i.e. $T^{\mathrm{tr}}T=I$), then $\det T=\pm1$ and hence $m$ is invariant under all rotations. This is not obvious from the definition of $m^d$ as a product measure!

*Example 11.25.* Suppose that $T(x) = x+b$ for some $b\in\mathbb{R}^d$. In this case $T'(x)=I$ and therefore it follows that

$$\int_{\mathbb{R}^d} f(x+b)\,dx = \int_{\mathbb{R}^d} f(y)\,dy$$

for all measurable $f:\mathbb{R}^d\to[0,\infty]$ or for any $f\in L^1(m)$. In particular Lebesgue measure is invariant under translations.

*Example 11.26 (Polar Coordinates).* Suppose $T:(0,\infty)\times(0,2\pi)\to\mathbb{R}^2$ is defined by

$$x = T(r,\theta) = (r\cos\theta, r\sin\theta),$$

i.e. we are making the change of variable,

$$x_1 = r\cos\theta \text{ and } x_2 = r\sin\theta \text{ for } 0 < r < \infty \text{ and } 0 < \theta < 2\pi.$$

In this case

$$T'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore

$$dx = |\det T'(r,\theta)|\, dr d\theta = r dr d\theta.$$

Observing that

$$\mathbb{R}^2 \setminus T\left((0,\infty)\times(0,2\pi)\right) = \ell := \{(x,0) : x \geq 0\}$$

has $m^2$ – measure zero, it follows from the change of variables Theorem 11.21 that

$$\int_{\mathbb{R}^2} f(x)dx = \int_0^{2\pi} d\theta \int_0^\infty dr\ r\cdot f(r(\cos\theta,\sin\theta)) \tag{11.39}$$

for any Borel measurable function $f : \mathbb{R}^2 \to [0,\infty]$.

*Example 11.27 (Holomorphic Change of Variables).* Suppose that $f : \Omega \subset_o \mathbb{C} \cong \mathbb{R}^2 \to \mathbb{C}$ is an injective holomorphic function such that $f'(z) \neq 0$ for all $z \in \Omega$. We may express $f$ as

$$f(x+iy) = U(x,y) + iV(x,y)$$

for all $z = x+iy \in \Omega$. Hence if we make the change of variables,

$$w = u + iv = f(x+iy) = U(x,y) + iV(x,y)$$

then

$$dudv = \left|\det\begin{bmatrix} U_x & U_y \\ V_x & V_y \end{bmatrix}\right| dxdy = |U_x V_y - U_y V_x|\, dxdy.$$

Recalling that $U$ and $V$ satisfy the Cauchy Riemann equations, $U_x = V_y$ and $U_y = -V_x$ with $f' = U_x + iV_x$, we learn

$$U_x V_y - U_y V_x = U_x^2 + V_x^2 = |f'|^2.$$

Therefore

$$dudv = |f'(x+iy)|^2\, dxdy.$$

*Example 11.28.* In this example we will evaluate the integral

$$I := \iint_\Omega \left(x^4 - y^4\right) dxdy$$

where

$$\Omega = \left\{(x,y) : 1 < x^2 - y^2 < 2,\ 0 < xy < 1\right\},$$

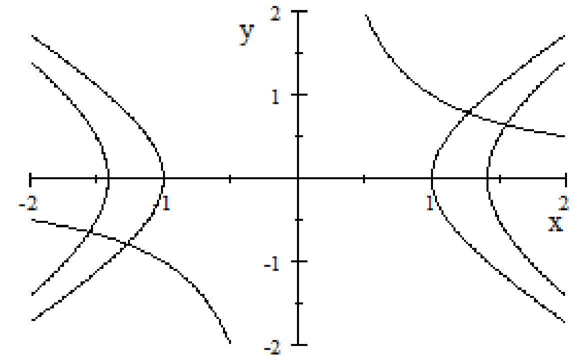



**Fig. 11.3.** The region $\Omega$ consists of the two curved rectangular regions shown.

see Figure 11.3. We are going to do this by making the change of variables,

$$(u,v) := T(x,y) = \left(x^2 - y^2, xy\right),$$

in which case

$$dudv = \left|\det\begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix}\right| dxdy = 2\left(x^2+y^2\right) dxdy$$

Notice that

$$\left(x^4 - y^4\right) = \left(x^2 - y^2\right)\left(x^2+y^2\right) = u\left(x^2+y^2\right) = \frac{1}{2}ududv.$$

The function $T$ is not injective on $\Omega$ but it is injective on each of its connected components. Let $D$ be the connected component in the first quadrant so that $\Omega = -D \cup D$ and $T(\pm D) = (1,2)\times(0,1)$. The change of variables theorem then implies

$$I_\pm := \iint_{\pm D} \left(x^4 - y^4\right) dxdy = \frac{1}{2}\iint_{(1,2)\times(0,1)} ududv = \frac{1}{2}\frac{u^2}{2}|_1^2 \cdot 1 = \frac{3}{4}$$

and therefore $I = I_+ + I_- = 2\cdot(3/4) = 3/2$.

**Exercise 11.3 (Spherical Coordinates).** Let $T : (0,\infty)\times(0,\pi)\times(0,2\pi) \to \mathbb{R}^3$ be defined by

$$\begin{aligned} T(r,\varphi,\theta) &= (r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi) \\ &= r(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi), \end{aligned}$$

see Figure 11.4. By making the change of variables $x = T(r,\varphi,\theta)$, show

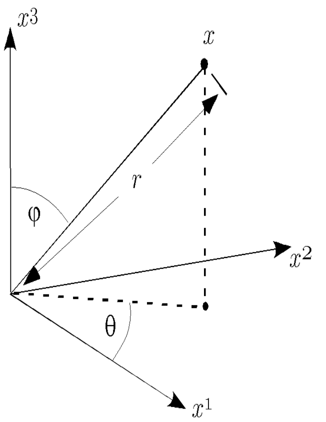



**Fig. 11.4.** The relation of $x$ to $(r,\phi,\theta)$ in spherical coordinates.

$$\int_{\mathbb{R}^3} f(x)dx = \int_0^\pi d\varphi \int_0^{2\pi} d\theta \int_0^\infty dr\ r^2 \sin\varphi \cdot f(T(r,\varphi,\theta))$$

for any Borel measurable function, $f:\mathbb{R}^3 \to [0,\infty]$.

**Lemma 11.29.** *Let $a>0$ and*

$$I_d(a) := \int_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

*Then $I_d(a) = (\pi/a)^{d/2}$.*

**Proof.** By Tonelli's theorem and induction,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \end{aligned} \tag{11.40}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)} dx_1 dx_2.$$

Using polar coordinates, see Eq. (11.39), we find,

$$\begin{aligned} I_2(a) &= \int_0^\infty dr\ r \int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi \int_0^\infty r e^{-ar^2} dr \\ &= 2\pi \lim_{M\to\infty} \int_0^M r e^{-ar^2} dr = 2\pi \lim_{M\to\infty} \frac{e^{-ar^2}}{-2a}\int_0^M = \frac{2\pi}{2a} = \pi/a. \end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (11.40). ■



## 11.6 Other change of variables proofs.

The proofs given below both rely on knowing the change of variables theorem for linear transformations summarized succinctly in the next theorem.

**Lemma 11.30.** *If $T$ is a real $d\times d$ matrix, then there exists $D = \mathrm{diag}(\lambda_1,\dots,\lambda_d)$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d \ge 0$ and two orthogonal matrices $R$ and $S$ such that $T = RDS^{\mathrm{tr}}$.*

**Proof.** Since $T^*T$ is symmetric, by the spectral theorem there exists an orthonormal basis $\{u_j\}_{j=1}^d$ of $\mathbb{R}^d$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d \ge 0$ such that $T^*Tu_j = \lambda_j^2 u_j$. Let $1 \le k \le d$ be the first index such that $\lambda_{k+1} = 0$ wherein by convention $\lambda_{d+1} = 0$. Then $\left\{v_j := \frac{1}{\lambda_j} T u_j\right\}_{j=1}^k$ is an orthonormal subset of $\mathbb{R}^d$ and $Tu_j = 0$ for all $j > d$. We now extend (if necessary, i.e. if $k < d$) $\{v_j\}_{j=1}^k$ arbitrarily to an orthonormal basis for $\mathbb{R}^d$. Now let $D = \mathrm{diag}(\lambda_1,\dots,\lambda_d)$ and let $R$ and $S$ be the orthogonal matrices defined by

$$Re_j = v_j \text{ and } Se_j = u_j \text{ for all } 1 \le j \le d.$$

We then have

$$TSe_j = Tu_j = \lambda_j v_j = \lambda_j Re_j = R\lambda_j e_j = RDe_j \text{ for all } 1 \le j \le d$$

and hence $TS = RD$, i.e. $T = RDS^{\mathrm{tr}}$. ■

**Theorem 11.31.** *If $T$ is a real $d\times d$ matrix, then $m\circ T = |\det T|\, m$.*

**Proof.** The first thing to note is that $T_*m$ is still a translation invariant measure which is finite on bounded sets and therefore $T_*m = km$ for some $k \ge 0$. So our only goal is to identify the constant $k$. We first consider two special cases.

1. If $T = R$ is orthogonal and $B$ is the unit ball in $\mathbb{R}^d$, then $km(B) = m(RB) = m(B)$ from which it follows $k = 1 = |\det R|$.
2. If $T = D = \mathrm{diag}(\lambda_1,\dots,\lambda_d)$ with $\lambda_i \ge 0$, then $D[0,1]^d = [0,\lambda_1]\times\cdots\times[0,\lambda_d]$ so that
$$k = km\left([0,1]^d\right) = m\left(D[0,1]^d\right) = \lambda_1\dots\lambda_d = \det D.$$
3. For the general case we use Lemma 11.30 to write $T = RDS^{\mathrm{tr}}$ and then find,
$$mT = mRDS^{\mathrm{tr}} = mDS^{\mathrm{tr}} = \det(D)\, mS^{\mathrm{tr}} = \det(D)\, m = |\det T|\, m.$$

■

**Theorem 11.32 (Linear Change of Variables Theorem).** *If $T \in GL(d,\mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d \times d$ invertible matrices, then the change of variables formula,*

$$\int_{\mathbb{R}^d} f(y)dy = \int_{\mathbb{R}^d} |\det T|\, f \circ T(x)dx, \tag{11.41}$$

*holds for all Riemann integrable functions $f : \mathbb{R}^d \to \mathbb{R}$.*

**Proof.** As usual it suffices to show Eq. (11.41) holds when $f = 1_A$ for some $A \in \mathcal{B}_{\mathbb{R}^d}$. In this cse,

$$\begin{aligned}\int_{\mathbb{R}^d} f \circ T(x)dx &= \int_{\mathbb{R}^d} 1_A \circ T(x)dx = \int_{\mathbb{R}^d} 1_{T^{-1}(A)}(x)\,dx \\ &= m\left(T^{-1}(A)\right) = \left|\det T^{-1}\right| m(A) \\ &= \frac{1}{|\det T|}\int_{\mathbb{R}^d} 1_A(x)\,dx = \frac{1}{|\det T|}\int_{\mathbb{R}^d} f(x)\,dx.\end{aligned}$$

■

The next exercise gives an alternative proof of Theorem 11.31.

**Exercise 11.4 (Change of variables for elementary matrices).** Let $R := (a,b] = (a_1,b_1] \times \cdots \times (a_d,b_d] \subset \mathbb{R}^d$ be a bounded half open rectangle. Show by direct calculation that;

$$|\det T| \int_{\mathbb{R}^d} 1_R \circ T(x)\,dx = m(R) = \int_{\mathbb{R}^d} 1_R(y)\,dy \tag{11.42}$$

for each of the following linear transformations;

1. Suppose that $i < k$ and
$$T(x_1, x_2 \dots, x_d) = (x_1, \dots, x_{i-1}, x_k, x_{i+1} \dots, x_{k-1}, x_i, x_{k+1}, \dots x_d),$$
i.e. $T$ swaps the $i$ and $k$ coordinates of $x$.
2. $T(x_1, \dots x_k, \dots, x_d) = (x_1, \dots, cx_k, \dots x_d)$ where $c \in \mathbb{R} \setminus \{0\}$.
3. $T(x_1, x_2 \dots, x_d) = (x_1, \dots, \overset{\text{i'th spot}}{x_i + cx_k}, \dots x_k, \dots x_d)$ where $c \in \mathbb{R}$.

**Hint:** you should use Fubini's theorem along with the one dimensional change of variables theorem.

### 11.6.1 $\delta$ – function localization proof of the change of variables Theorem 11.21

The proof we give here comes from [31][3] who attributes the idea to Cornea. Recall that we are trying to prove Eq. (11.30) which states,

$$\int_\Omega f(T(x))\,|\det T'(x)|dx = \int_{T(\Omega)} f(y)\,dy.$$

The key heuristic[4] idea of this proof is to first verify that the formula holds when $f(y) = \delta(y_0 - y)$ for all $y_0 \in T(\Omega)$, i.e. we wish to show (with $x_0 = T^{-1}(y_0)$) that

$$1 = \int_{T(\Omega)} \delta(y_0 - y)\,dy = \int_\Omega \delta(y_0 - T(x))\,|\det T'(x)|dx. \tag{11.43}$$

If we can verify Eq. (11.43), then multiplying this equation by $f(y_0)$ and integrating the result ,would give,

$$\begin{aligned}\int_{T(\Omega)} f(y_0)\,dy_0 &= \int_{T(\Omega)} dy_0 f(y_0) \int_\Omega dx \delta(y_0 - T(x))\,|\det T'(x)| \\ &= \int_\Omega dx \int_{T(\Omega)} dy_0 f(y_0)\,\delta(y_0 - T(x))\,|\det T'(x)| \\ &= \int_\Omega dx\ f(T(x))\,|\det T'(x)|\end{aligned}$$

which is the desired change of variables formula. So let us explain why Eq. (11.43) should be true.

As $y_0 - T(x) = 0$ iff $x = x_0 := T^{-1}(y_0)$, we should have

$$\begin{aligned}\int_\Omega \delta(y_0 - T(x))\,|\det T'(x)|dx &= \int_\Omega \delta(y_0 - T(x))\,|\det T'(x_0)|dx \\ &= |\det T'(x_0)| \cdot \int_\Omega \delta(y_0 - T(x))\,dx\end{aligned}$$

and so Eq. (11.43) is equivalent to showing

$$\int_\Omega \delta(y_0 - T(x))\,dx = \frac{1}{|\det T'(x_0)|}\ \forall\ x_0 \in \Omega. \tag{11.44}$$

[3] There are some mistakes in the arguments given in this reference which we have taken the opportunity to correct in the exposition below.

[4] If you are uncomfortable with the heuristic discussion to follows you may skip it and jump directly to Proposition 11.33.

The "reasons" to expect Eq. (11.44) is correct are; 1)

$$T(x) \cong T(x_0) + T'(x_0)(x - x_0) = y_0 + T'(x_0)(x - x_0)$$

for $x$ near $x_0$, 2) $\delta(y_0 - T(x))$ is supported in an "infinitesimal" neighborhood of $x_0$ and so we expect $\delta(y_0 - T(x)) = \delta(T'(x_0)(x - x_0))$ (not just approximately equal), and hence 3) making use of the change of variables theorem for linear transformations we should have

$$\begin{aligned}\int_\Omega \delta(y_0 - T(x))\,dx &= \int_\Omega \delta(T'(x_0)(x - x_0))\,dx = \int_{\mathbb{R}^d} \delta(T'(x_0)w)\,dw \\ &= \frac{1}{|\det[T'(x_0)]|}\int_{\mathbb{R}^d}\delta(z)\,dz = \frac{1}{|\det[T'(x_0)]|}.\end{aligned}$$

The above "proof" outline is of course not rigorous since there is no **honest** function satisfying the properties of a $\delta$ – function. To remedy this deficiency we are going to replace $\delta$ by an approximate $\delta$ – sequence $\{\delta_r\}_{r>0}$. So as usual we let $\delta_1 \in C_c^\infty(\mathbb{R}^d, [0,\infty))$ with $\operatorname{supp}(\delta_1) \subset B$ – the unit ball in $\mathbb{R}^d$ and $\int_{\mathbb{R}^d}\delta_1(x)\,dx = 1$ and then define, $\delta_r(z) = r^{-d}\delta_1(r^{-1}z)$ for all $r > 0$. The next key proposition is a rigorous version of Eq. (11.44).

**Proposition 11.33.** *If for each $r > 0$, $J_r : T(\Omega) \to [0,\infty]$ be defined by*

$$J_r(y) := \int_\Omega \delta_r(T(x) - y)\,dx, \tag{11.45}$$

*then $\lim_{r\downarrow 0} J_r(y) = 1/\left|\det T'(T^{-1}(y))\right|$ locally uniformly in $y \in \Omega$. In more detail we are claiming to any compact subset, $K \subset T(\Omega)$,*

$$\lim_{r\downarrow 0}\sup_{y\in K}\left|J_r(y) - \frac{1}{|\det T'(T^{-1}(y))|}\right| = 0. \tag{11.46}$$

We will give the proof of this result a little later in this section after Lemma 11.34. First let us show how to use Propositions 11.33 to prove Theorem 11.21.

**Proof of Theorem 11.21..** Let $f \in C_c(T(\Omega), [0,\infty))$ and $K := \operatorname{supp}(f)$. Then by Tonelli's theorem,

$$\int_{T(\Omega)} f(y) J_r(y)\,dy = \int_\Omega dx \int_{T(\Omega)} dy \delta_r(T(x) - y) f(y) = \int_\Omega f * \delta_r(T(x))\;dx. \tag{11.47}$$

By Propositions 11.33 and dominated convergence theorem,

$$\lim_{r\downarrow 0}\int_{T(\Omega)} f(y) J_r(y)\,dy = \int_{T(\Omega)} f(y)\frac{1}{|\det T'(T^{-1}(y))|}dy. \tag{11.48}$$

By simple approximate $\delta$ – function arguments we know $f * \delta_r \to f$ uniformly on $\Omega$ and moreover $\operatorname{supp}(f * \delta_r) \subset K_r$ where $K_r := \left\{x \in \mathbb{R}^d : d_K(x) \le r\right\}$. Hence if we let

$$\rho = \min\left(\frac{1}{2}dist\left(K, T(\Omega)^c\right), 1\right), \tag{11.49}$$

then for all $0 < r \le \rho$, $\operatorname{supp}((f * \delta_r) \circ T) \subset T^{-1}(K_\rho)$ (a compact subset of $\Omega$) and $(f * \delta_r) \circ T \to f \circ T$ uniformly as $r \downarrow 0$. Thus again by Dominated convergence theorem we conclude

$$\lim_{r\downarrow 0}\int_\Omega f * \delta_r(T(x))\;dx = \int_\Omega f(T(x))\,dx. \tag{11.50}$$

Combining Eqs. (11.47), (11.48), and (11.50) implies

$$\int_\Omega f(T(x))\,dx = \int_{T(\Omega)} f(y)\frac{1}{|\det T'(T^{-1}(y))|}dy$$

from which Theorem 11.21 follows upon replacing $f(y)$ by $f(y) \cdot \left|\det T'(T^{-1}(y))\right|$. ■

The next lemma spells out certain mapping and continuity properties of $T^{-1}$, $T$, and $T'$ which will be used in the proof of Proposition 11.33.

**Lemma 11.34.** *Let $K$ be a compact subset of $T(\Omega)$, $0 < \rho \le 1$ be as in Eq. (11.49), $K_\rho := \left\{x \in \mathbb{R}^d : d_K(x) \le \rho\right\}$,[5] and*

$$M := \sup_{y\in K_\rho}\left\|\left(T^{-1}\right)'(y)\right\|_{op} < \infty. \tag{11.51}$$

*Then*

$$\sup_{y\in K}\left\|T^{-1}(y+w) - T^{-1}(y)\right\| \le M\|w\| \;\; \forall\, w \in \rho B, \tag{11.52}$$

$$\inf_{y\in K}\left\|T'\left(T^{-1}(y)\right)w\right\| \ge \frac{1}{M}\|w\| \;\; \forall\, w \in \mathbb{R}^d, \tag{11.53}$$

*and $\lim_{r\downarrow 0}\varepsilon_K(r) = 0$ where*

$$\varepsilon_K(r) := \sup_{y\in K}\sup_{\|w\|\le rM}\left\|\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right] - T'\left(T^{-1}(y)\right)w\right\|_{op}. \tag{11.54}$$

**Proof.** To simplify notation in the proof, let $S := T^{-1} : T(\Omega) \to \Omega$. For $y \in K$ and $\|w\| \le \rho$ we have by the fundamental theorem of calculus,

[5] We have $K \subset K_\rho \subset T(\Omega)$ and $K_\rho$ is closed and bounded and hence compact.

$$S(y+w) - S(y) = \left[\int_0^1 S'(y+sw)\,ds\right] w \tag{11.55}$$

which along with standard estimates gives Eq. (11.52).

By the chain rule applied to the identity, $y = T \circ S(y)$ for all $y \in T(\Omega)$, we have $I = T'(S(y))\,S'(y)$, i.e. $S'(y) = T'(S(y))^{-1}$. By the very definition of $M$, we have

$$\left\|T'(S(y))^{-1} z\right\| = \|S'(y)\,z\| \le M\,\|z\| \ \ \forall\ y \in K \text{ and } z \in \mathbb{R}^d.$$

Taking $z = T'(S(y))\,w$ in this inequality leads directly to Eq. (11.53).

Let $\hat{\rho} \in (0,\rho]$ be chosen so that $M\hat{\rho} < dist(S(K), \Omega^c)$. In proving the last assertion we always assume that $r \in (0,\hat{\rho})$. By the the fundamental theorem of calculus, for $y \in K$ and $w \in rMB$, we have

$$\begin{aligned}
&\frac{1}{r}\left[T(S(y)+rw) - y\right] - T'(S(y))\,w \\
&= \left[\frac{1}{r}\int_0^r T'((S(y)+\sigma w))\,d\sigma\right] w - T'(S(y))\,w \\
&= \left[\frac{1}{r}\int_0^r \left[T'((S(y)+\sigma w)) - T'(S(y))\right] d\sigma\right] w \\
&= \int_0^r \left[T'((S(y)+\sigma w)) - T'(S(y))\right] \frac{w}{r} d\sigma.
\end{aligned}$$

It is now a simple matter to use this identity and simple estimates to show

$$\varepsilon_K(r) \le \sup_{y\in K} \sup_{\|w\|\le rM} \left\|T'((S(y)+w)) - T'(S(y))\right\|_{op} \cdot M$$

which tends to 0 as $r \downarrow 0$ by uniform continuity of $T'$ on

$$[S(K)]_{M\hat{\rho}} = \left\{x \in \mathbb{R}^d : d_{S(K)}(x) \le M\hat{\rho}\right\}.$$

■

We are now ready for the proof of Proposition 11.33.

**Proof of Proposition 11.33.** Let $K$ be a compact subset of $T(\Omega)$, $0 < \rho \le 1$ be as in Eq. (11.49), and let $y \in K$ and $0 < r \le \rho$ and let us continue to use the notation in Lemma 11.34. If $x \in \Omega$ is such that $\delta_r(T(x) - y) > 0$, then $\|T(x) - y\| \le r$ or equivalently that $T(x) \in y + rB$, i.e. $x \in T^{-1}(y + rB)$. Now the estimate in Eq. (11.52) implies[6]

$$T^{-1}(y+rB) \subset T^{-1}(y) + rMB \ \forall\ y \in K \text{ and } r \le \rho$$

and hence $J_r(y)$ defined in Eq. (11.45) may be described by;

$$J_r(y) := \int_{T^{-1}(y)+rMB} \delta_r(T(x) - y)\,dx \ \forall\ y \in K.$$

We now make the affine change of variables, $x = T^{-1}(y) + rw$ in the above integral to find

$$J_r(y) := \int_{w\in MB} \delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right]\right) dw \ \forall\ y \in K.$$

Using the dominated convergence theorem along with the continuity of $\delta_1$ it follows that

$$\begin{aligned}
\lim_{r\downarrow 0} J_r(y) &= \int_{MB} \lim_{r\downarrow 0} \delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right]\right) dw \\
&= \int_{MB} \delta_1\left(\lim_{r\downarrow 0}\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right]\right) dw \\
&= \int_{MB} \delta_1\left(T'\left(T^{-1}(y)\right) w\right) dw.
\end{aligned} \tag{11.56}$$

According to Eq. (11.53), $\left\|T'\left(T^{-1}(y)\right) w\right\| > 1$ if $\|w\| > M$ and therefore,

$$\begin{aligned}
\int_{MB} \delta_1\left(T'\left(T^{-1}(y)\right) w\right) dw &= \int_{\mathbb{R}^d} \delta_1\left(T'\left(T^{-1}(y)\right) w\right) dw \\
&= \int_{\mathbb{R}^d} \delta_1(z) \frac{1}{\left|\det T'\left(T^{-1}(y)\right)\right|} dz \\
&= \frac{1}{\left|\det T'\left(T^{-1}(y)\right)\right|}
\end{aligned}$$

wherein we have made the linear change of variables, $z = T'\left(T^{-1}(y)\right) w$ in the second equality[7] and used $\int_{\mathbb{R}^d} \delta_1(z)\,dz = 1$ in the last equality.

[6] At this point in [31], it is implicitly asserted that $w \to \delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right]\right)$ is supported in $B$. This is however falsie in general. For example if $M > 1$ and $T(x) = M^{-1}x$, then

$$w \to \delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y) + rw\right) - y\right]\right) = \delta_1\left(M^{-1}w\right)$$

which is supported in $MB$ in general and not in $B$.

[7] Note well that $y$ is fixed here and so $z$ is varying with $w$ only.

To finish the proof we still need to show the convergence in Eq. (11.56) is uniform over $y \in K$. However, for $y \in K$,

$$\begin{aligned}
&\left|J_r(y) - \int_{MB} \delta_1\left(T'\left(T^{-1}(y)\right)w\right)dw\right| \\
&= \left|\int_{MB}\left[\delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y)+rw\right)-y\right]\right) - \delta_1\left(T'\left(T^{-1}(y)\right)w\right)\right]dw\right| \\
&\leq \int_{MB}\left|\delta_1\left(\frac{1}{r}\left[T\left(T^{-1}(y)+rw\right)-y\right]\right) - \delta_1\left(T'\left(T^{-1}(y)\right)w\right)\right|dw \\
&\leq \|\nabla\delta_1\|_\infty \cdot m(MB)\cdot\varepsilon_K(r) \to 0 \text{ as } r \downarrow 0
\end{aligned}$$

wherein $\varepsilon_K(r)$ is as in Eq. (11.54). The proof is complete since we have already seen in Lemma 11.34 that $\lim_{r\downarrow 0}\varepsilon_K(r) = 0$. ■

### 11.6.2 Radon Nykodym proof.

*Remark 11.35 (A proof using the Radon Nykodym theorem).* As usual let $T : \Omega \to T(\Omega)$ be a $C^1$ – diffeomorpshism and assume both $T$ and $T^{-1}$ have globally bounded Lipschitz constants (this can be achieved by shrinking $\Omega$ if necessary). We will work in the $\ell^\infty$ – norm on $\mathbb{R}^d$.

1. If $f : T(\Omega) \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ we have
$$\int_\Omega f\circ T dm = \int_{T(\Omega)} fd\left(m\circ T^{-1}\right) \tag{11.57}$$
and
$$\int_{T(\Omega)} g\circ T^{-1}dm = \int_\Omega gd(m\circ T). \tag{11.58}$$
2. Referring to the math 140 notes, show $|m\circ T(A)| \leq Km(A)$ and similarly $\left|m\circ T^{-1}(A)\right| \leq Km(A)$. Therefore by the easiest version of the the Radon – Nykodym there are bounded non-negative functions, $\alpha$ and $\beta$ such that
$$\begin{aligned}
\frac{d(m\circ T)}{dm} &= \alpha : \Omega \to \mathbb{R} \text{ and} \\
\frac{d\left(m\circ T^{-1}\right)}{dm} &= \beta : T(\Omega) \to \mathbb{R}.
\end{aligned}$$
In other words we now have,
$$\int_\Omega f\circ T dm = \int_{T(\Omega)} f\beta dm \text{ and} \tag{11.59}$$
$$\int_{T(\Omega)} g\circ T^{-1}dm = \int_\Omega g\alpha dm. \tag{11.60}$$
3. There is a relationship between $\alpha$ and $\beta$. Indeed taking $g = (f\beta)\circ T$ shows
$$\begin{aligned}
\int_\Omega f\circ T dm = \int_{T(\Omega)} f\beta dm &= \int_{T(\Omega)} ((f\beta)\circ T)\circ T^{-1}dm \\
&= \int_\Omega (f\circ T)\cdot(\beta\circ T)\,\alpha dm
\end{aligned}$$
from which we conclude
$$(\beta\circ T)\,\alpha = 1 \text{ a.e.} \iff \alpha = \frac{1}{\beta\circ T} \tag{11.61}$$
We now wish to compute the functions $\alpha$ and $\beta$ by taking limits and for this we will use the Lebesgue differentiation theorem or the easier fact that $\delta_r * \alpha \to \alpha$ in $L^1_{loc}$ as $r \downarrow 0$ where $\delta_r(x) := \frac{1}{m(B_r(0))}1_{B_r(0)}$.
4. For $x \in \Omega$ and $y \in y \in B_r(x) = x + B_r(0)$,
$$\begin{aligned}
T(y) - T(x) &= \int_0^1 T'(x+t(y-x))(y-x)\,dt \\
&= \left[T'(x) + \int_0^1 \left[T'(x+t(y-x)) - T'(x)\right]dt\right](y-x) \\
&= T'(x)\left[I + \varepsilon(x,y)\right](y-x)
\end{aligned}$$
where
$$\varepsilon(x,y) := \int_0^1 \left[T'(x)^{-1}T'(x+t(y-x)) - I\right]dt.$$
The error term, $\varepsilon(x,y)$, satisfies
$$\|\varepsilon(x,y)\| \leq \int_0^1 \left\|T'(x)^{-1}T'(x+t(y-x)) - I\right\|dt = \varepsilon_0(r)$$
where $\varepsilon_0(r)$ denotes a function of $r \geq 0$ such that $\lim_{r\downarrow 0}\varepsilon_0(r) = \varepsilon_0(0) = 0$. Combining these identities and estimates while using
$$\|[I+\varepsilon(x,y)](y-x)\| \leq r(1+\varepsilon(r)) \;\; \forall\; y \in B_r(x),$$
implies
$$T(B_r(x)) \subset T(x) + T'(x)\,B_{r\cdot(1+\varepsilon(r))}(0). \tag{11.62}$$

5. We now give two arguments showing $\alpha(x) = \frac{d(m\circ T)}{dm}(x) \le |\det T'(x)|$ for a.e. $x \in \Omega$.
   a) From Eq. (11.62) it follows that
$$\begin{aligned}\frac{m(T(B_r(x)))}{m(B_r(x))} &\le \frac{m\left(T(x) + T'(x)B_{r(1+\varepsilon(r))}(0)\right)}{m(B_r(x))}\\ &= |\det T'(x)| \cdot (1+\varepsilon(r))^d.\end{aligned}$$
   Letting $r \downarrow 0$ and using the Lebesgue differentiation theorem (see Theorem **??**) which is a rather deep result!)
$$\alpha(x) = \frac{d(m\circ T)}{dm}(x) \le |\det T'(x)| \text{ for a.e. } x \in \Omega. \tag{11.63}$$
   Applying this result with $T$ replaced by $T^{-1}$ then shows,
$$\beta(y) = \frac{d\left(m\circ T^{-1}\right)}{dm}(y) \le \left|\det\left(T^{-1}\right)'(y)\right| \text{ for a.e. } y \in T(\Omega). \tag{11.64}$$
   b) Alternatively let us observe that
$$\frac{m(T(B_r(x)))}{m(B_r(x))} = \frac{1}{m(B_r(x))}\int_{B_r(0)} \alpha(x+y)\,dy = \delta_r * \alpha(x).$$
   By the easier approximate identity Theorem **??** we know $\delta_r * \alpha \to \alpha$ in $L^1_{loc}$ and so there exists $r_n \downarrow 0$ such that $\delta_{r_n} * \alpha \to \alpha$ a.e. as $n \to \infty$. Thus we again learn that for a.e. $x$,
$$\begin{aligned}\alpha(x) &= \lim_{n\to\infty}\delta_{r_n} * \alpha(x) = \lim_{n\to\infty}\frac{m(T(B_{r_n}(x)))}{m(B_{r_n}(x))}\\ &\le \lim_{n\to\infty}|\det T'(x)| \cdot (1+\varepsilon(r_n))^d = |\det T'(x)|.\end{aligned}$$
6. We now use $\alpha = \frac{1}{\beta\circ T}$ from Eq. (11.61) along with the inequality in Eq. (11.64) to learn
$$\alpha(x) = \frac{1}{\beta\circ T(x)} \ge \frac{1}{\left|\det\left(T^{-1}\right)'(T(x))\right|}.$$
   On the other hand, since $T^{-1}\circ T = I$, it follows by the chain rule that $\left(T^{-1}\right)'(T(x))\,T'(x) = I$ and therefore
$$\frac{1}{\left|\det\left(T^{-1}\right)'(T(x))\right|} = |\det T'(x)|$$
   and we may conclude $\alpha(x) \ge |\det T'(x)|$. This result along with the inequality in Eq. (11.63) shows
$$\alpha(x) = |\det T'(x)|.$$
7. Using this result back in Eq. (11.60) with $g = f\circ T$ for some function on $f : T(\Omega) \to \mathbb{R}$ gives,
$$\int_{T(\Omega)} f\,dm = \int_\Omega f\circ T \cdot |\det T'|\,dm.$$

**Note:** By working harder as in the inverse function Theorem **??**, we could have proved the stronger version of Eq. (11.62);
$$T(x) + T'(x)B_{r(1-\varepsilon(r))}(0) \subset T(B_r(x)) \subset T(x) + T'(x)B_{r(1+\varepsilon(r))}(0).$$
If we had done this we could have avoided discussing $\beta$ altogether. Indeed, the method of step 5. would then give
$$|\det T'(x)| \cdot (1-\varepsilon(r))^d \le \frac{m(T(B_r(x)))}{m(B_r(x))} \le |\det T'(x)| \cdot (1+\varepsilon(r))^d$$
which upon letting $r \downarrow 0$ would have shown $\frac{d[m\circ T]}{dm}(x) = |\det T'(x)|$.

## 11.7 The Polar Decomposition of Lebesgue Measure*

Let
$$S^{d-1} = \{x \in \mathbb{R}^d : |x|^2 := \sum_{i=1}^d x_i^2 = 1\}$$
be the unit sphere in $\mathbb{R}^d$ equipped with its Borel $\sigma$ − algebra, $\mathcal{B}_{S^{d-1}}$ and $\Phi : \mathbb{R}^d \setminus \{0\} \to (0,\infty) \times S^{d-1}$ be defined by $\Phi(x) := (|x|, |x|^{-1}x)$. The inverse map, $\Phi^{-1} : (0,\infty) \times S^{d-1} \to \mathbb{R}^d \setminus \{0\}$, is given by $\Phi^{-1}(r,\omega) = r\omega$. Since $\Phi$ and $\Phi^{-1}$ are continuous, they are both Borel measurable. For $E \in \mathcal{B}_{S^{d-1}}$ and $a > 0$, let
$$E_a := \{r\omega : r \in (0,a] \text{ and } \omega \in E\} = \Phi^{-1}((0,a] \times E) \in \mathcal{B}_{\mathbb{R}^d}.$$

**Definition 11.36.** *For $E \in \mathcal{B}_{S^{d-1}}$, let $\sigma(E) := d\cdot m(E_1)$. We call $\sigma$ the surface measure on $S^{d-1}$.*

It is easy to check that $\sigma$ is a measure. Indeed if $E \in \mathcal{B}_{S^{d-1}}$, then $E_1 = \Phi^{-1}((0,1] \times E) \in \mathcal{B}_{\mathbb{R}^d}$ so that $m(E_1)$ is well defined. Moreover if $E = \sum_{i=1}^\infty E_i$, then $E_1 = \sum_{i=1}^\infty (E_i)_1$ and
$$\sigma(E) = d\cdot m(E_1) = \sum_{i=1}^\infty m((E_i)_1) = \sum_{i=1}^\infty \sigma(E_i).$$
The intuition behind this definition is as follows. If $E \subset S^{d-1}$ is a set and $\varepsilon > 0$ is a small number, then the volume of

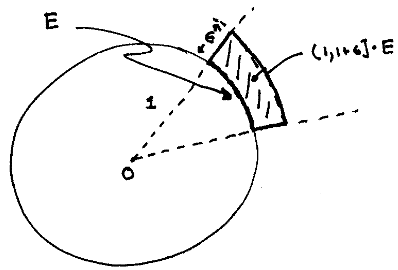



**Fig. 11.5.** Motivating the definition of surface measure for a sphere.

$$(1, 1+\varepsilon] \cdot E = \{r\omega : r \in (1, 1+\varepsilon] \text{ and } \omega \in E\}$$

should be approximately given by $m\left((1, 1+\varepsilon] \cdot E\right) \cong \sigma(E)\varepsilon$, see Figure 11.5 below. On the other hand

$$m\left((1, 1+\varepsilon]E\right) = m\left(E_{1+\varepsilon} \setminus E_1\right) = \left\{(1+\varepsilon)^d - 1\right\} m(E_1).$$

Therefore we expect the area of $E$ should be given by

$$\sigma(E) = \lim_{\varepsilon \downarrow 0} \frac{\left\{(1+\varepsilon)^d - 1\right\} m(E_1)}{\varepsilon} = d \cdot m(E_1).$$

The following theorem is motivated by Example 11.26 and Exercise 11.3.

**Theorem 11.37 (Polar Coordinates).** *If $f : \mathbb{R}^d \to [0, \infty]$ is a $(\mathcal{B}_{\mathbb{R}^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x)dm(x) = \int_{(0,\infty) \times S^{d-1}} f(r\omega) r^{d-1} \, dr d\sigma(\omega). \tag{11.65}$$

*In particular if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable then*

$$\int_{\mathbb{R}^d} f(|x|)dx = \int_0^\infty f(r)dV(r) \tag{11.66}$$

*where $V(r) = m\left(B(0, r)\right) = r^d m\left(B(0, 1)\right) = d^{-1}\sigma\left(S^{d-1}\right) r^d$.*

**Proof.** By Exercise 9.11,

$$\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d \setminus \{0\}} \left(f \circ \Phi^{-1}\right) \circ \Phi \, dm = \int_{(0,\infty) \times S^{d-1}} \left(f \circ \Phi^{-1}\right) \, d\left(\Phi_* m\right) \tag{11.67}$$

and therefore to prove Eq. (11.65) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right) \; \forall \; A \in \mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}. \tag{11.68}$$

If $A = (a, b] \times E$ with $0 < a < b$ and $E \in \mathcal{B}_{S^{d-1}}$, then

$$\Phi^{-1}(A) = \{r\omega : r \in (a, b] \text{ and } \omega \in E\} = bE_1 \setminus aE_1$$

wherein we have used $E_a = aE_1$ in the last equality. Therefore by the basic scaling properties of $m$ and the fundamental theorem of calculus,

$$\begin{aligned}\left(\Phi_* m\right)\left((a, b] \times E\right) &= m\left(bE_1 \setminus aE_1\right) = m(bE_1) - m(aE_1) \\ &= b^d m(E_1) - a^d m(E_1) = d \cdot m(E_1) \int_a^b r^{d-1} dr.\end{aligned} \tag{11.69}$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$\rho(J) = \int_J r^{d-1} dr \; \forall \; J \in \mathcal{B}_{(0,\infty)}, \tag{11.70}$$

Eq. (11.69) may be written as

$$\left(\Phi_* m\right)\left((a, b] \times E\right) = \rho((a, b]) \cdot \sigma(E) = \left(\rho \otimes \sigma\right)\left((a, b] \times E\right). \tag{11.71}$$

Since

$$\mathcal{E} = \left\{(a, b] \times E : 0 < a < b \text{ and } E \in \mathcal{B}_{S^{d-1}}\right\},$$

is a $\pi$ class (in fact it is an elementary class) such that $\sigma(\mathcal{E}) = \mathcal{B}_{(0,\infty)} \otimes \mathcal{B}_{S^{d-1}}$, it follows from the $\pi - \lambda$ Theorem and Eq. (11.71) that $\Phi_* m = \rho \otimes \sigma$. Using this result in Eq. (11.67) gives

$$\int_{\mathbb{R}^d} f dm = \int_{(0,\infty) \times S^{d-1}} \left(f \circ \Phi^{-1}\right) \, d\left(\rho \otimes \sigma\right)$$

which combined with Tonelli's Theorem 11.6 proves Eq. (11.67). ■

**Corollary 11.38.** *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1} \subset \mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{11.72}$$

*where $\Gamma$ is the gamma function is as in Example 9.47 and 9.50.*

**Proof.** Using Theorem 11.37 we find

$$I_d(1) = \int_0^\infty dr\, r^{d-1} e^{-r^2} \int_{S^{d-1}} d\sigma = \sigma(S^{d-1}) \int_0^\infty r^{d-1} e^{-r^2} dr.$$

We simplify this last integral by making the change of variables $u = r^2$ so that $r = u^{1/2}$ and $dr = \frac{1}{2} u^{-1/2} du$. The result is

$$\begin{aligned} \int_0^\infty r^{d-1} e^{-r^2} dr &= \int_0^\infty u^{\frac{d-1}{2}} e^{-u} \frac{1}{2} u^{-1/2} du \\ &= \frac{1}{2} \int_0^\infty u^{\frac{d}{2}-1} e^{-u} du = \frac{1}{2} \Gamma(d/2). \end{aligned} \tag{11.73}$$

Combing the the last two equations with Lemma 11.29 which states that $I_d(1) = \pi^{d/2}$, we conclude that

$$\pi^{d/2} = I_d(1) = \frac{1}{2} \sigma(S^{d-1}) \Gamma(d/2)$$

which proves Eq. (11.72). ■

## 11.8 More Spherical Coordinates*

In this section we will define spherical coordinates in all dimensions. Along the way we will develop an explicit method for computing surface integrals on spheres. As usual when $n = 2$ define spherical coordinates $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ so that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = T_2(\theta, r).$$

For $n = 3$ we let $x_3 = r\cos\varphi_1$ and then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = T_2(\theta, r\sin\varphi_1),$$

as can be seen from Figure 11.6, so that

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} T_2(\theta, r\sin\varphi_1) \\ r\cos\varphi_1 \end{pmatrix} = \begin{pmatrix} r\sin\varphi_1\cos\theta \\ r\sin\varphi_1\sin\theta \\ r\cos\varphi_1 \end{pmatrix} =: T_3(\theta, \varphi_1, r, ).$$

We continue to work inductively this way to define

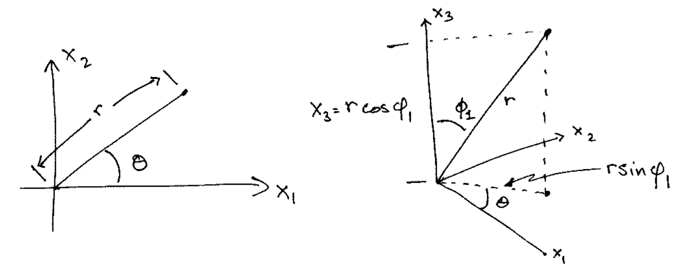



**Fig. 11.6.** Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1}, ) \\ r\cos\varphi_{n-1} \end{pmatrix} = T_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r).$$

So for example,

$$\begin{aligned} x_1 &= r\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_2\cos\varphi_1 \\ x_4 &= r\cos\varphi_2 \end{aligned}$$

and more generally,

$$\begin{aligned} x_1 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\cos\varphi_1 \\ &\vdots \\ x_{n-2} &= r\sin\varphi_{n-2}\sin\varphi_{n-3}\cos\varphi_{n-4} \\ x_{n-1} &= r\sin\varphi_{n-2}\cos\varphi_{n-3} \\ x_n &= r\cos\varphi_{n-2}. \end{aligned} \tag{11.74}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x) dm(x) \\ &= \int_0^\infty dr \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} d\varphi_1\dots d\varphi_{n-2} d\theta \left[ \begin{array}{c} \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) \\ \times f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)) \end{array} \right] \end{aligned} \tag{11.75}$$

where

$$\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) := \left|\det T_n'(\theta,\varphi_1,\dots,\varphi_{n-2},r)\right|.$$

**Proposition 11.39.** *The Jacobian, $\Delta_n$ is given by*

$$\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) = r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1. \quad (11.76)$$

*If $f$ is a function on $rS^{n-1}$ – the sphere of radius $r$ centered at 0 inside of $\mathbb{R}^n$, then*

$$\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1}\int_{S^{n-1}} f(r\omega)d\sigma(\omega)$$
$$= \int_{0\le\varphi_i\le\pi,0\le\theta\le2\pi} f(T_n(\theta,\varphi_1,\dots,\varphi_{n-2},r))\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)d\varphi_1\dots d\varphi_{n-2}d\theta \quad (11.77)$$

**Proof.** We are going to compute $\Delta_n$ inductively. Letting $\rho := r\sin\varphi_{n-1}$ and writing $\frac{\partial T_n}{\partial\xi}$ for $\frac{\partial T_n}{\partial\xi}(\theta,\varphi_1,\dots,\varphi_{n-2},\rho)$ we have

$$\begin{aligned}\Delta_{n+1}(\theta,&\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r)\\ &= \left|\begin{bmatrix} \frac{\partial T_n}{\partial\theta} & \frac{\partial T_n}{\partial\varphi_1} & \dots & \frac{\partial T_n}{\partial\varphi_{n-2}} & \frac{\partial T_n}{\partial\rho}r\cos\varphi_{n-1} & \frac{\partial T_n}{\partial\rho}\sin\varphi_{n-1} \\ 0 & 0 & \dots & 0 & -r\sin\varphi_{n-1} & \cos\varphi_{n-1}\end{bmatrix}\right|\\ &= r\left(\cos^2\varphi_{n-1}+\sin^2\varphi_{n-1}\right)\Delta_n(,\theta,\varphi_1,\dots,\varphi_{n-2},\rho)\\ &= r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}),\end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta,\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r) = r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}). \quad (11.78)$$

To arrive at this result we have expanded the determinant along the bottom row. Staring with $\Delta_2(\theta,r) = r$ already derived in Example 11.26, Eq. (11.78) implies,

$$\begin{aligned}\Delta_3(\theta,\varphi_1,r) &= r\Delta_2(\theta,r\sin\varphi_1) = r^2\sin\varphi_1\\ \Delta_4(\theta,\varphi_1,\varphi_2,r) &= r\Delta_3(\theta,\varphi_1,r\sin\varphi_2) = r^3\sin^2\varphi_2\sin\varphi_1\\ &\vdots\\ \Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) &= r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1\end{aligned}$$

which proves Eq. (11.76). Equation (11.77) now follows from Eqs. (11.65), (11.75) and (11.76). ■

As a simple application, Eq. (11.77) implies

$$\begin{aligned}\sigma(S^{n-1}) &= \int_{0\le\varphi_i\le\pi,0\le\theta\le2\pi}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1 d\varphi_1\dots d\varphi_{n-2}d\theta\\ &= 2\pi\prod_{k=1}^{n-2}\gamma_k = \sigma(S^{n-2})\gamma_{n-2}\end{aligned} \quad (11.79)$$

where $\gamma_k := \int_0^\pi \sin^k\varphi d\varphi$. If $k\ge1$, we have by integration by parts that,

$$\begin{aligned}\gamma_k &= \int_0^\pi \sin^k\varphi d\varphi = -\int_0^\pi \sin^{k-1}\varphi\ d\cos\varphi = 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \cos^2\varphi d\varphi\\ &= 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \left(1-\sin^2\varphi\right)d\varphi = 2\delta_{k,1} + (k-1)\left[\gamma_{k-2}-\gamma_k\right]\end{aligned}$$

and hence $\gamma_k$ satisfies $\gamma_0 = \pi$, $\gamma_1 = 2$ and the recursion relation

$$\gamma_k = \frac{k-1}{k}\gamma_{k-2} \text{ for } k\ge2.$$

Hence we may conclude

$$\gamma_0 = \pi,\ \gamma_1 = 2,\ \gamma_2 = \frac{1}{2}\pi,\ \gamma_3 = \frac{2}{3}2,\ \gamma_4 = \frac{3}{4}\frac{1}{2}\pi,\ \gamma_5 = \frac{4}{5}\frac{2}{3}2,\ \gamma_6 = \frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k} = \pi\frac{(2k-1)!!}{(2k)!!} \text{ and } \gamma_{2k+1} = 2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1} = \frac{2k+2}{2k+3}\gamma_{2k+1} = \frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!} = 2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)} = \frac{2k+1}{2k+1}\gamma_{2k} = \frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!} = \pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (11.79) may be written as

$$\sigma(S^n) = \sigma\left(S^{n-1}\right)\gamma_{n-1} \quad (11.80)$$

which combined with $\sigma\left(S^1\right) = 2\pi$ implies

$$\sigma\left(S^1\right) = 2\pi,$$
$$\sigma(S^2) = 2\pi \cdot \gamma_1 = 2\pi \cdot 2,$$
$$\sigma(S^3) = 2\pi \cdot 2 \cdot \gamma_2 = 2\pi \cdot 2 \cdot \frac{1}{2}\pi = \frac{2^2\pi^2}{2!!},$$
$$\sigma(S^4) = \frac{2^2\pi^2}{2!!} \cdot \gamma_3 = \frac{2^2\pi^2}{2!!} \cdot 2\frac{2}{3} = \frac{2^3\pi^2}{3!!}$$
$$\sigma(S^5) = 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi = \frac{2^3\pi^3}{4!!},$$
$$\sigma(S^6) = 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi \cdot \frac{4}{5}\frac{2}{3}2 = \frac{2^4\pi^3}{5!!}$$

and more generally that

$$\sigma(S^{2n}) = \frac{2\,(2\pi)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{(2\pi)^{n+1}}{(2n)!!} \tag{11.81}$$

which is verified inductively using Eq. (11.80). Indeed,

$$\sigma(S^{2n+1}) = \sigma(S^{2n})\gamma_{2n} = \frac{2\,(2\pi)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!} = \frac{(2\pi)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)}) = \sigma(S^{2n+2}) = \sigma(S^{2n+1})\gamma_{2n+1} = \frac{(2\pi)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!} = \frac{2\,(2\pi)^{n+1}}{(2n+1)!!}.$$

Using

$$(2n)!! = 2n\,(2(n-1))\ldots(2\cdot 1) = 2^n n!$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (11.65) and (11.81 are in agreement. We may also write the formula in Eq. (11.81) as

$$\sigma(S^n) = \begin{cases} \frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!} & \text{for } n \text{ odd.} \end{cases}$$

## 11.9 Gaussian Random Vectors

**Definition 11.40 (Gaussian Random Vectors).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}^d$ be a random vector. We say that $X$ is Gaussian if there exists an $d \times d$ – symmetric matrix $Q$ and a vector $\mu \in \mathbb{R}^d$ such that*

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = \exp\left(-\frac{1}{2}Q\lambda\cdot\lambda + i\mu\cdot\lambda\right) \text{ for all } \lambda \in \mathbb{R}^d. \tag{11.82}$$

*We will write $X \overset{d}{=} N(Q,\mu)$ to denote a Gaussian random vector such that Eq. (11.82) holds.*

Notice that if there exists a random variable satisfying Eq. (11.82) then its law is uniquely determined by $Q$ and $\mu$ because of Corollary 10.13. In the exercises below your will develop some basic properties of Gaussian random vectors – see Theorem 11.44 for a summary of what you will prove.

**Exercise 11.5.** Show that $Q$ must be non-negative in Eq. (11.82).

**Definition 11.41.** *Given a Gaussian random vector, $X$, we call the pair, $(Q,\mu)$ appearing in Eq. (11.82) the **characteristics** of $X$. We will also abbreviate the statement that $X$ is a Gaussian random vector with characteristics $(Q,\mu)$ by writing $X \overset{d}{=} N(Q,\mu)$.*

**Lemma 11.42.** *Suppose that $X \overset{d}{=} N(Q,\mu)$ and $A : \mathbb{R}^d \to \mathbb{R}^m$ is a $m \times d$ – real matrix and $\alpha \in \mathbb{R}^m$, then $AX + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. In short we might abbreviate this by saying, $AN(Q,\mu) + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$.*

**Proof.** Let $\xi \in \mathbb{R}^m$, then

$$\begin{aligned}\mathbb{E}\left[e^{i\xi\cdot(AX+\alpha)}\right] &= e^{i\xi\cdot\alpha}\mathbb{E}\left[e^{iA^{\mathrm{tr}}\xi\cdot X}\right] = e^{i\xi\cdot\alpha}\exp\left(-\frac{1}{2}QA^{\mathrm{tr}}\xi\cdot A^{\mathrm{tr}}\xi + i\mu\cdot A^{\mathrm{tr}}\xi\right) \\ &= e^{i\xi\cdot\alpha}\exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi\cdot\xi + iA\mu\cdot\xi\right) \\ &= \exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi\cdot\xi + i\,(A\mu+\alpha)\cdot\xi\right)\end{aligned}$$

from which it follows that $AX + \alpha \overset{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. ■

**Exercise 11.6.** Let $P$ be the probability measure on $\Omega := \mathbb{R}^d$ defined by

$$dP(x) := \left(\frac{1}{2\pi}\right)^{d/2} e^{-\frac{1}{2}x\cdot x}dx = \prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}}e^{-x_i^2/2}dx_i\right).$$

Show that $N : \Omega \to \mathbb{R}^d$ defined by $N(x) = x$ is Gaussian and satisfies Eq. (11.82) with $Q = I$ and $\mu = 0$. Also show

$$\mu_i = \mathbb{E}N_i \text{ and } \delta_{ij} = \mathrm{Cov}\,(N_i, N_j) \text{ for all } 1 \le i,j \le d. \tag{11.83}$$

**Hint:** use Exercise 9.15 and (of course) Fubini's theorem.

**Exercise 11.7.** Let $A$ be any real $m \times d$ matrix and $\mu \in \mathbb{R}^m$ and set $X := AN + \mu$ where $\Omega = \mathbb{R}^d$, $P$, and $N$ are as in Exercise 11.6. Show that $X$ is Gaussian by showing Eq. (11.82) holds with $Q = AA^{\text{tr}}$ ($A^{\text{tr}}$ is the transpose of the matrix $A$) and $\mu = \mu$. Also show that

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \operatorname{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{11.84}$$

*Remark 11.43 (Spectral Theorem).* Recall that if $Q$ is a real symmetric $d \times d$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u\}_{j=1}^d$, such that $Qu_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{R}$. Moreover, $\lambda_j \ge 0$ for all $j$ is equivalent to $Q$ being non-negative. When $Q \ge 0$ we may define $Q^{1/2}$ by

$$Q^{1/2} u_j := \sqrt{\lambda_j} u_j \text{ for } 1 \le j \le d.$$

Notice that $Q^{1/2} \ge 0$ and $Q = \left(Q^{1/2}\right)^2$ and $Q^{1/2}$ is still symmetric. If $Q$ is positive definite, we may also define, $Q^{-1/2}$ by

$$Q^{-1/2} u_j := \frac{1}{\sqrt{\lambda_j}} u_j \text{ for } 1 \le j \le d$$

so that $Q^{-1/2} = \left[Q^{1/2}\right]^{-1}$.

**Exercise 11.8.** Suppose that $Q$ is a positive definite (for simplicity) $d \times d$ real matrix and $\mu \in \mathbb{R}^d$ and let $\Omega = \mathbb{R}^d$, $P$, and $N$ be as in Exercise 11.6. By Exercise 11.7 we know that $X = Q^{1/2}N + \mu$ is a Gaussian random vector satisfying Eq. (11.82). Use the multi-dimensional change of variables formula to show

$$\operatorname{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}} \exp\left(-\frac{1}{2} Q^{-1}(y-\mu) \cdot (y-\mu)\right) dy.$$

Let us summarize some of what the preceding exercises have shown.

**Theorem 11.44.** *To each positive definite $d \times d$ real symmetric matrix $Q$ and $\mu \in \mathbb{R}^d$ there exist Gaussian random vectors, $X$, satisfying Eq. (11.82). Moreover for such an $X$,*

$$\operatorname{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}} \exp\left(-\frac{1}{2} Q^{-1}(y-\mu) \cdot (y-\mu)\right) dy$$

*where $Q$ and $\mu$ may be computed from $X$ using,*

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \operatorname{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{11.85}$$

When $Q$ is degenerate, i.e. $\operatorname{Nul}(Q) \ne \{0\}$, then $X = Q^{1/2}N + \mu$ is still a Gaussian random vectors satisfying Eq. (11.82). However now the $\operatorname{Law}_P(X)$ is a measure on $\mathbb{R}^d$ which is concentrated on the non-trivial subspace, $\operatorname{Nul}(Q)^\perp$ – the details of this are left to the reader for now.

**Exercise 11.9 (Gaussian random vectors are "highly" integrable.).** Suppose that $X : \Omega \to \mathbb{R}^d$ is a Gaussian random vector, say $X \overset{d}{=} N(Q, \mu)$. Let $\|x\| := \sqrt{x \cdot x}$ and $m := \max\{Qx \cdot x : \|x\| = 1\}$ be the largest eigenvalue[8] of $Q$. Then $\mathbb{E}\left[e^{\varepsilon \|X\|^2}\right] < \infty$ for every $\varepsilon < \frac{1}{2m}$.

*Remark 11.45.* We can in fact compute $\mathbb{E}\left[e^{\varepsilon \|X\|^2}\right]$ exactly – see Eq. (11.86) below for the final answer. To this end, let $\{u_i\}_{i=1}^d$ be a diagonalizing orthonormal basis (not necessarily the standard basis) for $Q$, i.e. $Qu = q_i u$ with $q_i \ge 0$ for $1 \le i \le d$. Further let $Z_i := N \cdot u_i$ and observe that $(Z_1, \dots, Z_d)$ is a linear transform of $N$, namely

$$Z := \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = RN$$

where $R$ is the rotation matrix with rows, $u_1^{\text{tr}}, u_2^{\text{tr}}, \dots, u_d^{\text{tr}}$ respectively. From this it follows that $Z \overset{d}{=} N(R^{\text{tr}}R, 0) = N(I, 0)$ is still a standard normal random vector. Moreover,

$$\begin{aligned} \|X\|^2 &= \|AN + \mu\|^2 = \|AN\|^2 + \|\mu\|^2 + 2AN \cdot \mu \\ &= \sum_{i=1}^d \left(q_i Z_i^2 + 2\sqrt{q_i}\mu_i Z_i\right) + \|\mu\|^2 \end{aligned}$$

where $\mu_i := \mu \cdot u_i$. Therefore using the independence of the $\{Z_i\}_{i=1}^d$,

$$\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right] = e^{\varepsilon\|\mu\|^2} \cdot \prod_{i=1}^d \mathbb{E}\left[e^{\varepsilon\left(q_i Z_i^2 + 2\sqrt{q_i}\mu_i Z_i\right)}\right].$$

Now observe that for any $\alpha < 1$, $t \in \mathbb{R}$, and $Z \overset{d}{=} N(0,1)$, that

$$\begin{aligned} \mathbb{E}\left[e^{\alpha Z^2 + tZ}\right] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\alpha z^2 + tz} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(1-2\alpha)z^2 + tz} dz \\ &= \frac{1}{\sqrt{(1-2\alpha)}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2 + \frac{t}{\sqrt{1-2\alpha}}x} dx \\ &= \frac{1}{\sqrt{(1-2\alpha)}} \exp\left(\frac{1}{2} \frac{t^2}{(1-2\alpha)}\right). \end{aligned}$$

wherein we have made the change of variables, $z = \frac{1}{\sqrt{1-2\alpha}}x$. So taking $\alpha = \varepsilon q_i$ and $t = 2\varepsilon\sqrt{q_i}\mu_i$ we may conclude

[8] For those who know about operator norms observe that $m = \|Q\|$ in this case.

$$\begin{aligned}\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right] &= e^{\varepsilon\|\mu\|^2}\cdot\prod_{i=1}^d \frac{1}{\sqrt{1-2\varepsilon q_i}}\exp\left(\frac{\left(2\varepsilon\sqrt{q_i}\mu_i\right)^2}{2\left(1-2\varepsilon q_i\right)}\right)\\ &= e^{\varepsilon\|\mu\|^2}\cdot\frac{1}{\sqrt{\det\left(I-2\varepsilon Q\right)}}\exp\left(2\varepsilon^2\sum_{i=1}^d\frac{q_i}{1-2\varepsilon q_i}\mu_i^2\right)\\ &= e^{\varepsilon\|\mu\|^2}\cdot\frac{1}{\sqrt{\det\left(I-2\varepsilon Q\right)}}\exp\left(2\varepsilon^2\sum_{i=1}^d\frac{q_i}{1-2\varepsilon q_i}\mu_i^2\right).\end{aligned}$$

To simplify this expression more observe that

$$\begin{aligned}\varepsilon\|\mu\|^2+2\varepsilon^2\sum_{i=1}^d\frac{q_i}{1-2\varepsilon q_i}\mu_i^2 &= \sum_{i=1}^d\left[2\varepsilon^2\frac{q_i}{1-2\varepsilon q_i}+\varepsilon\right]\mu_i^2\\ &= \sum_{i=1}^d\frac{\varepsilon}{1-2\varepsilon q_i}\mu_i^2=\varepsilon\left(1-2\varepsilon Q\right)^{-1}\mu\cdot\mu.\end{aligned}$$

Thus we have shown,

$$\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right]=\frac{1}{\sqrt{\det\left(I-2\varepsilon Q\right)}}\exp\left(\varepsilon\mu\cdot\left(1-2\varepsilon Q\right)^{-1}\mu\right). \tag{11.86}$$

Because of Eq. (11.85), for all $\lambda\in\mathbb{R}^d$ we have

$$\mu\cdot\lambda=\sum_{i=1}^d\mathbb{E}X_i\cdot\lambda_i=\mathbb{E}\left(\lambda\cdot X\right)$$

and

$$\begin{aligned}Q\lambda\cdot\lambda &= \sum_{i,j}Q_{ij}\lambda_i\lambda_j=\sum_{i,j}\lambda_i\lambda_j\operatorname{Cov}\left(X_i,X_j\right)\\ &= \operatorname{Cov}\left(\sum_i\lambda_iX_i,\sum_j\lambda_jX_j\right)=\operatorname{Var}\left(\lambda\cdot X\right).\end{aligned}$$

Therefore we may reformulate the definition of a Gaussian random vector as follows.

**Definition 11.46 (Gaussian Random Vectors).** *Let $(\Omega,\mathcal{B},P)$ be a probability space. A random vector, $X:\Omega\to\mathbb{R}^d$, is Gaussian iff for all $\lambda\in\mathbb{R}^d$,*

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right]=\exp\left(-\frac{1}{2}\operatorname{Var}\left(\lambda\cdot X\right)+i\mathbb{E}\left(\lambda\cdot X\right)\right). \tag{11.87}$$

*In short, $X$ is a Gaussian random vector iff $\lambda\cdot X$ is a Gaussian random variable for all $\lambda\in\mathbb{R}^d$.*

*Remark 11.47.* To conclude that a random vector, $X:\Omega\to\mathbb{R}^d$, is Gaussian it is **not** enough to check that each of its components, $\{X_i\}_{i=1}^d$, are Gaussian random variables. The following simple counter example was provided by Nate Eldredge. Let $(X,Y):\Omega\to\mathbb{R}^2$ be a Random vector such that $(X,Y)_*P=\mu\otimes\nu$ where $d\mu(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$ and $\nu=\frac{1}{2}\left(\delta_{-1}+\delta_1\right)$. Then $(X,YX):\Omega\to\mathbb{R}^2$ is a random vector such that both components, $X$ and $YX$, are Gaussian random variables but $(X,YX)$ is **not** a Gaussian random vector.

**Exercise 11.10.** Prove the assertion made in Remark 11.47. **Hint:** explicitly compute $\mathbb{E}\left[e^{i(\lambda_1X+\lambda_2XY)}\right]$.

### 11.9.1 *Gaussian measures with possibly degenerate covariances

The main aim of this subsection is to explicitly describe Gaussian measures with possibly degenerate covariances, $Q$. The case where $Q>0$ has already been done in Theorem 11.44.

*Remark 11.48.* Recall that if $Q$ is a real symmetric $N\times N$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u\}_{j=1}^N$, such that $Qu_j=\lambda_ju_j$ for some $\lambda_j\in\mathbb{R}$. Moreover, $\lambda_j\ge0$ for all $j$ is equivalent to $Q$ being non-negative. Hence if $Q\ge0$ and $f:\{\lambda_j:j=1,2,\dots,N\}\to\mathbb{R}$, we may define $f(Q)$ to be the unique linear transformation on $\mathbb{R}^N$ such that $f(Q)u_j=\lambda_ju_j$.

*Example 11.49.* When $Q\ge0$ and $f(x):=\sqrt{x}$, we write $Q^{1/2}$ or $\sqrt{Q}$ for $f(Q)$. Notice that $Q^{1/2}\ge0$ and $Q=Q^{1/2}Q^{1/2}$.

*Example 11.50.* When $Q$ is symmetric and

$$f(x)=\begin{cases}1/x \text{ if } x\neq0\\ \;\,0 \quad\text{if } x=0\end{cases}$$

we will denote $f(Q)$ by $Q^{-1}$. As the notation suggests, $f(Q)$ is the inverse of $Q$ when $Q$ is invertible which happens iff $\lambda_i\neq0$ for all $i$. When $Q$ is not invertible,

$$Q^{-1}:=f(Q)=Q|_{\operatorname{Ran}(Q)}^{-1}P, \tag{11.88}$$

where $P:\mathbb{R}^N\to\mathbb{R}^N$ be orthogonal projection onto the $\operatorname{Ran}(Q)$. Observe that $P=g(Q)$ where $g(x)=1_{x\neq0}$.

**Lemma 11.51.** *For any $Q\ge0$ we can find a matrix, $A$, such that $Q=AA^{\mathrm{tr}}$. In fact it suffices to take $A=Q^{1/2}$.*

**Proposition 11.52.** *Suppose $X\overset{d}{=}N(Q,c)$ (see Definition 11.41) where $c\in\mathbb{R}^N$ and $Q$ is a positive semi-definite $N\times N$ real matrix. If $\mu=\mu_{(Q,c)}=P\circ X^{-1}$, then*

$$\int_{\mathbb{R}^N} f(x)\,d\mu(x) = \frac{1}{Z}\int_{c+\operatorname{Ran}(Q)} f(x)\exp\left(-\frac{1}{2}Q^{-1}(x-c)\cdot(x-c)\right)dx$$

*where $dx$ is now "Lebesgue measure" on $c+\operatorname{Ran}(Q)$, $Q^{-1}$ is defined as in Eq. (11.88), and $Z := \sqrt{\det\left(2\pi Q|_{\operatorname{Ran}(Q)}\right)}$.*

**Proof.** Let $k = \dim \operatorname{Ran}(Q)$ and choose a linear transformation, $U : \mathbb{R}^k \to \mathbb{R}^N$, such that $\operatorname{Ran}(U) = \operatorname{Ran}(Q)$ and $U : \mathbb{R}^k \to \operatorname{Ran}(Q)$ is an isometric isomorphism. Letting $A := Q^{1/2}U$, we have

$$AA^{\mathrm{tr}} = Q^{1/2}UU^{\mathrm{tr}}Q^{1/2} = Q^{1/2}P_{\operatorname{Ran}(Q)}Q^{1/2} = Q.$$

Therefore, if $Y = N(I_{k\times k}, 0)$, then $X = AY + c \overset{d}{=} N(Q,c)$ by Lemma 11.42. Observe that $X - c = Q^{1/2}UY$ takes values in $\operatorname{Ran}(Q)$ and hence the Law of $(X-c)$ is a probability measure on $\mathbb{R}^N$ which is concentrated on $\operatorname{Ran}(Q)$. From this it follows that $\mu = P\circ X^{-1}$ is a probability measure on measure on $\mathbb{R}^N$ which is concentrated on the affine space, $c+\operatorname{Ran}(Q)$. At any rate from Theorem 11.44 we have

$$\begin{aligned}\int_{\mathbb{R}^N} f(x)\,d\mu(x) &= \int_{\mathbb{R}^k} f(Ay+c)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}|y|^2}dy \\ &= \int_{\mathbb{R}^k} f\left(Q^{1/2}Uy+c\right)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}|y|^2}dy.\end{aligned}$$

Since

$$Q^{1/2}Uy + c = UU^{\mathrm{tr}}Q^{1/2}Uy + c,$$

we may make the change of variables, $z = U^{\mathrm{tr}}Q^{1/2}Uy$, using

$$dz = \sqrt{\det Q|_{\operatorname{Ran}(Q)}}dy = \sqrt{\prod_{i:\lambda_i\neq 0}\lambda_i dy}$$

and

$$\begin{aligned}|y|^2 &= \left|\left(U^{\mathrm{tr}}Q^{1/2}U\right)^{-1}z\right|^2 = \left|U^{\mathrm{tr}}Q^{-1/2}Uz\right|^2 = \left(Q^{-1/2}Uz, Q^{-1/2}Uz\right)_{\mathbb{R}^N} \\ &= \left(Q^{-1}Uz, Uz\right)_{\mathbb{R}^N},\end{aligned}$$

to find



$$\begin{aligned}\int_{\mathbb{R}^N} f(x)\,d\mu(x) &= \int_{\mathbb{R}^k} f(Uz+c)\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}\left|\left(U^{\mathrm{tr}}Q^{1/2}U\right)^{-1}z\right|^2}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\left(\frac{1}{2\pi}\right)^{k/2}\frac{1}{\sqrt{\det Q|_{\operatorname{Ran}(Q)}}}e^{-\frac{1}{2}\left(Q^{-1}Uz,Uz\right)_{\mathbb{R}^N}}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\frac{1}{\sqrt{\det\left(2\pi Q|_{\operatorname{Ran}(Q)}\right)}}e^{-\frac{1}{2}\left(Q^{-1}Uz,Uz\right)_{\mathbb{R}^N}}dz \\ &= \int_{\mathbb{R}^k} f(Uz+c)\frac{1}{\sqrt{\det\left(2\pi Q|_{\operatorname{Ran}(Q)}\right)}}e^{-\frac{1}{2}\left(Q^{-1}(Uz+c-c),(Uz+c-c)\right)_{\mathbb{R}^N}}dz.\end{aligned}$$

This completes the proof, since $x = Uz + c \in c + \operatorname{Ran}(Q)$ is by definition distributed as Lebesgue measure on $c+\operatorname{Ran}(Q)$ when $z$ is distributed as Lebesgue measure on $\mathbb{R}^k$. ■

## 11.10 Kolmogorov's Extension Theorems

In this section we will extend the results of Section 6.10 to spaces which are not simply products of discrete spaces. We begin with a couple of results involving the topology on $\mathbb{R}^N$.

### 11.10.1 Regularity and compactness results

**Theorem 11.53 (Inner-Outer Regularity).** *Suppose $\mu$ is a probability measure on $\left(\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N}\right)$, then for all $B\in\mathcal{B}_{\mathbb{R}^N}$ we have*

$$\mu(B) = \inf\{\mu(V) : B\subset V \text{ and } V \text{ is open}\} \tag{11.89}$$

*and*

$$\mu(B) = \sup\{\mu(K) : K\subset B \text{ with } K \text{ compact}\}. \tag{11.90}$$

**Proof.** In this proof, $C$, and $C_i$ will always denote a closed subset of $\mathbb{R}^N$ and $V$, $V_i$ will always be open subsets of $\mathbb{R}^N$. Let $\mathcal{F}$ be the collection of sets, $A\in\mathcal{B}$, such that for all $\varepsilon > 0$ there exists an open set $V$ and a closed set, $C$, such that $C\subset A\subset V$ and $\mu(V\setminus C) < \varepsilon$. The key point of the proof is to show $\mathcal{F} = \mathcal{B}$ for this certainly implies Equation (11.89) and also that

$$\mu(B) = \sup\{\mu(C) : C\subset B \text{ with } C \text{ closed}\}. \tag{11.91}$$

Moreover, by MCT, we know that if $C$ is closed and $K_n := C\cap\left\{x\in\mathbb{R}^N : |x|\leq n\right\}$, then $\mu(K_n)\uparrow\mu(C)$. This observation along with Eq. (11.91) shows Eq. (11.90) is valid as well.

To prove $\mathcal{F} = \mathcal{B}$, it suffices to show $\mathcal{F}$ is a $\sigma$ – algebra which contains all closed subsets of $\mathbb{R}^N$. To the prove the latter assertion, given a closed subset, $C \subset \mathbb{R}^N$, and $\varepsilon > 0$, let

$$C_\varepsilon := \cup_{x\in C} B(x,\varepsilon)$$

where $B(x,\varepsilon) := \left\{y \in \mathbb{R}^N : |y-x| < \varepsilon\right\}$. Then $C_\varepsilon$ is an open set and $C_\varepsilon \downarrow C$ as $\varepsilon \downarrow 0$. (You prove.) Hence by the DCT, we know that $\mu(C_\varepsilon \setminus C) \downarrow 0$ form which it follows that $C \in \mathcal{F}$.

We will now show that $\mathcal{F}$ is an algebra. Clearly $\mathcal{F}$ contains the empty set and if $A \in \mathcal{F}$ with $C \subset A \subset V$ and $\mu(V \setminus C) < \varepsilon$, then $V^c \subset A^c \subset C^c$ with $\mu(C^c \setminus V^c) = \mu(V \setminus C) < \varepsilon$. This shows $A^c \in \mathcal{F}$. Similarly if $A_i \in \mathcal{F}$ for $i = 1, 2$ and $C_i \subset A_i \subset V_i$ with $\mu(V_i \setminus C_i) < \varepsilon$, then

$$C := C_1 \cup C_2 \subset A_1 \cup A_2 \subset V_1 \cup V_2 =: V$$

and

$$\begin{aligned}\mu(V \setminus C) &\le \mu(V_1 \setminus C) + \mu(V_2 \setminus C)\\ &\le \mu(V_1 \setminus C_1) + \mu(V_2 \setminus C_2) < 2\varepsilon.\end{aligned}$$

This implies that $A_1 \cup A_2 \in \mathcal{F}$ and we have shown $\mathcal{F}$ is an algebra.

We now show that $\mathcal{F}$ is a $\sigma$ – algebra. To do this it suffices to show $A := \sum_{n=1}^\infty A_n \in \mathcal{F}$ if $A_n \in \mathcal{F}$ with $A_n \cap A_m = \emptyset$ for $m \neq n$. Let $C_n \subset A_n \subset V_n$ with $\mu(V_n \setminus C_n) < \varepsilon 2^{-n}$ for all $n$ and let $C^N := \cup_{n\le N} C_n$ and $V := \cup_{n=1}^\infty V_n$. Then $C^N \subset A \subset V$ and

$$\begin{aligned}\mu\left(V \setminus C^N\right) &\le \sum_{n=0}^\infty \mu\left(V_n \setminus C^N\right) \le \sum_{n=0}^N \mu(V_n \setminus C_n) + \sum_{n=N+1}^\infty \mu(V_n)\\ &\le \sum_{n=0}^N \varepsilon 2^{-n} + \sum_{n=N+1}^\infty \left[\mu(A_n) + \varepsilon 2^{-n}\right]\\ &= \varepsilon + \sum_{n=N+1}^\infty \mu(A_n).\end{aligned}$$

The last term is less that $2\varepsilon$ for $N$ sufficiently large because $\sum_{n=1}^\infty \mu(A_n) = \mu(A) < \infty$. ■

**Notation 11.54** *Let $I := [0,1]$, $Q = I^{\mathbb{N}}$, $\pi_j : Q \to I$ be the projection map, $\pi_j(x) = x_j$ (where $x = (x_1, x_2, \dots, x_j, \dots)$ for all $j \in \mathbb{N}$, and $\mathcal{B}_Q := \sigma(\pi_j : j \in \mathbb{N})$ be the product $\sigma$ – algebra on $Q$. Let us further say that a sequence $\{x(m)\}_{m=1}^\infty \subset Q$, where $x(m) = (x_1(m), x_2(m), \dots)$, converges to $x \in Q$ iff $\lim_{m\to\infty} x_j(m) = x_j$ for all $j \in \mathbb{N}$. (This is just pointwise convergence.)*

**Lemma 11.55 (Baby Tychonoff's Theorem).** *The infinite dimensional cube, $Q$, is compact, i.e. every sequence $\{x(m)\}_{m=1}^\infty \subset Q$ has a convergent subsequence,$\{x(m_k)\}_{k=1}^\infty$.*

**Proof.** Since $I$ is compact, it follows that for each $j \in \mathbb{N}$, $\{x_j(m)\}_{m=1}^\infty$ has a convergent subsequence. It now follow by Cantor's diagonalization method, that there is a subsequence, $\{m_k\}_{k=1}^\infty$, of $\mathbb{N}$ such that $\lim_{k\to\infty} x_j(m_k) \in I$ exists for all $j \in \mathbb{N}$. ■

**Corollary 11.56 (Finite Intersection Property).** *Suppose that $K_m \subset Q$ are sets which are, (i) closed under taking sequential limits*[9]*, and (ii) have the finite intersection property, (i.e. $\cap_{m=1}^n K_m \neq \emptyset$ for all $m \in \mathbb{N}$), then $\cap_{m=1}^\infty K_m \neq \emptyset$.*

**Proof.** By assumption, for each $n \in \mathbb{N}$, there exists $x(n) \in \cap_{m=1}^n K_m$. Hence by Lemma 11.55 there exists a subsequence, $x(n_k)$, such that $x := \lim_{k\to\infty} x(n_k)$ exists in $Q$. Since $x(n_k) \in \cap_{m=1}^n K_m$ for all $k$ large, and each $K_m$ is closed under sequential limits, it follows that $x \in K_m$ for all $m$. Thus we have shown, $x \in \cap_{m=1}^\infty K_m$ and hence $\cap_{m=1}^\infty K_m \neq \emptyset$. ■

### 11.10.2 Kolmogorov's Extension Theorem and Infinite Product Measures

**Theorem 11.57 (Kolmogorov's Extension Theorem).** *Let $I := [0,1]$. For each $n \in \mathbb{N}$, let $\mu_n$ be a probability measure on $(I^n, \mathcal{B}_{I^n})$ such that $\mu_{n+1}(A \times I) = \mu_n(A)$. Then there exists a unique measure, $P$ on $(Q, \mathcal{B}_Q)$ such that*

$$P(A \times Q) = \mu_n(A) \tag{11.92}$$

*for all $A \in \mathcal{B}_{I^n}$ and $n \in \mathbb{N}$.*

**Proof.** Let $\mathcal{A} := \cup \mathcal{B}_n$ where $\mathcal{B}_n := \{A \times Q : A \in \mathcal{B}_{I^n}\} = \sigma(\pi_1, \dots, \pi_n)$, where $\pi_i(x) = x_i$ if $x = (x_1, x_2, \dots) \in Q$. Then define $P$ on $\mathcal{A}$ by Eq. (11.92) which is easily seen (Exercise 11.11) to be a well defined finitely additive measure on $\mathcal{A}$. So to finish the proof it suffices to show if $B_n \in \mathcal{A}$ is a decreasing sequence such that

$$\inf_n P(B_n) = \lim_{n\to\infty} P(B_n) = \varepsilon > 0,$$

then $B := \cap B_n \neq \emptyset$.

To simplify notation, we may reduce to the case where $B_n \in \mathcal{B}_n$ for all $n$. To see this is permissible, Let us choose $1 \le n_1 < n_2 < n_3 < \dots$. such that

[9] For example, if $K_m = K_m' \times Q$ with $K_m'$ being a closed subset of $I^m$, then $K_m$ is closed under sequential limits.

$B_k \in \mathcal{B}_{n_k}$ for all $k$. (This is possible since $\mathcal{B}_n$ is increasing in $n$.) We now define a new decreasing sequence of sets, $\left\{\tilde{B}_k\right\}_{k=1}^{\infty}$ as follows,

$$\left(\tilde{B}_1, \tilde{B}_2, \dots\right) = \left(\overbrace{Q, \dots, Q}^{n_1-1 \text{ times}}, \overbrace{B_1, \dots, B_1}^{n_2-n_1 \text{ times}}, \overbrace{B_2, \dots, B_2}^{n_3-n_2 \text{ times}}, \overbrace{B_3, \dots, B_3}^{n_4-n_3 \text{ times}}, \dots\right).$$

We then have $\tilde{B}_n \in \mathcal{B}_n$ for all $n$, $\lim_{n\to\infty} P\left(\tilde{B}_n\right) = \varepsilon > 0$, and $B = \cap_{n=1}^{\infty} \tilde{B}_n$. Hence we may replace $B_n$ by $\tilde{B}_n$ if necessary so as to have $B_n \in \mathcal{B}_n$ for all $n$.

Since $B_n \in \mathcal{B}_n$, there exists $B'_n \in \mathcal{B}_{I^n}$ such that $B_n = B'_n \times Q$ for all $n$. Using the regularity Theorem 11.53, there are compact sets, $K'_n \subset B'_n \subset I^n$, such that $\mu_n(B'_n \setminus K'_n) \le \varepsilon 2^{-n-1}$ for all $n \in \mathbb{N}$. Let $K_n := K'_n \times Q$, then $P(B_n \setminus K_n) \le \varepsilon 2^{-n-1}$ for all $n$. Moreover,

$$\begin{aligned} P\left(B_n \setminus [\cap_{m=1}^n K_m]\right) &= P\left(\cup_{m=1}^n [B_n \setminus K_m]\right) \le \sum_{m=1}^n P(B_n \setminus K_m) \\ &\le \sum_{m=1}^n P(B_m \setminus K_m) \le \sum_{m=1}^n \varepsilon 2^{-m-1} \le \varepsilon/2. \end{aligned}$$

So, for all $n \in \mathbb{N}$,

$$P\left(\cap_{m=1}^n K_m\right) = P(B_n) - P\left(B_n \setminus [\cap_{m=1}^n K_m]\right) \ge \varepsilon - \varepsilon/2 = \varepsilon/2,$$

and in particular, $\cap_{m=1}^n K_m \ne \emptyset$. An application of Corollary 11.56 now implies, $\emptyset \ne \cap_n K_n \subset \cap_n B_n$. ∎

**Exercise 11.11.** Show that Eq. (11.92) defines a well defined finitely additive measure on $\mathcal{A} := \cup \mathcal{B}_n$.

The next result is an easy corollary of Theorem 11.57.

**Theorem 11.58.** *Suppose $\{(X_n, \mathcal{M}_n)\}_{n\in\mathbb{N}}$ are standard Borel spaces (see Appendix 11.11 below), $X := \prod_{n\in\mathbb{N}} X_n$, $\pi_n : X \to X_n$ be the $n^{th}$ – projection map, $\mathcal{B}_n := \sigma(\pi_k : k \le n)$, $\mathcal{B} = \sigma(\pi_n : n \in \mathbb{N})$, and $T_n := X_{n+1} \times X_{n+2} \times \dots$. Further suppose that for each $n \in \mathbb{N}$ we are given a probability measure, $\mu_n$ on $\mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$ such that*

$$\mu_{n+1}(A \times X_{n+1}) = \mu_n(A) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n.$$

*Then there exists a unique probability measure, $P$, on $(X, \mathcal{B})$ such that $P(A \times T_n) = \mu_n(A)$ for all $A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$.*

**Proof.** Since each $(X_n, \mathcal{M}_n)$ is measure theoretic isomorphic to a Borel subset of $I$, we may assume that $X_n \in \mathcal{B}_I$ and $\mathcal{M}_n = (\mathcal{B}_I)_{X_n}$ for all $n$. Given $A \in \mathcal{B}_{I^n}$, let $\bar{\mu}_n(A) := \mu_n(A \cap [X_1 \times \dots \times X_n])$ – a probability measure on $\mathcal{B}_{I^n}$. Furthermore,

$$\begin{aligned} \bar{\mu}_{n+1}(A \times I) &= \mu_{n+1}([A \times I] \cap [X_1 \times \dots \times X_{n+1}]) \\ &= \mu_{n+1}((A \cap [X_1 \times \dots \times X_n]) \times X_{n+1}) \\ &= \mu_n((A \cap [X_1 \times \dots \times X_n])) = \bar{\mu}_n(A). \end{aligned}$$

Hence by Theorem 11.57, there is a unique probability measure, $\bar{P}$, on $I^{\mathbb{N}}$ such that

$$\bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n(A) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{B}_{I^n}.$$

We will now check that $P := \bar{P}|_{\otimes_{n=1}^{\infty}\mathcal{M}_n}$ is the desired measure. First off we have

$$\begin{aligned} \bar{P}(X) &= \lim_{n\to\infty} \bar{P}\left(X_1 \times \dots \times X_n \times I^{\mathbb{N}}\right) = \lim_{n\to\infty} \bar{\mu}_n(X_1 \times \dots \times X_n) \\ &= \lim_{n\to\infty} \mu_n(X_1 \times \dots \times X_n) = 1. \end{aligned}$$

Secondly, if $A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$, we have

$$\begin{aligned} P(A \times T_n) &= \bar{P}(A \times T_n) = \bar{P}\left(\left(A \times I^{\mathbb{N}}\right) \cap X\right) \\ &= \bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n(A) = \mu_n(A). \end{aligned}$$

∎

Here is an example of this theorem in action.

**Theorem 11.59 (Infinite Product Measures).** *Suppose that $\{\nu_n\}_{n=1}^{\infty}$ are a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\mathcal{B} := \otimes_{n\in\mathbb{N}} \mathcal{B}_{\mathbb{R}}$ is the product $\sigma$ – algebra on $\mathbb{R}^{\mathbb{N}}$. Then there exists a unique probability measure, $\nu$, on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$, such that*

$$\nu\left(A_1 \times A_2 \times \dots \times A_n \times \mathbb{R}^{\mathbb{N}}\right) = \nu_1(A_1) \dots \nu_n(A_n) \;\; \forall\, A_i \in \mathcal{B}_{\mathbb{R}} \;\&\; n \in \mathbb{N}. \tag{11.93}$$

*Moreover, this measure satisfies,*

$$\int_{\mathbb{R}^{\mathbb{N}}} f(x_1, \dots, x_n)\, d\nu(x) = \int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, d\nu_1(x_1) \dots d\nu_n(x_n) \tag{11.94}$$

*for all $n \in \mathbb{N}$ and $f : \mathbb{R}^n \to \mathbb{R}$ which are bounded and measurable or non-negative and measurable.*

**Proof.** The measure $\nu$ is created by apply Theorem 11.58 with $\mu_n := \nu_1 \otimes \dots \otimes \nu_n$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n} = \otimes_{k=1}^n \mathcal{B}_{\mathbb{R}})$ for each $n \in \mathbb{N}$. Observe that

$$\mu_{n+1}(A\times\mathbb{R}) = \mu_n(A)\cdot\nu_{n+1}(\mathbb{R}) = \mu_n(A),$$

so that $\{\mu_n\}_{n=1}^\infty$ satisfies the needed consistency conditions. Thus there exists a unique measure $\nu$ on $(\mathbb{R}^{\mathbb{N}},\mathcal{B})$ such that

$$\nu\left(A\times\mathbb{R}^{\mathbb{N}}\right) = \mu_n(A) \text{ for all } A\in\mathcal{B}_{\mathbb{R}^n} \text{ and } n\in\mathbb{N}.$$

Taking $A = A_1\times A_2\times\cdots\times A_n$ with $A_i\in\mathcal{B}_{\mathbb{R}}$ then gives Eq. (11.93). For this measure, it follows that Eq. (11.94) holds when $f = 1_{A_1\times\cdots\times A_n}$. Thus by an application of Theorem 10.2 with $\mathbb{M} = \{1_{A_1\times\cdots\times A_n} : A_i\in\mathcal{B}_{\mathbb{R}}\}$ and $\mathbb{H}$ being the set of bounded measurable functions, $f:\mathbb{R}^n\to\mathbb{R}$, for which Eq. (11.94) shows that Eq. (11.94) holds for all bounded and measurable functions, $f:\mathbb{R}^n\to\mathbb{R}$. The statement involving non-negative functions follows by a simple limiting argument involving the MCT. ■

It turns out that the existence of infinite product measures require no topological restrictions on the measure spaces involved. See Corollary 19.71 below.

## 11.11 Appendix: Standard Borel Spaces*

For more information along the lines of this section, see Royden [39] and Parthasarathy [34].

**Definition 11.60.** *Two measurable spaces, $(X,\mathcal{M})$ and $(Y,\mathcal{N})$ are said to be* ***isomorphic*** *if there exists a bijective map, $f:X\to Y$ such that $f(\mathcal{M}) = \mathcal{N}$ and $f^{-1}(\mathcal{N}) = \mathcal{M}$, i.e. both $f$ and $f^{-1}$ are measurable. In this case we say $f$ is a measure theoretic isomorphism and we will write $X\cong Y$.*

**Definition 11.61.** *A measurable space, $(X,\mathcal{M})$ is said to be a* ***standard Borel space*** *if $(X,\mathcal{M})\cong(B,\mathcal{B}_B)$ where $B$ is a Borel subset of $\left((0,1),\mathcal{B}_{(0,1)}\right)$.*

**Definition 11.62 (Polish spaces).** *A* ***Polish space*** *is a separable topological space $(X,\tau)$ which admits a complete metric, $\rho$, such that $\tau=\tau_\rho$.*

The main goal of this chapter is to prove every Borel subset of a Polish space is a standard Borel space, see Corollary 11.72 below. Along the way we will show a number of spaces, including $[0,1]$, $(0,1]$, $[0,1]^d$, $\mathbb{R}^d$, $\{0,1\}^{\mathbb{N}}$, and $\mathbb{R}^{\mathbb{N}}$, are all (measure theoretic) isomorphic to $(0,1)$. Moreover we also will see that a countable product of standard Borel spaces is again a standard Borel space, see Corollary 11.69.

*On first reading, you may wish to skip the rest of this section.

**Lemma 11.63.** *Suppose $(X,\mathcal{M})$ and $(Y,\mathcal{N})$ are measurable spaces such that $X=\sum_{n=1}^\infty X_n$, $Y=\sum_{n=1}^\infty Y_n$, with $X_n\in\mathcal{M}$ and $Y_n\in\mathcal{N}$. If $(X_n,\mathcal{M}_{X_n})$ is isomorphic to $(Y_n,\mathcal{N}_{Y_n})$ for all $n$ then $X\cong Y$. Moreover, if $(X_n,\mathcal{M}_n)$ and $(Y_n,\mathcal{N}_n)$ are isomorphic measure spaces, then $(X:=\prod_{n=1}^\infty X_n,\otimes_{n=1}^\infty\mathcal{M}_n)$ are $(Y:=\prod_{n=1}^\infty Y_n,\otimes_{n=1}^\infty\mathcal{N}_n)$ are isomorphic.*

**Proof.** For each $n\in\mathbb{N}$, let $f_n:X_n\to Y_n$ be a measure theoretic isomorphism. Then define $f:X\to Y$ by $f=f_n$ on $X_n$. Clearly, $f:X\to Y$ is a bijection and if $B\in\mathcal{N}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty f^{-1}(B\cap Y_n) = \cup_{n=1}^\infty f_n^{-1}(B\cap Y_n)\in\mathcal{M}.$$

This shows $f$ is measurable and by similar considerations, $f^{-1}$ is measurable as well. Therefore, $f:X\to Y$ is the desired measure theoretic isomorphism.

For the second assertion, let $f_n:X_n\to Y_n$ be a measure theoretic isomorphism of all $n\in\mathbb{N}$ and then define

$$f(x) = (f_1(x_1), f_2(x_2),\dots) \text{ with } x=(x_1,x_2,\dots)\in X.$$

Again it is clear that $f$ is bijective and measurable, since

$$f^{-1}\left(\prod_{n=1}^\infty B_n\right) = \prod_{n=1}^\infty f_n^{-1}(B_n)\in\otimes_{n=1}^\infty\mathcal{N}_n$$

for all $B_n\in\mathcal{M}_n$ and $n\in\mathbb{N}$. Similar reasoning shows that $f^{-1}$ is measurable as well. ■

**Proposition 11.64.** *Let $-\infty<a<b<\infty$. The following measurable spaces equipped with there Borel $\sigma$ – algebras are all isomorphic; $(0,1)$, $[0,1]$, $(0,1]$, $[0,1)$, $(a,b)$, $[a,b]$, $(a,b]$, $[a,b)$, $\mathbb{R}$, and $(0,1)\cup\Lambda$ where $\Lambda$ is a finite or countable subset of $\mathbb{R}\setminus(0,1)$.*

**Proof.** It is easy to see by that any bounded open, closed, or half open interval is isomorphic to any other such interval using an affine transformation. Let us now show $(-1,1)\cong[-1,1]$. To prove this it suffices, by Lemma 11.63,to observe that

$$(-1,1) = \{0\}\cup\sum_{n=0}^\infty\left((-2^{-n},-2^{-n}]\cup[2^{-n-1},2^{-n})\right)$$

and

$$[-1,1] = \{0\}\cup\sum_{n=0}^\infty\left([-2^{-n},-2^{-n-1})\cup(2^{-n-1},2^{-n}]\right).$$

Similarly $(0,1)$ is isomorphic to $(0,1]$ because

$$(0,1)=\sum_{n=0}^{\infty}[2^{-n-1},2^{-n}) \text{ and } (0,1]=\sum_{n=0}^{\infty}(2^{-n-1},2^{-n}].$$

The assertion involving $\mathbb{R}$ can be proved using the bijection, $\tan : (-\pi/2,\pi/2)\to\mathbb{R}$.

If $\Lambda=\{1\}$, then by Lemma 11.63 and what we have already proved, $(0,1)\cup\{1\}=(0,1]\cong(0,1)$. Similarly if $N\in\mathbb{N}$ with $N\geq 2$ and $\Lambda=\{2,\dots,N+1\}$, then

$$(0,1)\cup\Lambda\cong(0,1]\cup\Lambda=(0,2^{-N+1}]\cup\left[\sum_{n=1}^{N-1}(2^{-n},2^{-n-1}]\right]\cup\Lambda$$

while

$$(0,1)=\left(0,2^{-N+1}\right)\cup\left[\sum_{n=1}^{N-1}\left(2^{-n},2^{-n-1}\right)\right]\cup\left\{2^{-n}:n=1,2,\dots,N\right\}$$

and so again it follows from what we have proved and Lemma 11.63 that $(0,1)\cong(0,1)\cup\Lambda$. Finally if $\Lambda=\{2,3,4,\dots\}$ is a countable set, we can show $(0,1)\cong(0,1)\cup\Lambda$ with the aid of the identities,

$$(0,1)=\left[\sum_{n=1}^{\infty}\left(2^{-n},2^{-n-1}\right)\right]\cup\left\{2^{-n}:n\in\mathbb{N}\right\}$$

and

$$(0,1)\cup\Lambda\cong(0,1]\cup\Lambda=\left[\sum_{n=1}^{\infty}(2^{-n},2^{-n-1}]\right]\cup\Lambda.$$

■

**Notation 11.65** *Suppose $(X,\mathcal{M})$ is a measurable space and $A$ is a set. Let $\pi_a : X^A\to X$ denote projection operator onto the $a^{th}$ – component of $X^A$ (i.e. $\pi_a(\omega)=\omega(a)$ for all $a\in A$) and let $\mathcal{M}^{\otimes A}:=\sigma(\pi_a : a\in A)$ be the product $\sigma$ – algebra on $X^A$.*

**Lemma 11.66.** *If $\varphi : A\to B$ is a bijection of sets and $(X,\mathcal{M})$ is a measurable space, then $\left(X^A,\mathcal{M}^{\otimes A}\right)\cong\left(X^B,\mathcal{M}^{\otimes B}\right)$.*

**Proof.** The map $f : X^B\to X^A$ defined by $f(\omega)=\omega\circ\varphi$ for all $\omega\in X^B$ is a bijection with $f^{-1}(\alpha)=\alpha\circ\varphi^{-1}$. If $a\in A$ and $\omega\in X^B$, we have

$$\pi_a^{X^A}\circ f(\omega)=f(\omega)(a)=\omega(\varphi(a))=\pi_{\varphi(a)}^{X^B}(\omega),$$

where $\pi_a^{X^A}$ and $\pi_b^{X^B}$ are the projection operators on $X^A$ and $X^B$ respectively. Thus $\pi_a^{X^A}\circ f=\pi_{\varphi(a)}^{X^B}$ for all $a\in A$ which shows $f$ is measurable. Similarly, $\pi_b^{X^B}\circ f^{-1}=\pi_{\varphi^{-1}(b)}^{X^A}$ showing $f^{-1}$ is measurable as well. ■



**Proposition 11.67.** *Let $\Omega:=\{0,1\}^{\mathbb{N}}$, $\pi_i : \Omega\to\{0,1\}$ be projection onto the $i^{th}$ component, and $\mathcal{B}:=\sigma(\pi_1,\pi_2,\dots)$ be the product $\sigma$ – algebra on $\Omega$. Then $(\Omega,\mathcal{B})\cong\left((0,1),\mathcal{B}_{(0,1)}\right)$.*

**Proof.** We will begin by using a specific binary digit expansion of a point $x\in[0,1)$ to construct a map from $[0,1)\to\Omega$. To this end, let $r_1(x)=x$,

$$\gamma_1(x):=1_{x\geq 2^{-1}} \text{ and } r_2(x):=x-2^{-1}\gamma_1(x)\in(0,2^{-1}),$$

then let $\gamma_2:=1_{r_2\geq 2^{-2}}$ and $r_3=r_2-2^{-2}\gamma_2\in\left(0,2^{-2}\right)$. Working inductively, we construct $\{\gamma_k(x),r_k(x)\}_{k=1}^{\infty}$ such that $\gamma_k(x)\in\{0,1\}$, and

$$r_{k+1}(x)=r_k(x)-2^{-k}\gamma_k(x)=x-\sum_{j=1}^{k}2^{-j}\gamma_j(x)\in\left(0,2^{-k}\right) \tag{11.95}$$

for all $k$. Let us now define $g : [0,1)\to\Omega$ by $g(x):=(\gamma_1(x),\gamma_2(x),\dots)$. Since each component function, $\pi_j\circ g=\gamma_j : [0,1)\to\{0,1\}$, is measurable it follows that $g$ is measurable.

By construction,

$$x=\sum_{j=1}^{k}2^{-j}\gamma_j(x)+r_{k+1}(x)$$

and $r_{k+1}(x)\to 0$ as $k\to\infty$, therefore

$$x=\sum_{j=1}^{\infty}2^{-j}\gamma_j(x) \text{ and } r_{k+1}(x)=\sum_{j=k+1}^{\infty}2^{-j}\gamma_j(x). \tag{11.96}$$

Hence if we define $f : \Omega\to[0,1]$ by $f=\sum_{j=1}^{\infty}2^{-j}\pi_j$, then $f(g(x))=x$ for all $x\in[0,1)$. This shows $g$ is injective, $f$ is surjective, and $f$ in injective on the range of $g$.

We now claim that $\Omega_0:=g([0,1))$, the range of $g$, consists of those $\omega\in\Omega$ such that $\omega_i=0$ for infinitely many $i$. Indeed, if there exists an $k\in\mathbb{N}$ such that $\gamma_j(x)=1$ for all $j\geq k$, then (by Eq. (11.96)) $r_{k+1}(x)=2^{-k}$ which would contradict Eq. (11.95). Hence $g([0,1))\subset\Omega_0$. Conversely if $\omega\in\Omega_0$ and $x=f(\omega)\in[0,1)$, it is not hard to show inductively that $\gamma_j(x)=\omega_j$ for all $j$, i.e. $g(x)=\omega$. For example, if $\omega_1=1$ then $x\geq 2^{-1}$ and hence $\gamma_1(x)=1$. Alternatively, if $\omega_1=0$, then

$$x=\sum_{j=2}^{\infty}2^{-j}\omega_j<\sum_{j=2}^{\infty}2^{-j}=2^{-1}$$

so that $\gamma_1(x)=0$. Hence it follows that $r_2(x)=\sum_{j=2}^{\infty}2^{-j}\omega_j$ and by similar reasoning we learn $r_2(x)\geq 2^{-2}$ iff $\omega_2=1$, i.e. $\gamma_2(x)=1$ iff $\omega_2=1$. The full induction argument is now left to the reader.

Since single point sets are in $\mathcal{B}$ and

$$\Lambda := \Omega \setminus \Omega_0 = \cup_{n=1}^{\infty} \{\omega \in \Omega : \omega_j = 1 \text{ for } j \geq n\}$$

is a countable set, it follows that $\Lambda \in \mathcal{B}$ and therefore $\Omega_0 = \Omega \setminus \Lambda \in \mathcal{B}$. Hence we may now conclude that $g : \left([0,1), \mathcal{B}_{[0,1)}\right) \to (\Omega_0, \mathcal{B}_{\Omega_0})$ is a measurable bijection with measurable inverse given by $f|_{\Omega_0}$, i.e. $\left([0,1), \mathcal{B}_{[0,1)}\right) \cong (\Omega_0, \mathcal{B}_{\Omega_0})$. An application of Lemma 11.63 and Proposition 11.64 now implies

$$\Omega = \Omega_0 \cup \Lambda \cong [0,1) \cup \mathbb{N} \cong [0,1) \cong (0,1).$$

■

**Corollary 11.68.** *The following spaces are all isomorphic to* $\left((0,1), \mathcal{B}_{(0,1)}\right)$*;* $(0,1)^d$ *and* $\mathbb{R}^d$ *for any* $d \in \mathbb{N}$ *and* $[0,1]^{\mathbb{N}}$ *and* $\mathbb{R}^{\mathbb{N}}$ *where both of these spaces are equipped with their natural product* $\sigma$ *– algebras, .*

**Proof.** In light of Lemma 11.63 and Proposition 11.64 we know that $(0,1)^d \cong \mathbb{R}^d$ and $(0,1)^{\mathbb{N}} \cong [0,1]^{\mathbb{N}} \cong \mathbb{R}^{\mathbb{N}}$. So, using Proposition 11.67, it suffices to show $(0,1)^d \cong \Omega \cong (0,1)^{\mathbb{N}}$ and to do this it suffices to show $\Omega^d \cong \Omega$ and $\Omega^{\mathbb{N}} \cong \Omega$.

To reduce the problem further, let us observe that $\Omega^d \cong \{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}}$ and $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. For example, let $g : \Omega^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}^2}$ be defined by $g(\omega)(i,j) = \omega(i)(j)$ for all $\omega \in \Omega^{\mathbb{N}} = \left[\{0,1\}^{\mathbb{N}}\right]^{\mathbb{N}}$. Then $g$ is a bijection and since $\pi_{(i,j)}^{\{0,1\}^{\mathbb{N}^2}} \circ g(\omega) = \pi_j^{\Omega}\left(\pi_i^{\Omega^{\mathbb{N}}}(\omega)\right)$, it follows that $g$ is measurable. The inverse, $g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega^{\mathbb{N}}$, to $g$ is given by $g^{-1}(\alpha)(i)(j) = \alpha(i,j)$. To see this map is measurable, we have $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega = \{0,1\}^{\mathbb{N}}$ is given $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}(\alpha) = g^{-1}(\alpha)(i)(\cdot) = \alpha(i,\cdot)$ and hence

$$\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g(\alpha) = \alpha(i,j) = \pi_{i,j}^{\{0,1\}^{\mathbb{N}^2}}(\alpha)$$

from which it follows that $\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} = \pi^{\{0,1\}^{\mathbb{N}^2}}$ is measurable for all $i,j \in \mathbb{N}$ and hence $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}$ is measurable for all $i \in \mathbb{N}$ and hence $g^{-1}$ is measurable. This shows $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. The proof that $\Omega^d \cong \{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}}$ is analogous.

We may now complete the proof with a couple of applications of Lemma 11.66. Indeed $\mathbb{N}$, $\mathbb{N}\times\{1,2,\dots,d\}$, and $\mathbb{N}^2$ all have the same cardinality and therefore,

$$\{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}} \cong \{0,1\}^{\mathbb{N}^2} \cong \{0,1\}^{\mathbb{N}} = \Omega.$$

■

**Corollary 11.69.** *Suppose that* $(X_n, \mathcal{M}_n)$ *for* $n \in \mathbb{N}$ *are standard Borel spaces, then* $X := \prod_{n=1}^{\infty} X_n$ *equipped with the product* $\sigma$ *– algebra,* $\mathcal{M} := \otimes_{n=1}^{\infty}\mathcal{M}_n$ *is again a standard Borel space.*

**Proof.** Let $A_n \in \mathcal{B}_{[0,1]}$ be Borel sets on $[0,1]$ such that there exists a measurable isomorpohism, $f_n : X_n \to A_n$. Then $f : X \to A := \prod_{n=1}^{\infty} A_n$ defined by $f(x_1, x_2, \dots) = (f_1(x_1), f_2(x_2), \dots)$ is easily seen to me a measure theoretic isomorphism when $A$ is equipped with the product $\sigma -$ algebra, $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$. So according to Corollary 11.68, to finish the proof it suffice to show $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n} = \mathcal{M}_A$ where $\mathcal{M} := \otimes_{n=1}^{\infty}\mathcal{B}_{[0,1]}$ is the product $\sigma -$ algebra on $[0,1]^{\mathbb{N}}$.

The $\sigma -$ algebra, $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$, is generated by sets of the form, $B := \prod_{n=1}^{\infty} B_n$ where $B_n \in \mathcal{B}_{A_n} \subset \mathcal{B}_{[0,1]}$. On the other hand, the $\sigma -$ algebra, $\mathcal{M}_A$ is generated by sets of the form, $A \cap \tilde{B}$ where $\tilde{B} := \prod_{n=1}^{\infty} \tilde{B}_n$ with $\tilde{B}_n \in \mathcal{B}_{[0,1]}$. Since

$$A \cap \tilde{B} = \prod_{n=1}^{\infty}\left(\tilde{B}_n \cap A_n\right) = \prod_{n=1}^{\infty} B_n$$

where $B_n = \tilde{B}_n \cap A_n$ is the generic element in $\mathcal{B}_{A_n}$, we see that $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$ and $\mathcal{M}_A$ can both be generated by the same collections of sets, we may conclude that $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n} = \mathcal{M}_A$. ■

Our next goal is to show that any Polish space with its Borel $\sigma -$ algebra is a standard Borel space.

**Notation 11.70** *Let* $Q := [0,1]^{\mathbb{N}}$ *denote the (infinite dimensional)* ***unit cube*** *in* $\mathbb{R}^{\mathbb{N}}$*. For* $a, b \in Q$ *let*

$$d(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n}|a_n - b_n| = \sum_{n=1}^{\infty} \frac{1}{2^n}|\pi_n(a) - \pi_n(b)|. \tag{11.97}$$

**Exercise 11.12.** Show $d$ is a metric and that the Borel $\sigma$ – algebra on $(Q,d)$ is the same as the product $\sigma$ – algebra.

**Theorem 11.71.** *To every separable metric space* $(X,\rho)$*, there exists a continuous injective map* $G : X \to Q$ *such that* $G : X \to G(X) \subset Q$ *is a homeomorphism. Moreover if the metric,* $\rho$*, is also complete, then* $G(X)$ *is a* $G_\delta$ *–set, i.e. the* $G(X)$ *is the countable intersection of open subsets of* $(Q,d)$*. In short, any separable metrizable space* $X$ *is homeomorphic to a subset of* $(Q,d)$ *and if* $X$ *is a Polish space then* $X$ *is homeomorphic to a* $G_\delta$ *– subset of* $(Q,d)$*.*

**Proof.** (This proof follows that in Rogers and Williams [38, Theorem 82.5 on p. 106.].) By replacing $\rho$ by $\frac{\rho}{1+\rho}$ if necessary, we may assume that $0 \leq \rho < 1$. Let $D = \{a_n\}_{n=1}^{\infty}$ be a countable dense subset of $X$ and define

$$G(x) = (\rho(x,a_1), \rho(x,a_2), \rho(x,a_3), \dots) \in Q$$

and

$$\gamma(x,y) = d(G(x), G(y)) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\rho(x,a_n) - \rho(y,a_n)|$$

for $x, y \in X$. To prove the first assertion, we must show $G$ is injective and $\gamma$ is a metric on $X$ which is compatible with the topology determined by $\rho$.

If $G(x) = G(y)$, then $\rho(x,a) = \rho(y,a)$ for all $a \in D$. Since $D$ is a dense subset of $X$, we may choose $\alpha_k \in D$ such that

$$0 = \lim_{k\to\infty} \rho(x,\alpha_k) = \lim_{k\to\infty} \rho(y,\alpha_k) = \rho(y,x)$$

and therefore $x = y$. A simple argument using the dominated convergence theorem shows $y \to \gamma(x,y)$ is $\rho$ – continuous, i.e. $\gamma(x,y)$ is small if $\rho(x,y)$ is small. Conversely,

$$\begin{aligned}\rho(x,y) &\le \rho(x,a_n) + \rho(y,a_n) = 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n) \\ &\le 2\rho(x,a_n) + |\rho(x,a_n) - \rho(y,a_n)| \le 2\rho(x,a_n) + 2^n\gamma(x,y).\end{aligned}$$

Hence if $\varepsilon > 0$ is given, we may choose $n$ so that $2\rho(x,a_n) < \varepsilon/2$ and so if $\gamma(x,y) < 2^{-(n+1)}\varepsilon$, it will follow that $\rho(x,y) < \varepsilon$. This shows $\tau_\gamma = \tau_\rho$. Since $G : (X,\gamma) \to (Q,d)$ is isometric, $G$ is a homeomorphism.

Now suppose that $(X,\rho)$ is a complete metric space. Let $S := G(X)$ and $\sigma$ be the metric on $S$ defined by $\sigma(G(x), G(y)) = \rho(x,y)$ for all $x, y \in X$. Then $(S,\sigma)$ is a complete metric (being the isometric image of a complete metric space) and by what we have just prove, $\tau_\sigma = \tau_{d_S}$. Consequently, if $u \in S$ and $\varepsilon > 0$ is given, we may find $\delta'(\varepsilon)$ such that $B_\sigma(u,\delta'(\varepsilon)) \subset B_d(u,\varepsilon)$. Taking $\delta(\varepsilon) = \min(\delta'(\varepsilon), \varepsilon)$, we have $\operatorname{diam}_d(B_d(u,\delta(\varepsilon))) < \varepsilon$ and $\operatorname{diam}_\sigma(B_d(u,\delta(\varepsilon))) < \varepsilon$ where

$$\operatorname{diam}_\sigma(A) := \{\sup \sigma(u,v) : u,v \in A\} \text{ and}$$
$$\operatorname{diam}_d(A) := \{\sup d(u,v) : u,v \in A\}.$$

Let $\bar{S}$ denote the closure of $S$ inside of $(Q,d)$ and for each $n \in \mathbb{N}$ let

$$\mathcal{N}_n := \{N \in \tau_d : \operatorname{diam}_d(N) \vee \operatorname{diam}_\sigma(N \cap S) < 1/n\}$$

and let $U_n := \cup\mathcal{N}_n \in \tau_d$. From the previous paragraph, it follows that $S \subset U_n$ and therefore $S \subset \bar{S} \cap (\cap_{n=1}^\infty U_n)$.

Conversely if $u \in \bar{S} \cap (\cap_{n=1}^\infty U_n)$ and $n \in \mathbb{N}$, there exists $N_n \in \mathcal{N}_n$ such that $u \in N_n$. Moreover, since $N_1 \cap \dots \cap N_n$ is an open neighborhood of $u \in \bar{S}$, there exists $u_n \in N_1 \cap \dots \cap N_n \cap S$ for each $n \in \mathbb{N}$. From the definition of $\mathcal{N}_n$, we have $\lim_{n\to\infty} d(u,u_n) = 0$ and $\sigma(u_n,u_m) \le \max(n^{-1}, m^{-1}) \to 0$ as $m, n \to \infty$. Since $(S,\sigma)$ is complete, it follows that $\{u_n\}_{n=1}^\infty$ is convergent in $(S,\sigma)$ to some element $u_0 \in S$. Since $(S,d_S)$ has the same topology as $(S,\sigma)$ it follows that $d(u_n,u_0) \to 0$ as well and thus that $u = u_0 \in S$. We have now shown, $S = \bar{S} \cap (\cap_{n=1}^\infty U_n)$. This completes the proof because we may write $\bar{S} = \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ where $S_{1/n} := \{u \in Q : d(u,\bar{S}) < 1/n\}$ and therefore, $S = \left(\bigcap_{n=1}^\infty U_n\right) \cap \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ is a $G_\delta$ set. ∎

**Corollary 11.72.** *Every Polish space, $X$, with its Borel $\sigma$ – algebra is a standard Borel space. Consequently any Borel subset of $X$ is also a standard Borel space.*

**Proof.** Theorem 11.71 shows that $X$ is homeomorphic to a measurable (in fact a $G_\delta$) subset $Q_0$ of $(Q,d)$ and hence $X \cong Q_0$. Since $Q$ is a standard Borel space so is $Q_0$ and hence so is $X$. ∎

## 11.12 More Exercises

**Exercise 11.13.** Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1,2,3$ be $\sigma$ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1,x_2),x_3) = (x_1,x_2,x_3).$$

1. Show $F$ is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and $F^{-1}$ is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2) \times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3) \to (X_1 \times X_2 \times X_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_*[(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3](F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then $\pi$ is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0,\infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1\times X_2\times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1,x_2,x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

**Exercise 11.14.** Prove the second assertion of Theorem 11.20. That is show $m^d$ is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$. **Hint:** Look at the proof of Theorem 6.61.

**Exercise 11.15.** (Part of Folland Problem 2.46 on p. 69.) Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel $\sigma$ – field on $X$, $m$ be Lebesgue measure on $[0,1]$ and $\nu$ be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x,x) \in X^2 : x \in X\}$ be the diagonal in $X^2$. Show

$$\int_X \left[\int_X 1_D(x,y)d\nu(y)\right] dm(x) \neq \int_X \left[\int_X 1_D(x,y)dm(x)\right] d\nu(y)$$

by explicitly computing both sides of this equation.

**Exercise 11.16.** Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

**Exercise 11.17.** Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

**Exercise 11.18.** Folland Problem 2.55 on p. 77. (Explicit integrations.)

**Exercise 11.19.** Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a),dm)$, $g(x) = \int_x^a \frac{f(t)}{t}dt$ for $x \in (0,a)$, show $g \in L^1((0,a),dm)$ and

$$\int_0^a g(x)dx = \int_0^a f(t)dt.$$

**Exercise 11.20.** Show $\int_0^\infty \left|\frac{\sin x}{x}\right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty),m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

**Exercise 11.21.** Folland Problem 2.57 on p. 77.

**Exercise 11.22.** Folland Problem 2.58 on p. 77.

**Exercise 11.23.** Folland Problem 2.60 on p. 77. Properties of the $\Gamma$ – function.

**Exercise 11.24.** Folland Problem 2.61 on p. 77. Fractional integration.

**Exercise 11.25.** Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on $S^{n-1}$.

**Exercise 11.26.** Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log|x|\right|^b$ for $x$ near 0 and $x$ near $\infty$ in $\mathbb{R}^n$.

**Exercise 11.27.** Show, using Problem 11.25 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d}\delta_{ij}\sigma\left(S^{d-1}\right).$$

**Hint:** show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of $i$ and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

## 11.13 Exercises

Many of the following exercises are probably repeats of the exercises above.

**Exercise 11.28.** Prove Theorem **??**. Suggestion, to get started define

$$\pi(A) := \int_{X_1} d\mu(x_1) \cdots \int_{X_n} d\mu(x_n) 1_A(x_1,\dots,x_n)$$

and then show Eq. (**??**) holds. Use the case of two factors as the model of your proof.

**Exercise 11.29.** Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1,2,3$ be $\sigma$ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1,x_2),x_3) = (x_1,x_2,x_3).$$

1. Show $F$ is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and $F^{-1}$ is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2) \times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3) \to (X_1 \times X_2 \times X_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_*\left[(\mu_1 \otimes \mu_2) \otimes \mu_3\right]$, i.e. $\pi(A) = \left[(\mu_1 \otimes \mu_2) \otimes \mu_3\right](F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then $\pi$ is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0,\infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1 \times X_2 \times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1,x_2,x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

**Exercise 11.30.** Prove the second assertion of Theorem **??**. That is show $m^d$ is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$. **Hint:** Look at the proof of Theorem **??**.

**Exercise 11.31.** (Part of Folland Problem 2.46 on p. 69.) Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel $\sigma$ – field on $X$, $m$ be Lebesgue measure on $[0,1]$ and $\nu$ be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x,x) \in X^2 : x \in X\}$ be the diagonal in $X^2$. Show

$$\int_X \left[\int_X 1_D(x,y)d\nu(y)\right] dm(x) \neq \int_X \left[\int_X 1_D(x,y)dm(x)\right] d\nu(y)$$

by explicitly computing both sides of this equation.

**Exercise 11.32.** Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

**Exercise 11.33.** Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

**Exercise 11.34.** Folland Problem 2.55 on p. 77. (Explicit integrations.)

**Exercise 11.35.** Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a),dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0,a)$, show $g \in L^1((0,a),dm)$ and

$$\int_0^a g(x)dx = \int_0^a f(t)dt.$$

**Exercise 11.36.** Show $\int_0^\infty \left|\frac{\sin x}{x}\right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty),m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

**Exercise 11.37.** Folland Problem 2.57 on p. 77.

**Exercise 11.38.** Folland Problem 2.58 on p. 77.

**Exercise 11.39.** Folland Problem 2.60 on p. 77. Properties of the $\Gamma$ – function.

**Exercise 11.40.** Folland Problem 2.61 on p. 77. Fractional integration.

**Exercise 11.41.** Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on $S^{n-1}$.

**Exercise 11.42.** Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log|x|\right|^b$ for $x$ near 0 and $x$ near $\infty$ in $\mathbb{R}^n$.

**Exercise 11.43.** Show, using Problem 11.41 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d}\delta_{ij}\sigma\left(S^{d-1}\right).$$

**Hint:** show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of $i$ and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

# Independence

As usual, $(\Omega, \mathcal{B}, P)$ will be some fixed probability space. Recall that for $A, B \in \mathcal{B}$ with $P(B) > 0$ we let

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

which is to be read as; **the probability of $A$ given $B$.**

**Definition 12.1.** *We say that $A$ is independent of $B$ is $P(A|B) = P(A)$ or equivalently that*

$$P(A \cap B) = P(A) P(B).$$

*We further say a finite sequence of collection of sets, $\{\mathcal{C}_i\}_{i=1}^n$, are independent if*

$$P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

*for all $A_i \in \mathcal{C}_i$ and $J \subset \{1, 2, \dots, n\}$.*

## 12.1 Basic Properties of Independence

If $\{\mathcal{C}_i\}_{i=1}^n$, are independent classes then so are $\{\mathcal{C}_i \cup \{\Omega\}\}_{i=1}^n$. Moreover, if we assume that $\Omega \in \mathcal{C}_i$ for each $i$, then $\{\mathcal{C}_i\}_{i=1}^n$, are independent iff

$$P\left(\cap_{j=1}^n A_j\right) = \prod_{j=1}^n P(A_j) \text{ for all } (A_1, \dots, A_n) \in \mathcal{C}_1 \times \dots \times \mathcal{C}_n.$$

**Theorem 12.2.** *Suppose that $\{\mathcal{C}_i\}_{i=1}^n$ is a finite sequence of independent $\pi$ – classes. Then $\{\sigma(\mathcal{C}_i)\}_{i=1}^n$ are also independent.*

**Proof.** As mentioned above, we may always assume without loss of generality that $\Omega \in \mathcal{C}_i$. Fix, $A_j \in \mathcal{C}_j$ for $j = 2, 3, \dots, n$. We will begin by showing that

$$Q(A) := P(A \cap A_2 \cap \dots \cap A_n) = P(A) P(A_2) \dots P(A_n) \text{ for all } A \in \sigma(\mathcal{C}_1). \tag{12.1}$$

Since $Q(\cdot)$ and $P(A_2) \dots P(A_n) P(\cdot)$ are both finite measures agreeing on $\Omega$ and $A$ in the $\pi$ – system $\mathcal{C}_1$, Eq. (12.1) is a direct consequence of Proposition 6.18. Since $(A_2, \dots, A_n) \in \mathcal{C}_2 \times \dots \times \mathcal{C}_n$ were arbitrary we may now conclude that $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ are independent.

By applying the result we have just proved to the sequence, $\mathcal{C}_2, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ shows that $\sigma(\mathcal{C}_2), \mathcal{C}_3, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ are independent. Similarly we show inductively that

$$\sigma(\mathcal{C}_j), \mathcal{C}_{j+1}, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_{j-1})$$

are independent for each $j = 1, 2, \dots, n$. The desired result occurs at $j = n$. ■

**Definition 12.3.** *Let $(\Omega, , \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. The maps $\{Y_i\}_{i=1}^n$ are $P$ - independent iff $\{\mathcal{C}_i\}_{i=1}^n$ are $P$ – independent, where $\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i) = \sigma(Y_i) \subset \mathcal{B}$ for $1 \le i \le n$.*

**Theorem 12.4 (Independence and Product Measures).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. Further let $\mu_i := P \circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$. Then $\{Y_i\}_{i=1}^n$ are independent iff*

$$\mathrm{Law}_P(Y_1, \dots, Y_n) = \mu_1 \otimes \dots \otimes \mu_n,$$

*where $(Y_1, \dots, Y_n) : \Omega \to S_1 \times \dots \times S_n$ and*

$$\mathrm{Law}_P(Y_1, \dots, Y_n) = P \circ (Y_1, \dots, Y_n)^{-1} : \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n \to [0, 1]$$

*is the joint law of $Y_1, \dots, Y_n$.*

**Proof.** Recall that the general element of $\mathcal{C}_i$ is of the form $A_i = Y_i^{-1}(B_i)$ with $B_i \in \mathcal{S}_i$. Therefore for $A_i = Y_i^{-1}(B_i) \in \mathcal{C}_i$ we have

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P((Y_1, \dots, Y_n) \in B_1 \times \dots \times B_n) \\ &= ((Y_1, \dots, Y_n)_* P)(B_1 \times \dots \times B_n). \end{aligned}$$

If $(Y_1, \dots, Y_n)_* P = \mu_1 \otimes \dots \otimes \mu_n$ it follows that

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= \mu_1 \otimes \dots \otimes \mu_n (B_1 \times \dots \times B_n) \\ &= \mu_1(B_1) \cdots \mu(B_n) = P(Y_1 \in B_1) \cdots P(Y_n \in B_n) \\ &= P(A_1) \dots P(A_n) \end{aligned}$$

and therefore $\{\mathcal{C}_i\}$ are $P$ – independent and hence $\{Y_i\}$ are $P$ – independent.

Conversely if $\{Y_i\}$ are $P$ – independent, i.e. $\{\mathcal{C}_i\}$ are $P$ – independent, then

$$\begin{aligned} P\left((Y_1, \dots, Y_n) \in B_1 \times \cdots \times B_n\right) &= P\left(A_1 \cap \cdots \cap A_n\right) \\ &= P\left(A_1\right) \dots P\left(A_n\right) \\ &= P\left(Y_1 \in B_1\right) \cdots P\left(Y_n \in B_n\right) \\ &= \mu_1\left(B_1\right) \cdots \mu\left(B_n\right) \\ &= \mu_1 \otimes \cdots \otimes \mu_n\left(B_1 \times \cdots \times B_n\right). \end{aligned}$$

Since

$$\pi := \{B_1 \times \cdots \times B_n : B_i \in \mathcal{S}_i \text{ for } 1 \leq i \leq n\}$$

is a $\pi$ – system which generates $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$ and

$$\left(Y_1, \dots, Y_n\right)_* P = \mu_1 \otimes \cdots \otimes \mu_n \text{ on } \pi,$$

it follows that $\left(Y_1, \dots, Y_n\right)_* P = \mu_1 \otimes \cdots \otimes \mu_n$ on all of $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$. ■

*Remark 12.5.* When have a collection of not necessarily independent random functions, $Y_i : \Omega \to S_i$, like in Theorem 12.4 it is **not** in general possible to recover the joint distribution, $\pi := \mathrm{Law}_P\left(Y_1, \dots, Y_n\right)$, from the individual distributions, $\mu_i = \mathrm{Law}_P\left(Y_i\right)$ for all $1 \leq i \leq n$. For example suppose that $S_i = \mathbb{R}$ for $i = 1, 2$. $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $(Y_1, Y_2)$ have joint distribution, $\pi$, given by,

$$\pi(C) = \int_{\mathbb{R}} 1_C(x, x)\, d\mu(x) \text{ for all } C \in \mathcal{B}_{\mathbb{R}^2}.$$

If we let $\mu_i = \mathrm{Law}_P\left(Y_i\right)$, then for all $A \in \mathcal{B}_{\mathbb{R}}$ we have

$$\begin{aligned} \mu_1(A) &= P\left(Y_1 \in A\right) = P\left((Y_1, Y_2) \in A \times \mathbb{R}\right) \\ &= \pi(A \times \mathbb{R}) = \int_{\mathbb{R}} 1_{A \times \mathbb{R}}(x, x)\, d\mu(x) = \mu(A). \end{aligned}$$

Similarly we show that $\mu_2 = \mu$. On the other hand if $\mu$ is not concentrated on one point, $\mu \otimes \mu$ is another probability measure on $\left(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}\right)$ with the same **marginals** as $\pi$, i.e. $\pi(A \times \mathbb{R}) = \mu(A) = \pi(\mathbb{R} \times A)$ for all $A \in \mathcal{B}_{\mathbb{R}}$.

**Lemma 12.6.** *Let $(\Omega, , \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ and $\{(T_i, \mathcal{T}_i)\}_{i=1}^n$ be two collection of measurable spaces, $F_i : S_i \to T_i$ be a measurable map for each $i$ and $Y_i : \Omega \to S_i$ be a collection of $P$ – independent measurable maps. Then $\{F_i \circ Y_i\}_{i=1}^n$ are also $P$ – independent.*

**Proof.** Notice that

$$\sigma\left(F_i \circ Y_i\right) = \left(F_i \circ Y_i\right)^{-1}(\mathcal{T}_i) = Y_i^{-1}\left(F_i^{-1}(\mathcal{T}_i)\right) \subset Y_i^{-1}(\mathcal{S}_i) = \mathcal{C}_i.$$

The fact that $\{\sigma\left(F_i \circ Y_i\right)\}_{i=1}^n$ is independent now follows easily from the assumption that $\{\mathcal{C}_i\}$ are $P$ – independent. ■

*Example 12.7.* If $\Omega := \prod_{i=1}^n S_i$, $\mathcal{B} := \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$, $Y_i\left(s_1, \dots, s_n\right) = s_i$ for all $(s_1, \dots, s_n) \in \Omega$, and $\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i)$ for all $i$. Then the probability measures, $P$, on $(\Omega, \mathcal{B})$ for which $\{\mathcal{C}_i\}_{i=1}^n$ are independent are precisely the product measures, $P = \mu_1 \otimes \cdots \otimes \mu_n$ where $\mu_i$ is a probability measure on $(S_i, \mathcal{S}_i)$ for $1 \leq i \leq n$. Notice that in this setting,

$$\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i) = \{S_1 \times \cdots \times S_{i-1} \times B \times S_{i+1} \times \cdots \times S_n : B \in \mathcal{S}_i\} \subset \mathcal{B}.$$

**Proposition 12.8.** *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space and $\{Z_j\}_{j=1}^n$ are independent integrable random variables. Then $\prod_{j=1}^n Z_j$ is also integrable and*

$$\mathbb{E}\left[\prod_{j=1}^n Z_j\right] = \prod_{j=1}^n \mathbb{E} Z_j.$$

**Proof.** Let $\mu_j := P \circ Z_j^{-1} : \mathcal{B}_{\mathbb{R}} \to [0, 1]$ be the law of $Z_j$ for each $j$. Then we know $\left(Z_1, \dots, Z_n\right)_* P = \mu_1 \otimes \cdots \otimes \mu_n$. Therefore by Example 9.52 and Tonelli's theorem,

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^n |Z_j|\right] &= \int_{\mathbb{R}^n} \left[\prod_{j=1}^n |z_j|\right] d\left(\otimes_{j=1}^n \mu_j\right)(z) \\ &= \prod_{j=1}^n \int_{\mathbb{R}^n} |z_j|\, d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E}|Z_j| < \infty \end{aligned}$$

which shows that $\prod_{j=1}^n Z_j$ is integrable. Thus again by Example 9.52 and Fubini's theorem,

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^n Z_j\right] &= \int_{\mathbb{R}^n} \left[\prod_{j=1}^n z_j\right] d\left(\otimes_{j=1}^n \mu_j\right)(z) \\ &= \prod_{j=1}^n \int_{\mathbb{R}} z_j d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E} Z_j. \end{aligned}$$

■

**Theorem 12.9.** *Let $(\Omega, , \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \leq i \leq n$. Further let $\mu_i := P \circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$ and $\pi := P \circ (Y_1, \dots, Y_n)^{-1} : \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$ be the joint distribution of*

$$(Y_1, \dots, Y_n) : \Omega \to S_1 \times \cdots \times S_n.$$

*Then the following are equivalent,*

1. $\{Y_i\}_{i=1}^n$ *are independent,*
2. $\pi = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$
3. *for all bounded measurable functions,* $f : (S_1 \times \cdots \times S_n, \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,
$$\mathbb{E}f(Y_1, \dots, Y_n) = \int_{S_1 \times \cdots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \quad (12.2)$$
*( where the integrals may be taken in any order),*
4. $\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(Y_i)]$ *for all bounded (or non-negative) measurable functions,* $f_i : S_i \to \mathbb{R}$ *or* $\mathbb{C}$.

**Proof.** ($1 \iff 2$) has already been proved in Theorem 12.4. The fact that (2. $\implies$ 3.) now follows from Exercise 9.11 and Fubini's theorem. Similarly, (3. $\implies$ 4.) follows from Exercise 9.11 and Fubini's theorem after taking $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$. Lastly for (4. $\implies$ 1.), let $A_i \in \mathcal{S}_i$ and take $f_i := 1_{A_i}$ in 4. to learn,
$$P\left(\cap_{i=1}^n \{Y_i \in A_i\}\right) = \mathbb{E}\left[\prod_{i=1}^n 1_{A_i}(Y_i)\right] = \prod_{i=1}^n \mathbb{E}[1_{A_i}(Y_i)] = \prod_{i=1}^n P(Y_i \in A_i)$$
which shows that the $\{Y_i\}_{i=1}^n$ are independent. ■

**Corollary 12.10.** *Suppose that* $(\Omega, \mathcal{B}, P)$ *is a probability space and* $\{Y_j : \Omega \to \mathbb{R}\}_{j=1}^n$ *is a sequence of random variables with countable ranges, say* $\Lambda \subset \mathbb{R}$. *Then* $\{Y_j\}_{j=1}^n$ *are independent iff*
$$P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) = \prod_{j=1}^n P(Y_j = y_j) \quad (12.3)$$
*for all choices of* $y_1, \dots, y_n \in \Lambda$.

**Proof.** If the $\{Y_j\}$ are independent then clearly Eq. (12.3) holds by definition as $\{Y_j = y_j\} \in Y_j^{-1}(\mathcal{B}_{\mathbb{R}})$. Conversely if Eq. (12.3) holds and $f_i : \mathbb{R} \to [0, \infty)$ are measurable functions then,
$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] &= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) \\ &= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot \prod_{j=1}^n P(Y_j = y_j) \\ &= \prod_{i=1}^n \sum_{y_i \in \Lambda} f_i(y_i) \cdot P(Y_j = y_j) \\ &= \prod_{i=1}^n \mathbb{E}[f_i(Y_i)]\end{aligned}$$

wherein we have used Tonelli's theorem for sum in the third equality. It now follows that $\{Y_i\}$ are independent using item 4. of Theorem 12.9. ■

**Definition 12.11 (i.i.d.).** *A sequences of random variables,* $\{X_n\}_{n=1}^\infty$, *on a probability space,* $(\Omega, \mathcal{B}, P)$, *are* ***i.i.d. (= independent and identically distributed)*** *if they are independent and* $(X_n)_* P = (X_k)_* P$ *for all* $k, n$. *That is we should have*
$$P(X_n \in A) = P(X_k \in A) \text{ for all } k, n \in \mathbb{N} \text{ and } A \in \mathcal{B}_{\mathbb{R}}.$$

If $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables iff
$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{j=1}^n P(X_i \in A_i) = \prod_{j=1}^n P(X_1 \in A_i) = \prod_{j=1}^n \mu(A_i) \quad (12.4)$$
where $\mu = (X_1)_* P$. The identity in Eq. (12.4) is to hold for all $n \in \mathbb{N}$ and all $A_i \in \mathcal{B}_{\mathbb{R}}$. If we choose $\mu_n = \mu$ in Example 12.16, the $\{Y_n\}_{n=1}^\infty$ there are i.i.d. with $\mathrm{Law}_P(Y_n) = P \circ Y_n^{-1} = \mu$ for all $n \in \mathbb{N}$. We will give another proof of the existence of arbitrary sequences of independent random variables in Proposition 12.20 below which will rely on the following very special case.

**Exercise 12.1.** Suppose that $\Omega = [0,1)$, $\mathcal{B} = \mathcal{B}_{[0,1)}$, and $P = m$ is Lebesgue measure on $\mathcal{B}$. Let $Y_i(\omega) := \omega_i$ be the $i^{th}$ – digit in the base two expansion of $\omega$. To be more precise, $Y_i(\omega) \in \{0,1\}$ is chosen so that
$$\omega = \sum_{i=1}^\infty Y_i(\omega) 2^{-i} \text{ for all } \omega_i \in \{0,1\}.$$

As long as $\omega \neq k2^{-n}$ for some $0 < k \le n$, the above equation uniquely determines the $\{Y_i(\omega)\}$. Owing to the fact that $\sum_{l=n+1}^\infty 2^{-l} = 2^{-n}$, if $\omega = k2^{-n}$, there is some ambiguity in the definitions of the $Y_i(\omega)$ for large $i$ which we resolve by agreeing that $Y_i(\omega) = 1$ for a.a. $i$ is never allowed. Show the random variables, $\{Y_i\}_{i=1}^n$, are i.i.d. for each $n \in \mathbb{N}$ with $P(Y_i = 1) = 1/2 = P(Y_i = 0)$ for all $i$.

**Hint:** the idea is that knowledge of $(Y_1(\omega), \dots, Y_n(\omega))$ is equivalent to knowing for which $k \in \mathbb{N}_0 \cap [0, 2^n)$ that $\omega \in (2^{-n}k, 2^{-n}(k+1))$ and that this knowledge in no way helps you predict the value of $Y_{n+1}(\omega)$. More formally, you might start by showing,
$$P\left(\{Y_{n+1} = 1\} \,|[2^{-n}k, 2^{-n}(k+1))\right) = \frac{1}{2} = P\left(\{Y_{n+1} = 0\} \,|[2^{-n}k, 2^{-n}(k+1))\right).$$

See Section 12.9 if you need more guidancde with this exercise.

**Exercise 12.2.** Let $X, Y$ be two random variables on $(\Omega, \mathcal{B}, P)$.

1. Show that $X$ and $Y$ are independent iff $\operatorname{Cov}(f(X), g(Y)) = 0$ (i.e. $f(X)$ and $g(Y)$ are **uncorrelated**) for bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$.
2. If $X, Y \in L^2(P)$ and $X$ and $Y$ are independent, then $\operatorname{Cov}(X, Y) = 0$.
3. Show by example that if $X, Y \in L^2(P)$ and $\operatorname{Cov}(X, Y) = 0$ does not necessarily imply that $X$ and $Y$ are independent. **Hint:** try taking $(X, Y) = (X, ZX)$ where $X$ and $Z$ are independent simple random variables such that $\mathbb{E}Z = 0$ similar to Remark 11.47.

**Exercise 12.3 (A correlation inequality).** Suppose that $X$ is a random variable and $f, g : \mathbb{R} \to \mathbb{R}$ are two increasing functions such that both $f(X)$ and $g(X)$ are square integrable, i.e. $\mathbb{E}|f(X)|^2 + \mathbb{E}|g(X)|^2 < \infty$. Show $\operatorname{Cov}(f(X), g(X)) \geq 0$. **Hint:** let $Y$ be another random variable which has the same law as $X$ and is independent of $X$. Then consider

$$\mathbb{E}\left[(f(Y) - f(X)) \cdot (g(Y) - g(X))\right].$$

Let us now specialize to the case where $S_i = \mathbb{R}^{m_i}$ and $\mathcal{S}_i = \mathcal{B}_{\mathbb{R}^{m_i}}$ for some $m_i \in \mathbb{N}$.

**Theorem 12.12.** *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $m_j \in \mathbb{N}$, $S_j = \mathbb{R}^{m_j}$, $\mathcal{S}_j = \mathcal{B}_{\mathbb{R}^{m_j}}$, $Y_j : \Omega \to S_j$ be random vectors, and $\mu_j := \operatorname{Law}_P(Y_j) = P \circ Y_j^{-1} : \mathcal{S}_j \to [0,1]$ for $1 \leq j \leq n$. The the following are equivalent;*

1. *$\{Y_j\}_{j=1}^n$ are independent,*
2. *$\operatorname{Law}_P(Y_1, \dots, Y_n) = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$*
3. *for all bounded measurable functions, $f : (S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,*
$$\mathbb{E}f(Y_1, \dots, Y_n) = \int_{S_1 \times \dots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \quad (12.5)$$
*( where the integrals may be taken in any order),*
4. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}[f_j(Y_j)]$ for all bounded (or non-negative) measurable functions, $f_j : S_j \to \mathbb{R}$ or $\mathbb{C}$.*
5. *$P\left(\cap_{j=1}^n \{Y_j \leq y_j\}\right) = \prod_{j=1}^n P(\{Y_j \leq y_j\})$ for all $y_j \in S_j$, where we say that $Y_j \leq y_j$ iff $(Y_j)_k \leq (y_j)_k$ for $1 \leq k \leq m_j$.*
6. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}[f_j(Y_j)]$ for all $f_j \in C_c(S_j, \mathbb{R})$,*
7. *$\mathbb{E}\left[e^{i\sum_{j=1}^n \lambda_j \cdot Y_j}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{i\lambda_j \cdot Y_j}\right]$ for all $\lambda_j \in S_j = \mathbb{R}^{m_j}$.*

**Proof.** The equivalence of 1. – 4. has already been proved in Theorem 12.9. It is also clear that item 4. implies any of the items 5. –7. upon noting that item 5. may be written as,

$$\mathbb{E}\left[\prod_{j=1}^n 1_{(-\infty, y_j]}(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[1_{(-\infty, y_j]}(Y_j)\right]$$

where

$$(-\infty, y_j] := (-\infty, (y_j)_1] \times \dots \times (-\infty, (y_j)_{m_j}].$$

The proofs that either 5. or 6. or 7. implies item 3. is a simple application of the multiplicative system theorem in the form of either Corollary 10.3 or Corollary 10.10. In each case, let $\mathbb{H}$ denote the linear space of bounded measurable functions such that Eq. (12.5) holds. To complete the proof I will simply give you the multiplicative system, $\mathbb{M}$, to use in each of the cases. To describe $\mathbb{M}$, let $N = m_1 + \dots + m_n$ and

$$y = (y_1, \dots, y_n) = \left(y^1, y^2, \dots, y^N\right) \in \mathbb{R}^N \text{ and}$$
$$\lambda = (\lambda_1, \dots, \lambda_n) = \left(\lambda^1, \lambda^2, \dots, \lambda^N\right) \in \mathbb{R}^N$$

For showing 5. $\implies$ 3.take $\mathbb{M} = \left\{1_{(-\infty, y]} : y \in \mathbb{R}^N\right\}$.

For showing 6. $\implies$ 3. take $\mathbb{M}$ to be a those functions on $\mathbb{R}^N$ which are of the form, $f(y) = \prod_{l=1}^N f_l\left(y^l\right)$ with each $f_l \in C_c(\mathbb{R})$.

For showing 7. $\implies$ 3. take $\mathbb{M}$ to be the functions of the form,

$$f(y) = \exp\left(i \sum_{j=1}^n \lambda_j \cdot y_j\right) = \exp(i\lambda \cdot y).$$

**Alternatively** we could show 7. $\implies$ 2. as follows. Let $N = m_1 + \dots + m_n$, $\pi := \operatorname{Law}(Y_1, \dots, Y_n)$ and $\lambda := (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^N$. Item 7. states,

$$\begin{aligned}\hat{\pi}(\lambda) &:= \int_{\mathbb{R}^N} e^{i\lambda \cdot y} d\pi(y) = \mathbb{E}\left[e^{i\sum_{j=1}^n \lambda_j \cdot Y_j}\right] \\ &= \prod_{j=1}^n \mathbb{E}\left[e^{i\lambda_j \cdot Y_j}\right] = \prod_{j=1}^n \int_{\mathbb{R}^{m_j}} e^{i\lambda_j \cdot y_j} d\mu_j(y_j) \\ &= (\mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n)^{\hat{}}(\lambda)\end{aligned}$$

and so $\pi = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$ by Corollary 10.13. ■

**Definition 12.13.** *A collection of subsets of $\mathcal{B}$, $\{\mathcal{C}_t\}_{t \in T}$ is said to be independent iff $\{\mathcal{C}_t\}_{t \in \Lambda}$ are independent for all finite subsets, $\Lambda \subset T$. More explicitly, we are requiring*

$$P(\cap_{t \in \Lambda} A_t) = \prod_{t \in \Lambda} P(A_t)$$

*whenever $\Lambda$ is a finite subset of $T$ and $A_t \in \mathcal{C}_t$ for all $t \in \Lambda$.*

**Corollary 12.14.** *If $\{\mathcal{C}_t\}_{t\in T}$ is a collection of independent classes such that each $\mathcal{C}_t$ is a $\pi$ – system, then $\{\sigma(\mathcal{C}_t)\}_{t\in T}$ are independent as well.*

**Definition 12.15.** *A collections of random variables, $\{X_t : t \in T\}$ are* ***independent*** *iff $\{\sigma(X_t) : t \in T\}$ are independent.*

*Example 12.16.* Suppose that $\{\mu_n\}_{n=1}^\infty$ is any sequence of probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Let $\Omega = \mathbb{R}^\mathbb{N}$, $\mathcal{B} := \otimes_{n=1}^\infty \mathcal{B}_\mathbb{R}$ be the product $\sigma$ – algebra on $\Omega$, and $P := \otimes_{n=1}^\infty \mu_n$ be the product measure. Then the random variables, $\{Y_n\}_{n=1}^\infty$ defined by $Y_n(\omega) = \omega_n$ for all $\omega \in \Omega$ are independent with $\mathrm{Law}_P(Y_n) = \mu_n$ for each $n$.

**Lemma 12.17 (Independence of groupings).** *Suppose that $\{\mathcal{B}_t : t \in T\}$ is an independent family of $\sigma$ – fields. Suppose further that $\{T_s\}_{s\in S}$ is a partition of $T$ (i.e. $T = \sum_{s\in S} T_s$) and let*

$$\mathcal{B}_{T_s} = \vee_{t\in T_s}\mathcal{B}_t := \sigma(\cup_{t\in T_s}\mathcal{B}_t).$$

*Then $\{\mathcal{B}_{T_s}\}_{s\in S}$ is again independent family of $\sigma$ fields.*

**Proof.** Let

$$\mathcal{C}_s = \{\cap_{\alpha\in K} B_\alpha : B_\alpha \in \mathcal{B}_\alpha,\ K \subset_f T_s\}.$$

It is now easily checked that $\mathcal{B}_{T_s} = \sigma(\mathcal{C}_s)$ and that $\{\mathcal{C}_s\}_{s\in S}$ is an independent family of $\pi$ – systems. Therefore $\{\mathcal{B}_{T_s}\}_{s\in S}$ is an independent family of $\sigma$ – algebras by Corollary 12.14. ■

*Remark 12.18.* To better understand the last proof it is instructive to write out a special case in more detail. Suppose that $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$ are independent sub-$\sigma$– algebras relative to $(\Omega, \mathcal{F}, P)$. This independence implies

$$\begin{aligned} P(A\cap B\cap C\cap D\cap E) &= P(A)P(B)P(C)P(D)P(E) \\ &= P(A\cap B)P(C\cap D\cap E) \end{aligned}$$

for all $A \in \mathcal{A}$, $B \in \mathcal{B}$, $C \in \mathcal{C}$, $D \in \mathcal{D}$, and $E \in \mathcal{E}$. Therefore,

$$\begin{aligned} \mathcal{C}_1 &:= \{A\cap B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\} \text{ and} \\ \mathcal{C}_2 &:= \{C\cap D\cap E : C \in \mathcal{C},\ D \in \mathcal{D}, \text{ and } E \in \mathcal{E}\} \end{aligned}$$

are independent. As $\mathcal{C}_1$ and $\mathcal{C}_2$ are multiplicative we may conclude from Theorem 12.2 that $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{C}_1)$ and $\mathcal{C} \vee \mathcal{D} \vee \mathcal{E} = \sigma(\mathcal{C}_2)$ are independent as well.

**Corollary 12.19.** *Suppose that $\{Y_n\}_{n=1}^\infty$ is a sequence of independent random variables (or vectors) and $\Lambda_1, \dots, \Lambda_m$ is a collection of pairwise disjoint subsets of $\mathbb{N}$. Further suppose that $f_i : \mathbb{R}^{\Lambda_i} \to \mathbb{R}$ is a measurable function for each $1 \le i \le m$, then $Z_i := f_i(\{Y_l\}_{l\in\Lambda_i})$ is again a collection of independent random variables.*

**Proof.** Notice that $\sigma(Z_i) \subset \sigma(\{Y_l\}_{l\in\Lambda_i}) = \sigma(\cup_{l\in\Lambda_i}\sigma(Y_l))$. Since $\{\sigma(Y_l)\}_{l=1}^\infty$ are independent by assumption, it follows from Lemma 12.17 that $\{\sigma(\{Y_l\}_{l\in\Lambda_i})\}_{i=1}^m$ are independent and therefore so is $\{\sigma(Z_i)\}_{i=1}^m$, i.e. $\{Z_i\}_{i=1}^m$ are independent. ■

**Proposition 12.20.** *Given any sequence, $\{\mu_n\}_{n=1}^\infty$, of probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, there exists a independent sequence, $\{X_n\}_{n=1}^\infty$, of random variables on $([0,1), \mathcal{B}_{[0,1)}, P = m)$ such that $\mathrm{Law}_P(X_n) = \mu_n$ for all $n$.*

**Proof.** Let $\{Y_n\}_{n=1}^\infty$ be i.i.d. Bernoulli random variables on $([0,1), \mathcal{B}_{[0,1)}, P = m)$ as in Exercise 12.1. Let $\varphi : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ be a bijection and set $Z_{l,n} := Y_{\varphi(l,n)}$ so that $\{Z_{l,n}\}_{l,n=1}^\infty$ are i.i.d. Bernoulli random variables on $([0,1), \mathcal{B}_{[0,1)}, P = m)$. For each $n \in \mathbb{N}$, let $U_n := \sum_{l=1}^\infty Z_{l,n}2^{-l}$. According to Corollary 12.19, $\{U_n\}_{n=1}^\infty$ are independent random variables. Moreover the reader should verify that $\mathrm{Law}_P(U_n) = m$ for all $n$ so that $\{U_n\}_{n=1}^\infty$ are i.i.d. uniformly distributed random variables on $([0,1), \mathcal{B}_{[0,1)}, P = m)$. We now choose measurable functions, $G_n : (0,1) \to \mathbb{R}$ as in Theorem 8.48 so that $\mathrm{Law}_P(G_n(U_n)) = \mu_n$ for each $n$. Then $\{X_n := G_n(U_n)\}_{n=1}^\infty$ is the desired collection of independent random variables on $([0,1), \mathcal{B}_{[0,1)}, P = m)$. [See Section 12.9 for more details on this type of construction.] ■

The following theorem follows immediately from the definitions and Theorem 12.12.

**Theorem 12.21.** *Let $\mathbb{X} := \{X_t : t \in T\}$ be a collection of random variables. Then the following are equivalent:*

1. *The collection $\mathbb{X}$ is independent,*
2. $$P(\cap_{t\in\Lambda}\{X_t \in A_t\}) = \prod_{t\in\Lambda} P(X_t \in A_t)$$
   *for all finite subsets, $\Lambda \subset T$, and all $\{A_t\}_{t\in\Lambda} \subset \mathcal{B}_\mathbb{R}$.*
3. $$P(\cap_{t\in\Lambda}\{X_t \le x_t\}) = \prod_{t\in\Lambda} P(X_t \le x_t)$$
   *for all finite subsets, $\Lambda \subset T$, and all $\{x_t\}_{t\in\Lambda} \subset \mathbb{R}$.*
4. *For all $\Gamma \subset_f T$ and $f_t : \mathbb{R}^n \to \mathbb{R}$ which are bounded an measurable for all $t \in \Gamma$,*
   $$\mathbb{E}\left[\prod_{t\in\Gamma} f_t(X_t)\right] = \prod_{t\in\Gamma}\mathbb{E}f_t(X_t) = \int_{\mathbb{R}^\Gamma}\prod_{t\in\Gamma} f_t(x_t)\prod_{t\in\Gamma} d\mu_t(x_t).$$
5. $\mathbb{E}\left[\prod_{t\in\Gamma}\exp\left(e^{i\lambda_t\cdot X_t}\right)\right] = \prod_{t\in\Gamma}\hat{\mu}_t(\lambda)$.

6. *For all $\Gamma \subset_f T$ and $f : (\mathbb{R}^n)^\Gamma \to \mathbb{R}$,*

$$\mathbb{E}\left[f\left(X_\Gamma\right)\right] = \int_{(\mathbb{R}^n)^\Gamma} f\left(x\right) \prod_{t\in\Gamma} d\mu_t\left(x_t\right).$$

7. *For all $\Gamma \subset_f T$, $\operatorname{Law}_P\left(X_\Gamma\right) = \otimes_{t\in\Gamma}\mu_t$.*
8. $\operatorname{Law}_P\left(X\right) = \otimes_{t\in T}\mu_t$.

*Moreover, if $\mathcal{B}_t$ is a sub-$\sigma$ - algebra of $\mathcal{B}$ for $t \in T$, then $\{\mathcal{B}_t\}_{t\in T}$ are independent iff for all $\Gamma \subset_f T$,*

$$\mathbb{E}\left[\prod_{t\in\Gamma} X_t\right] = \prod_{t\in\Gamma} \mathbb{E}X_t \textit{ for all } X_t \in L^\infty\left(\Omega, \mathcal{B}_t, P\right).$$

**Proof.** The equivalence of 1. and 2. follows almost immediately form the definition of independence and the fact that $\sigma\left(X_t\right) = \{\{X_t \in A\} : A \in \mathcal{B}_\mathbb{R}\}$. Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 12.14 with $\mathcal{C}_t := \{\{X_t \le a\} : a \in \mathbb{R}\}$ and making use the observations that $\mathcal{C}_t$ is a $\pi$ – system for all $t$ and that $\sigma\left(\mathcal{C}_t\right) = \sigma\left(X_t\right)$. The remaining equivalence are also easy to check. ■

## 12.2 Examples of Independence

### 12.2.1 An Example of Ranks

**Lemma 12.22 (No Ties).** *Suppose that $X$ and $Y$ are independent random variables on a probability space $(\Omega, \mathcal{B}, P)$. If $F(x) := P(X \le x)$ is continuous, then $P(X = Y) = 0$.*

**Proof.** Let $\mu(A) := P(X \in A)$ and $\nu(A) = P(Y \in A)$. Because $F$ is continuous, $\mu(\{y\}) = F(y) - F(y-) = 0$, and hence

$$\begin{aligned} P\left(X = Y\right) &= \mathbb{E}\left[1_{\{X=Y\}}\right] = \int_{\mathbb{R}^2} 1_{\{x=y\}} d\left(\mu\otimes\nu\right)\left(x, y\right) \\ &= \int_\mathbb{R} d\nu\left(y\right) \int_\mathbb{R} d\mu\left(x\right) 1_{\{x=y\}} = \int_\mathbb{R} \mu\left(\{y\}\right) d\nu\left(y\right) \\ &= \int_\mathbb{R} 0 \; d\nu\left(y\right) = 0. \end{aligned}$$

**Second Proof.** For sake of comparison, lets give a proof where we do not allow ourselves to use Fubini's theorem. To this end let $\left\{a_l := \frac{l}{N}\right\}_{l=-\infty}^\infty$ (or for the moment any sequence such that, $a_l < a_{l+1}$ for all $l \in \mathbb{Z}$, $\lim_{l\to\pm\infty} a_l = \pm\infty$). Then

$$\{(x, x) : x \in \mathbb{R}\} \subset \cup_{l\in\mathbb{Z}} \left[(a_l, a_{l+1}] \times (a_l, a_{l+1}]\right]$$

and therefore,

$$\begin{aligned} P\left(X = Y\right) &\le \sum_{l\in\mathbb{Z}} P\left(X \in (a_l, a_{l+1}], Y \in (a_l, a_{l+1}]\right) = \sum_{l\in\mathbb{Z}} \left[F\left(a_{l+1}\right) - F\left(a_l\right)\right]^2 \\ &\le \sup_{l\in\mathbb{Z}} \left[F\left(a_{l+1}\right) - F\left(a_l\right)\right] \sum_{l\in\mathbb{Z}} \left[F\left(a_{l+1}\right) - F\left(a_l\right)\right] = \sup_{l\in\mathbb{Z}} \left[F\left(a_{l+1}\right) - F\left(a_l\right)\right]. \end{aligned}$$

Since $F$ is continuous and $F(\infty+) = 1$ and $F(\infty-) = 0$, it is easily seen that $F$ is uniformly continuous on $\mathbb{R}$. Therefore, if we choose $a_l = \frac{l}{N}$, we have

$$P\left(X = Y\right) \le \limsup_{N\to\infty} \sup_{l\in\mathbb{Z}} \left[F\left(\frac{l+1}{N}\right) - F\left(\frac{l}{N}\right)\right] = 0.$$

■

Let $\{X_n\}_{n=1}^\infty$ be i.i.d. with common continuous distribution function, $F$. So by Lemma 12.22 we know that

$$P\left(X_i = X_j\right) = 0 \text{ for all } i \neq j.$$

Let $R_n$ denote the "**rank**" of $X_n$ in the list $(X_1, \dots, X_n)$, i.e.

$$R_n := \sum_{j=1}^n 1_{X_j \ge X_n} = \#\left\{j \le n : X_j \ge X_n\right\}.$$

Thus $R_n = k$ if $X_n$ is the $k^{\text{th}}$ – largest element in the list, $(X_1, \dots, X_n)$. For example if $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, \dots) = (9, -8, 3, 7, 23, 0, -11, \dots)$, we have $R_1 = 1$, $R_2 = 2$, $R_3 = 2$, $R_4 = 2$, $R_5 = 1$, $R_6 = 5$, and $R_7 = 7$. Observe that rank order, from lowest to highest, of $(X_1, X_2, X_3, X_4, X_5)$ is $(X_2, X_3, X_4, X_1, X_5)$. This can be determined by the values of $R_i$ for $i = 1, 2, \dots, 5$ as follows. Since $R_5 = 1$, we must have $X_5$ in the last slot, i.e. $(*, *, *, *, X_5)$. Since $R_4 = 2$, we know out of the remaining slots, $X_4$ must be in the second from the far most right, i.e. $(*, *, X_4, *, X_5)$. Since $R_3 = 2$, we know that $X_3$ is again the second from the right of the remaining slots, i.e. we now know, $(*, X_3, X_4, *, X_5)$. Similarly, $R_2 = 2$ implies $(X_2, X_3, X_4, *, X_5)$ and finally $R_1 = 1$ gives, $(X_2, X_3, X_4, X_1, X_5)$ $(= (-8, 4, 7, 9, 23)$ in the example). As another example, if $R_i = i$ for $i = 1, 2, \dots, n$, then $X_n < X_{n-1} < \cdots < X_1$.

**Theorem 12.23 (Renyi Theorem).** *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. and assume that $F(x) := P(X_n \le x)$ is continuous. Then $\{R_n\}_{n=1}^\infty$ is an independent sequence,*

$$P\left(R_n = k\right) = \frac{1}{n} \textit{ for } k = 1, 2, \dots, n,$$

*and the events, $A_n = \{X_n \text{ is a record}\} = \{R_n = 1\}$ are independent as $n$ varies and*

$$P\left(A_n\right) = P\left(R_n = 1\right) = \frac{1}{n}.$$

**Proof.** By Problem 6 on p. 110 of Resnick or by Fubini's theorem, $(X_1,\dots,X_n)$ and $(X_{\sigma 1},\dots,X_{\sigma n})$ have the same distribution for any permutation $\sigma$.

Since $F$ is continuous, it now follows that up to a set of measure zero,

$$\Omega = \sum_\sigma \{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}$$

and therefore

$$1 = P(\Omega) = \sum_\sigma P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}).$$

Since $P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\})$ is independent of $\sigma$ we may now conclude that

$$P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}) = \frac{1}{n!}$$

for all $\sigma$. As observed before the statement of the theorem, to each realization $(\varepsilon_1,\dots,\varepsilon_n)$, (here $\varepsilon_i \in \mathbb{N}$ with $\varepsilon_i \le i$) of $(R_1,\dots,R_n)$ there is a uniquely determined permutation, $\sigma = \sigma(\varepsilon_1,\dots,\varepsilon_n)$, such that $X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}$. (Notice that there are $n!$ permutations of $\{1,2,\dots,n\}$ and there are also $n!$ choices for the $\{(\varepsilon_1,\dots,\varepsilon_n) : 1 \le \varepsilon_i \le i\}$.) From this it follows that

$$\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\} = \{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}$$

and therefore,

$$P(\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\}) = P(X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}) = \frac{1}{n!}.$$

Since

$$\begin{aligned} P(\{R_n = \varepsilon_n\}) &= \sum_{(\varepsilon_1,\dots\varepsilon_{n-1})} P(\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\}) \\ &= \sum_{(\varepsilon_1,\dots\varepsilon_{n-1})} \frac{1}{n!} = (n-1)! \cdot \frac{1}{n!} = \frac{1}{n} \end{aligned}$$

we have shown that

$$P(\{(R_1,\dots,R_n) = (\varepsilon_1,\dots,\varepsilon_n)\}) = \frac{1}{n!} = \prod_{j=1}^n \frac{1}{j} = \prod_{j=1}^n P(\{R_j = \varepsilon_j\}).$$

■



*Remark 12.24.* Renyi's Theorem is really about random permutations with the uniform distribution. The point is that to each realization, $(X_1,\dots,X_n)$, there is a permutation $\sigma$ such that $(X_{\sigma 1},\dots,X_{\sigma n})$ is listed in increasing order. What we showed above is that $\sigma$ is a random permutation with the uniform distribution. We then defined

$$R_k(\sigma) := \text{rank of } \sigma(k) \text{ in the list } (\sigma(1),\dots,\sigma(k)).$$

Thus if $n = 5$ and $(\sigma(1),\dots,\sigma(5)) = (3,5,2,1,4)$ we have $R_1(\sigma) = 1$, $R_2(\sigma) = 1$, $R_3(\sigma) = 3$, $R_4(\sigma) = 4$ and $R_5(\sigma) = 2$. We then showed that given these ranks we could recover $\sigma$, for example

$$\begin{aligned} R_5(\sigma) = 2 &\implies (\sigma(1),\dots,\sigma(5)) = (*,*,*,*,4) \\ R_4(\sigma) = 4 &\implies (\sigma(1),\dots,\sigma(5)) = (*,*,*,1,4) \\ R_3(\sigma) = 3 &\implies (\sigma(1),\dots,\sigma(5)) = (*,*,2,1,4) \\ R_2(\sigma) = 1 &\implies (\sigma(1),\dots,\sigma(5)) = (*,5,2,1,4) \\ R_1(\sigma) = 1 &\implies (\sigma(1),\dots,\sigma(5)) = (3,5,2,1,4). \end{aligned}$$

Thus it follows that

$$P(R_i = \varepsilon_i : 1 \le i \le n) = P(\sigma = \tau(\varepsilon_1,\dots,\varepsilon_n)) = \frac{1}{n!}.$$

Further observe that

$$P(R_k = \varepsilon_k) = \frac{1}{n!} \cdot (k-1)! \cdot n \cdot (n-1)\dots(n-k+1) = \frac{1}{k}.$$

Therefore we see that

$$P(R_i = \varepsilon_i : 1 \le i \le n) = \prod_{k=1}^n P(R_k = \varepsilon_k).$$

## 12.3 Independence for Gaussian Random Vectors

As you saw in Exercise 12.2, uncorrelated random variables are typically not independent. However, if the random variables involved are jointly Gaussian (see Definition 11.40), then independence and uncorrelated are actually the same thing!

**Lemma 12.25.** *Suppose that $Z = (X,Y)^{\mathrm{tr}}$ is a Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$. Then $X$ is independent of $Y$ iff* $\operatorname{Cov}(X_i,Y_j) = 0$ *for all $1 \le i \le k$ and $1 \le j \le l$. This lemma also holds more generally. Namely if $\{X^l\}_{l=1}^n$ is a sequence of random vectors such that $(X^1,\dots,X^n)$ is a Gaussian random vector. Then $\{X^l\}_{l=1}^n$ are independent iff* $\operatorname{Cov}\left(X_i^l, X_k^{l'}\right) = 0$ *for all $l \ne l'$ and $i$ and $k$.*

**Proof.** We know by Exercise 12.2 that if $X_i$ and $Y_j$ are independent, then $\operatorname{Cov}(X_i, Y_j) = 0$. For the converse direction, if $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$ and $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned}\operatorname{Var}(x \cdot X + y \cdot Y) &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y) + 2\operatorname{Cov}(x \cdot X, y \cdot Y) \\ &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y).\end{aligned}$$

Therefore using the fact that $(X, Y)$ is a Gaussian random vector,

$$\begin{aligned}\mathbb{E}\left[e^{ix\cdot X}e^{iy\cdot Y}\right] &= \mathbb{E}\left[e^{i(x\cdot X + y\cdot Y)}\right] \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x \cdot X + y \cdot Y) + \mathbb{E}(x \cdot X + y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X) - \frac{1}{2}\operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right],\end{aligned}$$

and because $x$ and $y$ were arbitrary, we may conclude from Theorem 12.12 that $X$ and $Y$ are independent. ■

**Corollary 12.26.** *Suppose that $X : \Omega \to \mathbb{R}^k$ and $Y : \Omega \to \mathbb{R}^l$ are two independent random Gaussian vectors, then $(X, Y)$ is also a Gaussian random vector. This corollary generalizes to multiple independent random Gaussian vectors.*

**Proof.** Let $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned}\mathbb{E}\left[e^{i(x,y)\cdot(X,Y)}\right] &= \mathbb{E}\left[e^{i(x\cdot X + y\cdot Y)}\right] = \mathbb{E}\left[e^{ix\cdot X}e^{iy\cdot Y}\right] = \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right] \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X)\right) \\ &\qquad \times \exp\left(-\frac{1}{2}\operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X) - \frac{1}{2}\operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2}\operatorname{Var}(x \cdot X + y \cdot Y) + i\mathbb{E}(x \cdot X + y \cdot Y)\right)\end{aligned}$$

which shows that $(X, Y)$ is again Gaussian. ■

**Notation 12.27** *Suppose that $\{X_i\}_{i=1}^n$ is a collection of $\mathbb{R}$ – valued variables or $\mathbb{R}^d$ – valued random vectors. We will write $X_1 \overset{\perp\perp}{+} X_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} X_n$ for $X_1 + \dots + X_n$ under the additional assumption that the $\{X_i\}_{i=1}^n$ are independent.*

**Corollary 12.28.** *Suppose that $\{X_i\}_{i=1}^n$ are independent Gaussian random variables, then $S_n := \sum_{i=1}^n X_i$ is a Gaussian random variables with :*

$$\operatorname{Var}(S_n) = \sum_{i=1}^n \operatorname{Var}(X_i) \text{ and } \mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i, \tag{12.6}$$

*i.e.*

$$X_1 \overset{\perp\perp}{+} X_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} X_n \overset{d}{=} N\left(\sum_{i=1}^n \operatorname{Var}(X_i), \sum_{i=1}^n \mathbb{E}X_i\right).$$

*In particular if $\{X_i\}_{i=1}^\infty$ are i.i.d. Gaussian random variables with $\mathbb{E}X_i = \mu$ and $\sigma^2 = \operatorname{Var}(X_i)$, then*

$$\frac{S_n}{n} - \mu \overset{d}{=} N\left(0, \frac{\sigma^2}{n}\right) \text{ and } \tag{12.7}$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{d}{=} N(0, 1). \tag{12.8}$$

*Equation (12.8) is a very special case of the central limit theorem while Eq. (12.7) leads to a very special case of the strong law of large numbers, see Corollary 12.29.*

**Proof.** The fact that $S_n$, $\frac{S_n}{n} - \mu$, and $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ are all Gaussian follows from Corollary 12.28 and Lemma 11.42 or by direct calculation. The formulas for the variances and means of these random variables are routine to compute. ■

Recall the first Borel Cantelli-Lemma 9.14 states that if $\{A_n\}_{n=1}^\infty$ are measurable sets, then

$$\sum_{n=1}^\infty P(A_n) < \infty \implies P(\{A_n \text{ i.o.}\}) = 0. \tag{12.9}$$

**Corollary 12.29.** *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. Gaussian random variables with $\mathbb{E}X_i = \mu$ and $\sigma^2 = \operatorname{Var}(X_i)$. Then $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s. and moreover for every $\alpha < \frac{1}{2}$, there exists $N_\alpha : \Omega \to \mathbb{N} \cup \{\infty\}$, such that $P(N_\alpha = \infty) = 0$ and*

$$\left|\frac{S_n}{n} - \mu\right| \le n^{-\alpha} \text{ for } n \ge N_\alpha.$$

*In particular, $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s.*

**Proof.** Let $Z \overset{d}{=} N(0, 1)$ so that $\frac{\sigma}{\sqrt{n}} Z \overset{d}{=} N\left(0, \frac{\sigma^2}{n}\right)$. From the Eq. (12.7) and Eq. (9.50),

$$P\left(\left|\frac{S_n}{n}-\mu\right|\geq\varepsilon\right)=P\left(\left|\frac{\sigma}{\sqrt{n}}Z\right|\geq\varepsilon\right)=P\left(|Z|\geq\frac{\sqrt{n}\varepsilon}{\sigma}\right)$$
$$\leq\exp\left(-\frac{1}{2}\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)^2\right)=\exp\left(-\frac{\varepsilon^2}{2\sigma^2}n\right).$$

Taking $\varepsilon=n^{-\alpha}$ with $1-2\alpha>0$, it follows that

$$\sum_{n=1}^{\infty}P\left(\left|\frac{S_n}{n}-\mu\right|\geq n^{-\alpha}\right)\leq\sum_{n=1}^{\infty}\exp\left(-\frac{1}{2\sigma^2}n^{1-2\alpha}\right)<\infty$$

and so by the first Borel-Cantelli lemma,

$$P\left(\left\{\left|\frac{S_n}{n}-\mu\right|\geq n^{-\alpha}\text{ i.o.}\right\}\right)=0.$$

Therefore, $P$ – a.s., $\left|\frac{S_n}{n}-\mu\right|\leq n^{-\alpha}$ a.a., and in particular $\lim_{n\to\infty}\frac{S_n}{n}=\mu$ a.s. ■

## 12.4 Summing independent random variables

**Exercise 12.4.** Suppose that $X\overset{d}{=}N\left(0,a^2\right)$ and $Y\overset{d}{=}N\left(0,b^2\right)$ and $X$ and $Y$ are independent. Show by direct computation using the formulas for the distributions of $X$ and $Y$ that $X+Y=N\left(0,a^2+b^2\right)$.

**Exercise 12.5.** Show that the sum, $N_1+N_2$, of two independent Poisson random variables, $N_1$ and $N_2$, with parameters $\lambda_1$ and $\lambda_2$ respectively is again a Poisson random variable with parameter $\lambda_1+\lambda_2$. (You could use generating functions or do this by hand.) In short $\mathrm{Poi}(\lambda_1)\overset{\perp\perp}{+}\mathrm{Poi}(\lambda_2)\overset{d}{=}\mathrm{Poi}(\lambda_1+\lambda_2)$.

*Example 12.30 (Gamma Distribution Sums).* We will show here that $\mathrm{Gamma}(k,\theta)\overset{\perp\perp}{+}\mathrm{Gamma}(l,\theta)=\mathrm{Gamma}(k+l,\theta)$. In Exercise 9.13 you showed if $k,\theta>0$ then

$$\mathbb{E}\left[e^{tX}\right]=(1-\theta t)^{-k}\text{ for }t<\theta^{-1}$$

where $X$ is a positive random variable with $X\overset{d}{=}\mathrm{Gamma}(k,\theta)$, i.e.

$$(X_*P)(dx)=x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}dx\text{ for }x>0.$$

Suppose that $X$ and $Y$ are independent Random variables with $X\overset{d}{=}\mathrm{Gamma}(k,\theta)$ and $Y\overset{d}{=}\mathrm{Gamma}(l,\theta)$ for some $l>0$. It now follows that



$$\mathbb{E}\left[e^{t(X+Y)}\right]=\mathbb{E}\left[e^{tX}e^{tY}\right]=\mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right]$$
$$=(1-\theta t)^{-k}(1-\theta t)^{-l}=(1-\theta t)^{-(k+l)}.$$

Therefore it follows from Exercise 10.3 that $X+Y\overset{d}{=}\mathrm{Gamma}(k+l,\theta)$.

*Example 12.31 (Exponential Distribution Sums).* If $\{T_k\}_{k=1}^n$ are independent random variables such that $T_k\overset{d}{=}E(\lambda_k)$ for all $k$, then

$$T_1\overset{\perp\perp}{+}T_2\overset{\perp\perp}{+}\dots\overset{\perp\perp}{+}T_n=\mathrm{Gamma}\left(n,\lambda^{-1}\right).$$

This follows directly from Example 12.30 using $E(\lambda)=\mathrm{Gamma}\left(1,\lambda^{-1}\right)$ and induction. We will verify this directly later on in Corollary 13.9.

Example 12.30 may also be verified using brute force. To this end, suppose that $f:\mathbb{R}_+\to\mathbb{R}_+$ is a measurable function, then

$$\mathbb{E}[f(X+Y)]=\int_{\mathbb{R}_+^2}f(x+y)\,x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}y^{l-1}\frac{e^{-y/\theta}}{\theta^l\Gamma(l)}dxdy$$
$$=\frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)}\int_{\mathbb{R}_+^2}f(x+y)\,x^{k-1}y^{l-1}e^{-(x+y)/\theta}dxdy.$$

Let us now make the change of variables, $x=x$ and $z=x+y$, so that $dxdy=dxdz$, to find,

$$\mathbb{E}[f(X+Y)]=\frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)}\int 1_{0\leq x\leq z<\infty}f(z)\,x^{k-1}(z-x)^{l-1}e^{-z/\theta}dxdz. \tag{12.10}$$

To finish the proof we must now do that $x$ integral and show,

$$\int_0^z x^{k-1}(z-x)^{l-1}dx=z^{k+l-1}\frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}.$$

(In fact we already know this must be correct from our Laplace transform computations above.) First make the change of variable, $x=zt$ to find,

$$\int_0^z x^{k-1}(z-x)^{l-1}dx=z^{k+l-1}B(k,l)$$

where $B(k,l)$ is the **beta − function** defined by;

$$B(k,l):=\int_0^1 t^{k-1}(1-t)^{l-1}dt\text{ for }\operatorname{Re}k,\operatorname{Re}l>0. \tag{12.11}$$

Combining these results with Eq. (12.10) then shows,

$$\mathbb{E}\left[f\left(X+Y\right)\right]=\frac{B\left(k,l\right)}{\theta^{k+l}\Gamma\left(k\right)\Gamma\left(l\right)}\int_0^\infty f\left(z\right)z^{k+l-1}e^{-z/\theta}dz. \tag{12.12}$$

Since we already know that

$$\int_0^\infty z^{k+l-1}e^{-z/\theta}dz=\theta^{k+l}\Gamma\left(k+l\right)$$

it follows by taking $f=1$ in Eq. (12.12) that

$$1=\frac{B\left(k,l\right)}{\theta^{k+l}\Gamma\left(k\right)\Gamma\left(l\right)}\theta^{k+l}\Gamma\left(k+l\right)$$

which implies,

$$B\left(k,l\right)=\frac{\Gamma\left(k\right)\Gamma\left(l\right)}{\Gamma\left(k+l\right)}. \tag{12.13}$$

Therefore, using this back in Eq. (12.12) implies

$$\mathbb{E}\left[f\left(X+Y\right)\right]=\frac{1}{\theta^{k+l}\Gamma\left(k+l\right)}\int_0^\infty f\left(z\right)z^{k+l-1}e^{-z/\theta}dz$$

from which it follows that $X+Y\overset{d}{=}\text{Gamma}(k+l,\theta)$.

Let us pause to give a direct verification of Eq. (12.13). By definition of the gamma function,

$$\begin{aligned}\Gamma\left(k\right)\Gamma\left(l\right)&=\int_{\mathbb{R}_+^2}x^{k-1}e^{-x}y^{l-1}e^{-y}dxdy=\int_{\mathbb{R}_+^2}x^{k-1}y^{l-1}e^{-(x+y)}dxdy.\\&=\int_{0\le x\le z<\infty}x^{k-1}\left(z-x\right)^{l-1}e^{-z}dxdz\end{aligned}$$

Making the change of variables, $x=x$ and $z=x+y$ it follows,

$$\Gamma\left(k\right)\Gamma\left(l\right)=\int_{0\le x\le z<\infty}x^{k-1}\left(z-x\right)^{l-1}e^{-z}dxdz.$$

Now make the change of variables, $x=zt$ to find,

$$\begin{aligned}\Gamma\left(k\right)\Gamma\left(l\right)&=\int_0^\infty dze^{-z}\int_0^1 dt\left(zt\right)^{k-1}\left(z-tz\right)^{l-1}z\\&=\int_0^\infty e^{-z}z^{k+l-1}dz\cdot\int_0^1 t^{k-1}\left(1-t\right)^{l-1}dt\\&=\Gamma\left(k+l\right)B\left(k,l\right).\end{aligned}$$

**Definition 12.32 (Beta distribution).** *The $\beta$ – distribution is*

$$d\mu_{x,y}\left(t\right)=\frac{t^{x-1}\left(1-t\right)^{y-1}dt}{B\left(x,y\right)}.$$

Observe that

$$\int_0^1 td\mu_{x,y}\left(t\right)=\frac{B\left(x+1,y\right)}{B\left(x,y\right)}=\frac{\frac{\Gamma(x+1)\Gamma(y)}{\Gamma(x+y+1)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}}=\frac{x}{x+y}$$

and

$$\int_0^1 t^2d\mu_{x,y}\left(t\right)=\frac{B\left(x+2,y\right)}{B\left(x,y\right)}=\frac{\frac{\Gamma(x+2)\Gamma(y)}{\Gamma(x+y+2)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}}=\frac{\left(x+1\right)x}{\left(x+y+1\right)\left(x+y\right)}.$$

## 12.5 A Strong Law of Large Numbers

**Theorem 12.33 (A simple form of the strong law of large numbers).** *If $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables such that $\mathbb{E}\left[\left|X_n\right|^4\right]<\infty$, then*

$$\lim_{n\to\infty}\frac{S_n}{n}=\mu\ a.s.$$

*where $S_n:=\sum_{k=1}^n X_k$ and $\mu:=\mathbb{E}X_n=\mathbb{E}X_1$.*

**Exercise 12.6.** Use the following outline to give a proof of Theorem 12.33.

1. First show that $x^p\le 1+x^4$ for all $x\ge 0$ and $1\le p\le 4$. Use this to conclude;
$$\mathbb{E}\left|X_n\right|^p\le 1+\mathbb{E}\left|X_n\right|^4<\infty \text{ for } 1\le p\le 4.$$
Thus $\gamma:=\mathbb{E}\left[\left|X_n-\mu\right|^4\right]$ and the standard deviation $\left(\sigma^2\right)$ of $X_n$ defined by,
$$\sigma^2:=\mathbb{E}\left[X_n^2\right]-\mu^2=\mathbb{E}\left[\left(X_n-\mu\right)^2\right]<\infty,$$
are finite constants independent of $n$.
2. Show for all $n\in\mathbb{N}$ that
$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{n}-\mu\right)^4\right]&=\frac{1}{n^4}\left(n\gamma+3n(n-1)\sigma^4\right)\\&=\frac{1}{n^2}\left[n^{-1}\gamma+3\left(1-n^{-1}\right)\sigma^4\right].\end{aligned}$$
(Thus $\frac{S_n}{n}\to\mu$ in $L^4\left(P\right)$.)
3. Use item 2. and Chebyshev's inequality to show
$$P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\le\frac{n^{-1}\gamma+3\left(1-n^{-1}\right)\sigma^4}{\varepsilon^4n^2}.$$
4. Use item 3. and the first Borel Cantelli Lemma 9.14 to conclude $\lim_{n\to\infty}\frac{S_n}{n}=\mu$ a.s.

## 12.6 A Central Limit Theorem

In this section we will give a preliminary a couple versions of the central limit theorem following [28, Chapter 2.14]. The next lemma contains the key to the results in this section.

**Lemma 12.34.** *Suppose that $\{U, X, Y\}$ are independent random variables such that $\mathbb{E}\left[|U|^3 + |X|^3 + |Y|^3\right] < \infty$, $\mathbb{E}X = \mathbb{E}Y$, and $\mathbb{E}X^2 = \mathbb{E}Y^2$. Then for every function, $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}} \left|f^{(3)}(x)\right| < \infty$,*

$$|\mathbb{E}[f(U+X) - f(U+Y)]| \le \frac{M}{3!} \cdot \mathbb{E}\left[|X|^3 + |Y|^3\right].$$

**Proof.** By Taylor's theorem for $u, \Delta \in \mathbb{R}$ there exists $c(u,\Delta)$ between $u$ and $\Delta$ so that

$$f(u+\Delta) = f(u) + f'(u)\Delta + \frac{1}{2}f''(u)\Delta^2 + r(u,\Delta) \tag{12.14}$$

where

$$|r(u,\Delta)| = \left|\frac{1}{3!}f^{(3)}(c(u,\Delta))\Delta^3\right| \le \frac{M}{3!}|\Delta|^3.$$

Using Eq. (12.14) with $u = U$ and $\Delta = X$ and $u = U$ and $\Delta = Y$ shows,

$$f(U+X) - f(U+Y) = f'(U)[X-Y] + \frac{1}{2}f''(U)\left[X^2 - Y^2\right] + r(U,X) - r(U,Y).$$

Taking expectations of this equation making use of the independence of $\{U, X, Y\}$ and the assumptions that $\mathbb{E}X = \mathbb{E}Y$ and $\mathbb{E}X^2 = \mathbb{E}Y^2$ shows,

$$|\mathbb{E}[f(U+X) - f(U+Y)]| = |\mathbb{E}[r(U,X) - r(U,Y)]| \le \frac{M}{3!} \cdot \mathbb{E}\left[|X|^3 + |Y|^3\right].$$

■

**Notation 12.35** *Given a square integrable random variable $Y$, let*

$$\bar{Y} := \frac{Y - \mathbb{E}Y}{\sigma(Y)} \text{ where } \sigma(Y) := \sqrt{\mathbb{E}(Y - \mathbb{E}Y)^2} = \sqrt{\operatorname{Var}(Y)}.$$

Let us also recall that if $Z = N\left(0, \sigma^2\right)$, then $Z \stackrel{d}{=} \sqrt{\sigma} N(0,1)$ and so by Eq. (9.48) with $\beta = 3$ we have,

$$\mathbb{E}\left|Z^3\right| = \sigma^3 \mathbb{E}|N(0,1)|^3 = \sqrt{8/\pi}\,\sigma^3. \tag{12.15}$$

**Theorem 12.36 (A CLT proof w/o Fourier).** *Suppose that $\{X_k\}_{k=1}^\infty \subset L^3(P)$ is a sequence of independent random variables such that*

$$C := \sup_k \mathbb{E}\left|X_k - \mathbb{E}X_k\right|^3 < \infty$$

*Then for every function, $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}} \left|f^{(3)}(x)\right| < \infty$ we have*

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \frac{M}{3!}\left(1 + \sqrt{8/\pi}\right)\frac{C}{\sigma(S_n)^3} \cdot n, \tag{12.16}$$

*where*

$$\bar{S}_n := \frac{1}{\sigma(S_n)}\left[S_n - \mathbb{E}S_n\right] \text{ with } S_n := X_1 + \cdots + X_n,$$

*and $N \stackrel{d}{=} N(0,1)$. In particular if we further assume that*

$$\delta := \liminf_{n\to\infty} \frac{1}{n}\sigma(S_n)^2 = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \operatorname{Var}(X_i) > 0, \tag{12.17}$$

*Then it follows that*

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| = O\left(\frac{1}{\sqrt{n}}\right) \text{ as } n \to \infty \tag{12.18}$$

*which is to say, $\bar{S}_n$ is "close" in distribution to $N$, which we abbreviate by $\bar{S}_n \stackrel{d}{\cong} N$ for large $n$. (It should be noted that the estimate in Eq. (12.16) is valid for any finite collection of random variables, $\{X_k\}_{k=1}^n$.)*

**Proof.** Let $n \in \mathbb{N}$ be fixed and then Let $\{Y_k, N_k\}_{k=1}^\infty$ be a collection of independent random variables such that

$$Y_k \stackrel{d}{=} \bar{X}_k = \frac{X_k - \mathbb{E}X_k}{\sigma(S_n)} \text{ and } N_k \stackrel{d}{=} N(0, \operatorname{Var}(Y_k)) \text{ for } 1 \le k \le n.$$

Let $S_n^Y = Y_1 + \cdots + Y_n \stackrel{d}{=} \bar{S}_n$ and $T_n := N_1 + \cdots + N_n$. Since

$$\sum_{k=1}^n \operatorname{Var}(N_k) = \sum_{k=1}^n \operatorname{Var}(Y_k) = \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \operatorname{Var}(X_k - \mathbb{E}X_k)$$
$$= \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \operatorname{Var}(X_k) = 1,$$

it follows by Corollary 12.28) that $T_n \stackrel{d}{=} N(0,1)$.

To compare $\mathbb{E}f\left(\bar{S}_n\right)$ with $\mathbb{E}f(N)$ we may compare $\mathbb{E}f\left(S_n^Y\right)$ with $\mathbb{E}f(T_n)$ which we will do by interpolating between $S_n^Y$ and $T_n$. To this end, for $0 \le k \le n$, let

$$V_k := N_1 + \dots + N_k + Y_{k+1} + \dots + Y_n$$

with the convention that $V_n = T_n$ and $V_0 = S_n^Y$. Then by a telescoping series argument, it follows that

$$\mathbb{E}f(T_n) - \mathbb{E}f\left(S_n^Y\right) = \mathbb{E}\left[f(V_n) - f(V_0)\right] = \sum_{k=1}^n \mathbb{E}\left[f(V_k) - f(V_{k-1})\right]. \quad (12.19)$$

If we set

$$U_k := N_1 + \dots + N_{k-1} + Y_{k+1} + \dots + Y_n$$

we have $V_k = U_k + N_k$ and $V_{k-1} = U_k + Y_k$. Therefore it follows by an application of Lemma 12.34 that

$$\left|\mathbb{E}\left[f(V_k) - f(V_{k-1})\right]\right| \le \frac{M}{3!}\left[\mathbb{E}\left|N_k\right|^3 + \mathbb{E}\left|Y_k\right|^3\right]$$

which combined with Eq. (12.19) shows

$$\begin{aligned}\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| &= \left|\mathbb{E}f(T_n) - \mathbb{E}f\left(S_n^Y\right)\right| \\ &\le \sum_{k=1}^n \left|\mathbb{E}\left[f(V_k) - f(V_{k-1})\right]\right| \\ &\le \frac{M}{3!}\sum_{k=1}^n \left[\mathbb{E}\left|N_k\right|^3 + \mathbb{E}\left|Y_k\right|^3\right]. \quad (12.20)\end{aligned}$$

We further have the estimates;

$$\begin{aligned}\mathbb{E}\left|Y_k\right|^3 &= \mathbb{E}\left|\left(\frac{X_k - \mathbb{E}X_k}{\sigma(S_n)}\right)^3\right| \le \frac{C}{\sigma(S_n)^3} \text{ and} \\ \mathbb{E}\left|N_k\right|^3 &= \sqrt{8/\pi}\cdot\operatorname{Var}(N_k)^{3/2} = \sqrt{8/\pi}\cdot\operatorname{Var}(Y_k)^{3/2} = \sqrt{8/\pi}\cdot\left(\mathbb{E}Y_k^2\right)^{3/2} \\ &\le \sqrt{8/\pi}\cdot\mathbb{E}\left|Y_k\right|^3 \le \sqrt{8/\pi}\frac{C}{\sigma(S_n)^3},\end{aligned}$$

wherein we have used Eq. (12.15) and Jensen's (or Hölder's) inequality (see Chapter 14 below) in the last two lines. or the last inequality. Combining these estimates with Eq. (12.20) gives the estimate in Eq. (12.16). ■

**Corollary 12.37.** *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables in $L^3(P)$, $C := \mathbb{E}\left|X_1 - \mathbb{E}X_1\right|^3 < \infty$, $S_n := X_1 + \dots + X_n$, and $N \overset{d}{=} N(0,1)$. Then for every function, $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}}\left|f^{(3)}(x)\right| < \infty$ we have*

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \frac{M}{3!\sqrt{n}}\left(1 + \sqrt{8/\pi}\right)\frac{C}{\operatorname{Var}(X_1)^{3/2}}. \quad (12.21)$$

*(This is a specialized form of the "Berry–Esseen theorem.")*

*Example 12.38.* Let $\lambda > 0$. Recall from Exercise 12.5, if $\{X_i\}_{i=1}^n$ are i.i.d. with $X_i \overset{d}{=} \operatorname{Poi}(\lambda/n)$, then $X = S_n = \sum_{i=1}^n X_i \overset{d}{=} \operatorname{Poi}(\lambda)$. From Exercise 9.5 we know that

$$\mathbb{E}X = \lambda = \operatorname{Var}(X)$$

and so we might expect from Corollary 12.37 that $\frac{X-\lambda}{\sqrt{\lambda}} = \overline{S_n}$ should be close to the standard normal and in fact by letting $n \to \infty$ we might errantly conclude that $\frac{X-\lambda}{\sqrt{\lambda}} \overset{d}{=} N(0,1)$. To see what is going on let us consider the estimate in Eq. (12.21) which states,

$$\begin{aligned}\left|\mathbb{E}f(N) - \mathbb{E}f\left(\frac{X-\lambda}{\sqrt{\lambda}}\right)\right| &\le \frac{M}{3!\sqrt{n}}\left(1 + \sqrt{8/\pi}\right)\frac{\mathbb{E}\left|X_1 - \mathbb{E}X_1\right|^3}{\operatorname{Var}(X_1)^{3/2}} \\ &= \frac{M}{3!\sqrt{n}}\left(1 + \sqrt{8/\pi}\right)\frac{\mathbb{E}\left|X_1 - \mathbb{E}X_1\right|^3}{\left(\frac{\lambda}{n}\right)^{3/2}} \\ &= \frac{M}{3!\lambda^{3/2}}\left(1 + \sqrt{8/\pi}\right)\left(\mathbb{E}\left|X_1 - \mathbb{E}X_1\right|^3\right)n \quad (12.22)\end{aligned}$$

To see what is happening as $n \to \infty$ we need to first show

$$\mathbb{E}X^3 = \lambda^3 + 3\lambda^2 + \lambda.$$

To see this recall from Exercise 9.5, if $X \overset{d}{=} \operatorname{Poi}(\lambda)$ then

$$\mathbb{E}X = \lambda = \operatorname{Var}(X) \text{ and}$$

$$\mathbb{E}\left[z^X\right] = \sum_{k=0}^\infty z^k \frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda(z-1)}.$$

Using $\mathbb{E}X^k = \left(z\frac{d}{dz}\right)^k|_{z=1}e^{\lambda(z-1)}$, we learn

$$\left(z\frac{d}{dz}\right)e^{\lambda(z-1)} = \lambda z e^{\lambda(z-1)} \implies \mathbb{E}X = \lambda,$$

$$\left(z\frac{d}{dz}\right)^2 e^{\lambda(z-1)} = \left(\lambda^2 z^2 + \lambda z\right)e^{\lambda(z-1)} \implies \mathbb{E}X^2 = \lambda^2 + \lambda$$

$$\begin{aligned}\left(z\frac{d}{dz}\right)^3 e^{\lambda(z-1)} &= \left[\left(\lambda^2 z^2 + \lambda z\right)\lambda z + 2\lambda^2 z^2 + \lambda z\right]e^{\lambda(z-1)} \\ &\implies \mathbb{E}X^3 = \lambda^3 + 3\lambda^2 + \lambda.\end{aligned}$$

We now see that

$$|\mathbb{E}[X-\lambda]|^3 \le \mathbb{E}|X-\lambda|^3 \le \mathbb{E}(X+\lambda)^3.$$

Since

$$\mathbb{E}(X\pm\lambda)^3 = \mathbb{E}\left[X^3 \pm \binom{3}{2}\left[\lambda X^2 \pm \lambda^2 X\right] \pm \lambda^3\right] = \lambda + O\left(\lambda^2\right),$$

we conclude that

$$\mathbb{E}|X-\mathbb{E}X|^3 = \lambda + O\left(\lambda^2\right) \text{ as } \lambda \downarrow 0.$$

Using this result in Eq. (12.22) then shows with $K = \frac{M}{3!}\left(1+\sqrt{8/\pi}\right)$ that

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\frac{X-\lambda}{\sqrt{\lambda}}\right)\right| \le \frac{K}{\lambda^{3/2}} n \cdot \left[\frac{\lambda}{n} + O\left((\lambda/n)^2\right)\right].$$

Letting $n\to\infty$ implies

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\frac{X-\lambda}{\sqrt{\lambda}}\right)\right| \le \frac{K}{\sqrt{\lambda}}. \tag{12.23}$$

We certainly can not conclude from this estimate that $X$ is Gaussian which is good as it is not. On the other hand if $X_\lambda \overset{d}{=} \mathrm{Poi}(\lambda)$, then Eq. (12.23) does imply

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \implies N(0,1) \text{ as } \lambda\to\infty,$$

where the convergence " $\implies$ " is weak convergence as described in Lemma 12.40 below.

By a slight modification of the proof of Theorem 12.36 we have the following central limit theorem.

**Theorem 12.39 (A CLT proof w/o Fourier).** *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables in $L^2(P)$, $S_n := X_1+\dots+X_n$, and $N \overset{d}{=} N(0,1)$. Then for every function, $f\in C^2(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}}\left|f^{(2)}(x)\right| < \infty$ and $f''$ being uniformly continuous on $\mathbb{R}$ we have,*

$$\lim_{n\to\infty}\mathbb{E}f\left(\bar{S}_n\right) = \mathbb{E}f(N).$$

**Proof.** In this proof we use the following form of Taylor's theorem;

$$f(x+\Delta) - f(x) = f'(x)\Delta + \frac{1}{2}f''(x)\Delta^2 + r(x,\Delta)\Delta^2 \tag{12.24}$$

where

$$r(x,\Delta) = \int_0^1 \left[f''(x+t\Delta) - f''(x)\right](1-t)\,dt.$$

Taking Eq. (12.24) with $\Delta$ replaced by $\delta$ and subtracting the results then implies

$$f(x+\Delta) - f(x+\delta) = f'(x)(\Delta-\delta) + \frac{1}{2}f''(x)\left(\Delta^2-\delta^2\right) + \rho(x,\Delta,\delta)$$

where now,

$$\rho(x,\Delta,\delta) = r(x,\Delta)\Delta^2 - r(x,\delta)\delta^2.$$

Since $f''$ is uniformly continuous it follows that

$$\varepsilon(\Delta) := \frac{1}{2}\sup\{|f''(x+t\Delta) - f(x)| : x\in\mathbb{R} \text{ and } 0\le t\le 1\} \to 0$$

Thus we may conclude that

$$|r(x,\Delta)| \le \int_0^1 |f''(x+t\Delta) - f''(x)|(1-t)\,dt \le \int_0^1 2\varepsilon(\Delta)(1-t)\,dt = \varepsilon(\Delta).$$

and therefore that

$$|\rho(x,\Delta,\delta)| \le \varepsilon(\Delta)\Delta^2 + \varepsilon(\delta)\delta^2.$$

So working just as in the proof of Theorem 12.36 we may conclude,

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \sum_{k=1}^n \mathbb{E}|R_k|$$

where now,

$$|R_k| = \varepsilon(N_k)N_k^2 + \varepsilon(Y_k)Y_k^2.$$

Since the $\{Y_k\}_{k=1}^n$ and the $\{N_k\}_{k=1}^n$ are *i.i.d.* now it follows that

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le n\cdot\mathbb{E}\left[\varepsilon(N_1)N_1^2 + \varepsilon(Y_1)Y_1^2\right].$$

Since $\mathrm{Var}(S_n) = n\cdot\mathrm{Var}(X_1)$, we have $Y_1 = \frac{X_1-\mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}$, $\mathrm{Var}(N_1) = \mathrm{Var}(Y_1) = \frac{1}{n}$ and therefore $N_1 \overset{d}{=} \sqrt{\frac{1}{n}}N$. Combining these observations shows,

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \mathbb{E}\left[\varepsilon\left(\sqrt{\frac{1}{n}}N\right)N^2 + \varepsilon\left(\frac{X_1-\mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}\right)\frac{(X_1-\mathbb{E}X_1)^2}{\sigma^2(X_1)}\right]$$

which goes to zero as $n\to\infty$ by the DCT. ■

**Lemma 12.40.** *Suppose that $\{W\}\cup\{W_n\}_{n=1}^\infty$ is a collection of random variables such that $\lim_{n\to\infty}\mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all $f\in C_c^\infty(\mathbb{R})$, then $W_n \implies W$ as $n\to\infty$ where " $\implies$ " is used to denote weak convergence, i.e. $\lim_{n\to\infty}\mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all bounded continuous functions, $f:\mathbb{R}\to\mathbb{R}$.*

**Proof.** According to Theorem 23.32 below it suffices to show $\lim_{n\to\infty}\mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all $f \in C_c(\mathbb{R})$. For such a function, $f \in C_c(\mathbb{R})$, we may find[1] $f_k \in C_c^\infty(\mathbb{R})$ with all supports being contained in a compact subset of $\mathbb{R}$ such that $\varepsilon_k := \sup_{x\in\mathbb{R}}|f(x) - f_k(x)| \to 0$ as $k \to \infty$. We then have,

$$\begin{aligned}|\mathbb{E}f(W) - \mathbb{E}f(W_n)| \le& |\mathbb{E}f(W) - \mathbb{E}f_k(W)| \\ &+ |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| + |\mathbb{E}f_k(W_n) - \mathbb{E}f(W_n)| \\ \le& \mathbb{E}|f(W) - f_k(W)| \\ &+ |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| + \mathbb{E}|f_k(W_n) - f(W_n)| \\ \le& 2\varepsilon_k + |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)|.\end{aligned}$$

Therefore it follows that

$$\begin{aligned}\limsup_{n\to\infty}|\mathbb{E}f(W) - \mathbb{E}f(W_n)| &\le 2\varepsilon_k + \limsup_{n\to\infty}|\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| \\ &= 2\varepsilon_k \to 0 \text{ as } k \to \infty.\end{aligned}$$

■

**Corollary 12.41.** *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent random variables, then under the hypothesis on this sequence in either of Theorem 12.36 or Theorem 12.39 we have $\bar{S}_n \implies N(0,1)$, i.e. $\lim_{n\to\infty}\mathbb{E}f(\bar{S}_n) = \mathbb{E}f(N(0,1))$ for all $f : \mathbb{R} \to \mathbb{R}$ which are bounded and continuous.*

For more on the methods employed in this section the reader is advised to look up "Stein's method." In Chapters 24 and 25 below, we will relax the assumptions in the above theorem. The proofs later will be based in the characteristic functional or equivalently the Fourier transform.

## 12.7 The Second Borel-Cantelli Lemma

**Lemma 12.42.** *If $0 \le x \le \frac{1}{2}$, then*

$$e^{-2x} \le 1 - x \le e^{-x}. \tag{12.25}$$

*Moreover, the upper bound in Eq. (12.25) is valid for all $x \in \mathbb{R}$.*

**Proof.** The upper bound follows by the convexity of $e^{-x}$, see Figure 12.1. For the lower bound we use the convexity of $\varphi(x) = e^{-2x}$ to conclude that the line joining $(0,1) = (0,\varphi(0))$ and $(1/2, e^{-1}) = (1/2, \varphi(1/2))$ lies above $\varphi(x)$ for $0 \le x \le 1/2$. Then we use the fact that the line $1 - x$ lies above this line

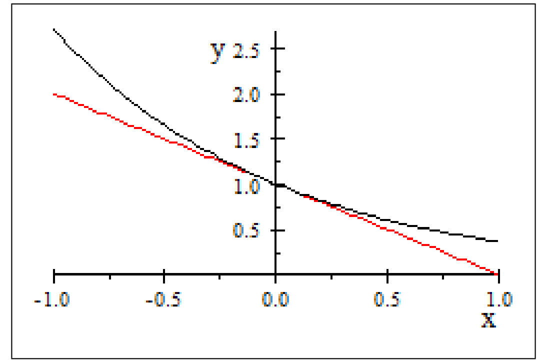



**Fig. 12.1.** A graph of $1 - x$ and $e^{-x}$ showing that $1 - x \le e^{-x}$ for all $x$.

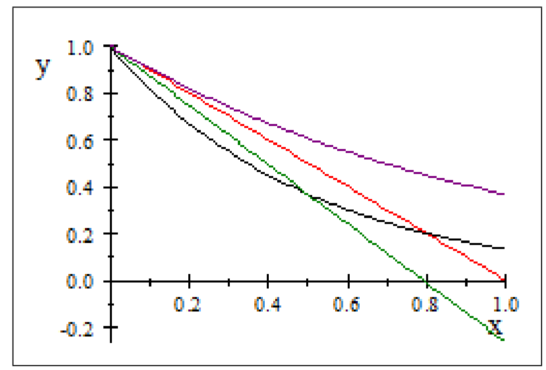



**Fig. 12.2.** A graph of $1-x$ (in red), the line joining $(0,1)$ and $(1/2, e^{-1})$ (in green), $e^{-x}$ (in purple), and $e^{-2x}$ (in black) showing that $e^{-2x} \le 1 - x \le e^{-x}$ for all $x \in [0, 1/2]$.

to conclude the lower bound in Eq. (12.25), see Figure 12.2. See Example 14.62 below for a more formal proof of this lemma. ■

For $\{a_n\}_{n=1}^\infty \subset [0,1]$, let

$$\prod_{n=1}^\infty (1 - a_n) := \lim_{N\to\infty}\prod_{n=1}^N (1 - a_n).$$

The limit exists since, $\prod_{n=1}^N (1 - a_n)$ decreases as $N$ increases.

**Corollary 12.43.** *If $\{a_n\}_{n=1}^\infty \subset [0,1]$ then*

$$\lim_{N\to\infty}\prod_{n=N}^\infty (1 - a_n) = \begin{cases} 1 \text{ if } \sum_{n=1}^\infty a_n < \infty \\ 0 \text{ if } \sum_{n=1}^\infty a_n = \infty \end{cases}.$$

[1] We will eventually prove this standard real analysis fact later in the course.

**Exercise 12.7.** Use Lemma 12.42 to prove Corollary 12.43.

**Exercise 12.8.** Show; if $\{a_n\}_{n=1}^{\infty} \subset [0,1)$, then

$$\prod_{n=1}^{\infty}(1-a_n) = 0 \iff \sum_{n=1}^{\infty} a_n = \infty.$$

The implication, $\Longleftarrow$ , holds even if $a_n = 1$ is allowed.

**Lemma 12.44 (Second Borel-Cantelli Lemma).** *Suppose that $\{A_n\}_{n=1}^{\infty}$ are independent sets. If*

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \tag{12.26}$$

*then*

$$P(\{A_n \text{ i.o.}\}) = 1. \tag{12.27}$$

*Combining this with the first Borel Cantelli Lemma 9.14 gives the **(Borel) Zero-One law**,*

$$P(A_n \text{ i.o.}) = \begin{cases} 0 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}.$$

**Proof.** We are going to prove Eq. (12.27) by showing,

$$0 = P(\{A_n \text{ i.o.}\}^c) = P(\{A_n^c \text{ a.a}\}) = P(\cup_{n=1}^{\infty} \cap_{k \ge n} A_k^c).$$

Since $\cap_{k\ge n} A_k^c \uparrow \cup_{n=1}^{\infty} \cap_{k\ge n} A_k^c$ as $n \to \infty$ and $\cap_{k=n}^m A_k^c \downarrow \cap_{n=1}^{\infty} \cup_{k\ge n} A_k$ as $m \to \infty$,

$$P(\cup_{n=1}^{\infty} \cap_{k\ge n} A_k^c) = \lim_{n\to\infty} P(\cap_{k\ge n} A_k^c) = \lim_{n\to\infty}\lim_{m\to\infty} P(\cap_{m\ge k\ge n} A_k^c).$$

Making use of the independence of $\{A_k\}_{k=1}^{\infty}$ and hence the independence of $\{A_k^c\}_{k=1}^{\infty}$, we have

$$P(\cap_{m\ge k\ge n} A_k^c) = \prod_{m\ge k\ge n} P(A_k^c) = \prod_{m\ge k\ge n} (1 - P(A_k)). \tag{12.28}$$

Using the upper estimate in Eq. (12.25) along with Eq. (12.28) shows

$$P(\cap_{m\ge k\ge n} A_k^c) \le \prod_{m\ge k\ge n} e^{-P(A_k)} = \exp\left(-\sum_{k=n}^{m} P(A_k)\right).$$

Using Eq. (12.26), we find from the above inequality that $\lim_{m\to\infty} P(\cap_{m\ge k\ge n} A_k^c) = 0$ and hence

$$P(\cup_{n=1}^{\infty} \cap_{k\ge n} A_k^c) = \lim_{n\to\infty}\lim_{m\to\infty} P(\cap_{m\ge k\ge n} A_k^c) = \lim_{n\to\infty} 0 = 0$$

Note: we could also appeal to Exercise 12.8 above to give a proof of the Borel Zero-One law without appealing to the first Borel Cantelli Lemma. ∎

*Example 12.45 (Example 9.15 continued).* Suppose that $\{X_n\}$ are now independent Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. Then $P(\lim_{n\to\infty} X_n = 0) = 1$ iff $\sum p_n < \infty$. Indeed, $P(\lim_{n\to\infty} X_n = 0) = 1$ iff $P(X_n = 0 \text{ a.a.}) = 1$ iff $P(X_n = 1 \text{ i.o.}) = 0$ iff $\sum p_n = \sum P(X_n = 1) < \infty$.

**Proposition 12.46 (Extremal behaviour of iid random variables).** *Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables and $c_n$ is an increasing sequence of positive real numbers such that for all $\alpha > 1$ we have*

$$\sum_{n=1}^{\infty} P\left(X_1 > \alpha^{-1} c_n\right) = \infty \tag{12.29}$$

*while*

$$\sum_{n=1}^{\infty} P(X_1 > \alpha c_n) < \infty. \tag{12.30}$$

*Then*

$$\limsup_{n\to\infty} \frac{X_n}{c_n} = 1 \text{ a.s.} \tag{12.31}$$

**Proof.** By the second Borel-Cantelli Lemma, Eq. (12.29) implies

$$P\left(X_n > \alpha^{-1} c_n \text{ i.o. } n\right) = 1$$

from which it follows that

$$\limsup_{n\to\infty} \frac{X_n}{c_n} \ge \alpha^{-1} \text{ a.s.}.$$

Taking $\alpha = \alpha_k = 1 + 1/k$, we find

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} \ge 1\right) = P\left(\cap_{k=1}^{\infty}\left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \ge \frac{1}{\alpha_k}\right\}\right) = 1.$$

Similarly, by the first Borel-Cantelli lemma, Eq. (12.30) implies

$$P(X_n > \alpha c_n \text{ i.o. } n) = 0$$

or equivalently,

$$P(X_n \le \alpha c_n \text{ a.a. } n) = 1.$$

That is to say,

$$\limsup_{n\to\infty} \frac{X_n}{c_n} \le \alpha \text{ a.s.}$$

and hence working as above,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} \le 1\right) = P\left(\cap_{k=1}^\infty \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \le \alpha_k\right\}\right) = 1.$$

Hence,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} = 1\right) = P\left(\left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \ge 1\right\} \cap \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \le 1\right\}\right) = 1.$$

■

*Example 12.47.* Let $\{X_n\}_{n=1}^\infty$ be i.i.d. standard normal random variables. Then by Mills' ratio (see Lemma 9.71),

$$P(X_n \ge \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha c_n} e^{-\alpha^2 c_n^2/2}.$$

Now, suppose that we take $c_n$ so that

$$e^{-c_n^2/2} = \frac{1}{n} \implies c_n = \sqrt{2\ln(n)}.$$

It then follows that

$$P(X_n \ge \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha\sqrt{2\ln(n)}} e^{-\alpha^2 \ln(n)} = \frac{1}{2\alpha\sqrt{\pi\ln(n)}} \frac{1}{n^{-\alpha^2}}$$

and therefore

$$\sum_{n=1}^\infty P(X_n \ge \alpha c_n) = \infty \text{ if } \alpha < 1$$

and

$$\sum_{n=1}^\infty P(X_n \ge \alpha c_n) < \infty \text{ if } \alpha > 1.$$

Hence an application of Proposition 12.46 shows

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \text{ a.s.}.$$

*Example 12.48.* Let $\{E_n\}_{n=1}^\infty$ be a sequence of i.i.d. random variables with exponential distributions determined by

$$P(E_n > x) = e^{-(x\vee 0)} \text{ or } P(E_n \le x) = 1 - e^{-(x\vee 0)}.$$

(Observe that $P(E_n \le 0) = 0$) so that $E_n > 0$ a.s.) Then for $c_n > 0$ and $\alpha > 0$, we have

$$\sum_{n=1}^\infty P(E_n > \alpha c_n) = \sum_{n=1}^\infty e^{-\alpha c_n} = \sum_{n=1}^\infty \left(e^{-c_n}\right)^\alpha.$$

Hence if we choose $c_n = \ln n$ so that $e^{-c_n} = 1/n$, then we have

$$\sum_{n=1}^\infty P(E_n > \alpha \ln n) = \sum_{n=1}^\infty \left(\frac{1}{n}\right)^\alpha$$

which is convergent iff $\alpha > 1$. So by Proposition 12.46, it follows that

$$\limsup_{n\to\infty} \frac{E_n}{\ln n} = 1 \text{ a.s.}$$

*Example 12.49.* * Suppose now that $\{X_n\}_{n=1}^\infty$ are i.i.d. distributed by the Poisson distribution with intensity, $\lambda$, i.e.

$$P(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In this case we have

$$P(X_1 \ge n) = e^{-\lambda} \sum_{k=n}^\infty \frac{\lambda^k}{k!} \ge \frac{\lambda^n}{n!} e^{-\lambda}$$

and

$$\begin{aligned}\sum_{k=n}^\infty \frac{\lambda^k}{k!} e^{-\lambda} &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=n}^\infty \frac{n!}{k!} \lambda^{k-n} \\ &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^\infty \frac{n!}{(k+n)!} \lambda^k \le \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^\infty \frac{1}{k!} \lambda^k = \frac{\lambda^n}{n!}.\end{aligned}$$

Thus we have shown that

$$\frac{\lambda^n}{n!} e^{-\lambda} \le P(X_1 \ge n) \le \frac{\lambda^n}{n!}.$$

Thus in terms of convergence issues, we may assume that

$$P(X_1 \ge x) \sim \frac{\lambda^x}{x!} \sim \frac{\lambda^x}{\sqrt{2\pi x} e^{-x} x^x}$$

wherein we have used Stirling's formula,

$$x! \sim \sqrt{2\pi x} e^{-x} x^x.$$

Now suppose that we wish to choose $c_n$ so that

$$P(X_1 \geq c_n) \sim 1/n.$$

This suggests that we need to solve the equation, $x^x = n$. Taking logarithms of this equation implies that

$$x = \frac{\ln n}{\ln x}$$

and upon iteration we find,

$$\begin{aligned} x &= \frac{\ln n}{\ln\left(\frac{\ln n}{\ln x}\right)} = \frac{\ln n}{\ell_2(n) - \ell_2(x)} = \frac{\ln n}{\ell_2(n) - \ell_2\left(\frac{\ln n}{\ln x}\right)} \\ &= \frac{\ln n}{\ell_2(n) - \ell_3(n) + \ell_3(x)}. \end{aligned}$$

where $\ell_k = \overbrace{\ln \circ \ln \circ \cdots \circ \ln}^{k \text{ - times}}$. Since, $x \leq \ln(n)$, it follows that $\ell_3(x) \leq \ell_3(n)$ and hence

$$x = \frac{\ln(n)}{\ell_2(n) + O(\ell_3(n))} = \frac{\ln(n)}{\ell_2(n)}\left(1 + O\left(\frac{\ell_3(n)}{\ell_2(n)}\right)\right).$$

Thus we are lead to take $c_n := \frac{\ln(n)}{\ell_2(n)}$. We then have, for $\alpha \in (0,\infty)$ that

$$\begin{aligned} (\alpha c_n)^{\alpha c_n} &= \exp\left(\alpha c_n\left[\ln\alpha + \ln c_n\right]\right) \\ &= \exp\left(\alpha\frac{\ln(n)}{\ell_2(n)}\left[\ln\alpha + \ell_2(n) - \ell_3(n)\right]\right) \\ &= \exp\left(\alpha\left[\frac{\ln\alpha - \ell_3(n)}{\ell_2(n)} + 1\right]\ln(n)\right) \\ &= n^{\alpha(1+\varepsilon_n(\alpha))} \end{aligned}$$

where

$$\varepsilon_n(\alpha) := \frac{\ln\alpha - \ell_3(n)}{\ell_2(n)}.$$

Hence we have

$$P(X_1 \geq \alpha c_n) \sim \frac{\lambda^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}e^{-\alpha c_n}(\alpha c_n)^{\alpha c_n}} \sim \frac{(\lambda/e)^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}}\frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}}.$$

Since

$$\ln(\lambda/e)^{\alpha c_n} = \alpha c_n \ln(\lambda/e) = \alpha\frac{\ln n}{\ell_2(n)}\ln(\lambda/e) = \ln n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}},$$

it follows that

$$(\lambda/e)^{\alpha c_n} = n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}}.$$

Therefore,

$$P(X_1 \geq \alpha c_n) \sim \frac{n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}}}{\sqrt{\frac{\ln(n)}{\ell_2(n)}}}\frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}} = \sqrt{\frac{\ell_2(n)}{\ln(n)}}\frac{1}{n^{\alpha(1+\delta_n(\alpha))}}$$

where $\delta_n(\alpha) \to 0$ as $n \to \infty$. From this observation, we may show,

$$\sum_{n=1}^{\infty} P(X_1 \geq \alpha c_n) < \infty \text{ if } \alpha > 1 \text{ and}$$
$$\sum_{n=1}^{\infty} P(X_1 \geq \alpha c_n) = \infty \text{ if } \alpha < 1$$

and so by Proposition 12.46 we may conclude that

$$\limsup_{n\to\infty}\frac{X_n}{\ln(n)/\ell_2(n)} = 1 \text{ a.s.}$$

## 12.8 Kolmogorov and Hewitt-Savage Zero-One Laws

Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables on a measurable space, $(\Omega, \mathcal{B})$. Let $\mathcal{B}_n := \sigma(X_1, \dots, X_n), \mathcal{B}_\infty := \sigma(X_1, X_2, \dots)$, $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \dots)$, and $\mathcal{T} := \cap_{n=1}^\infty \mathcal{T}_n \subset \mathcal{B}_\infty$. We call $\mathcal{T}$ the **tail** $\sigma$ − **field** and events, $A \in \mathcal{T}$, are called **tail events.**

*Example 12.50.* Let $S_n := X_1 + \cdots + X_n$ and $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ such that $b_n \uparrow \infty$. Here are some example of tail events and tail measurable random variables:

1. $\{\sum_{n=1}^\infty X_n \text{ converges}\} \in \mathcal{T}$. Indeed,
$$\left\{\sum_{k=1}^\infty X_k \text{ converges}\right\} = \left\{\sum_{k=n+1}^\infty X_k \text{ converges}\right\} \in \mathcal{T}_n$$
for all $n \in \mathbb{N}$.
2. Both $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are $\mathcal{T}$ − measurable as are $\limsup_{n\to\infty}\frac{S_n}{b_n}$ and $\liminf_{n\to\infty}\frac{S_n}{b_n}$.
3. $\{\lim X_n \text{ exists in } \bar{\mathbb{R}}\} = \left\{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\right\} \in \mathcal{T}$ and similarly,
$$\left\{\lim\frac{S_n}{b_n} \text{ exists in } \bar{\mathbb{R}}\right\} = \left\{\limsup_{n\to\infty}\frac{S_n}{b_n} = \liminf_{n\to\infty}\frac{S_n}{b_n}\right\} \in \mathcal{T}$$
and
$$\left\{\lim\frac{S_n}{b_n} \text{ exists in } \mathbb{R}\right\} = \left\{-\infty < \limsup_{n\to\infty}\frac{S_n}{b_n} = \liminf_{n\to\infty}\frac{S_n}{b_n} < \infty\right\} \in \mathcal{T}.$$

4. $\left\{\lim_{n\to\infty}\frac{S_n}{b_n}=0\right\}\in\mathcal{T}$. Indeed, for any $k\in\mathbb{N}$,

$$\lim_{n\to\infty}\frac{S_n}{b_n}=\lim_{n\to\infty}\frac{(X_{k+1}+\dots+X_n)}{b_n}$$

from which it follows that $\left\{\lim_{n\to\infty}\frac{S_n}{b_n}=0\right\}\in\mathcal{T}_k$ for all $k$.

**Definition 12.51.** *Let $(\Omega,\mathcal{B},P)$ be a probability space. A $\sigma$ – field, $\mathcal{F}\subset\mathcal{B}$ is* ***almost trivial*** *iff $P(\mathcal{F})=\{0,1\}$, i.e. $P(A)\in\{0,1\}$ for all $A\in\mathcal{F}$.*

The following conditions on a sub-$\sigma$-algebra, $\mathcal{F}\subset\mathcal{B}$ are equivalent; 1) $\mathcal{F}$ is almost trivial, 2) $P(A)=P(A)^2$ for all $A\in\mathcal{F}$, and 3) $\mathcal{F}$ is independent of itself. For example if $\mathcal{F}$ is independent of itself, then $P(A)=P(A\cap A)=P(A)P(A)$ for all $A\in\mathcal{F}$ which implies $P(A)=0$ or 1. If $\mathcal{F}$ is almost trivial and $A,B\in\mathcal{F}$, then $P(A\cap B)=1=P(A)P(B)$ if $P(A)=P(B)=1$ and $P(A\cap B)=0=P(A)P(B)$ if either $P(A)=0$ or $P(B)=0$. Therefore $\mathcal{F}$ is independent of itself.

**Lemma 12.52.** *Suppose that $X:\Omega\to\bar{\mathbb{R}}$ is a random variable which is $\mathcal{F}$ measurable, where $\mathcal{F}\subset\mathcal{B}$ is almost trivial. Then there exists $c\in\bar{\mathbb{R}}$ such that $X=c$ a.s.*

**Proof.** Since $\{X=\infty\}$ and $\{X=-\infty\}$ are in $\mathcal{F}$, if $P(X=\infty)>0$ or $P(X=-\infty)>0$, then $P(X=\infty)=1$ or $P(X=-\infty)=1$ respectively. Hence, it suffices to finish the proof under the added condition that $P(X\in\mathbb{R})=1$.

For each $x\in\mathbb{R}$, $\{X\le x\}\in\mathcal{F}$ and therefore, $P(X\le x)$ is either 0 or 1. Since the function, $F(x):=P(X\le x)\in\{0,1\}$ is right continuous, non-decreasing and $F(-\infty)=0$ and $F(+\infty)=1$, there is a unique point $c\in\mathbb{R}$ where $F(c)=1$ and $F(c-)=0$. At this point, we have $P(X=c)=1$.

**Alternatively** if $X:\Omega\to\mathbb{R}$ is an integrable $\mathcal{F}$ measurable random variable, we know that $X$ is independent of itself and therefore $X^2$ is integrable and $\mathbb{E}X^2=(\mathbb{E}X)^2=:c^2$. Thus it follows that $\mathbb{E}\left[(X-c)^2\right]=0$, i.e. $X=c$ a.s. For general $X:\Omega\to\mathbb{R}$, let $X_M:=(M\wedge X)\vee(-M)$, then $X_M=\mathbb{E}X_M$ a.s. For sufficiently large $M$ we know by MCT that $P(|X|<M)>0$ and since $X=X_M=\mathbb{E}X_M$ a.s. on $\{|X|<M\}$, it follows that $c=\mathbb{E}X_M$ is constant independent of $M$ for $M$ large. Therefore, $X=\lim_{M\to\infty}X_M\overset{\text{a.s.}}{=}\lim_{M\to\infty}c=c$. ■

**Proposition 12.53 (Kolmogorov's Zero-One Law).** *Suppose that $P$ is a probability measure on $(\Omega,\mathcal{B})$ such that $\{X_n\}_{n=1}^\infty$ are independent random variables. Then $\mathcal{T}$ is almost trivial, i.e. $P(A)\in\{0,1\}$ for all $A\in\mathcal{T}$. In particular the tail events in Example 12.50 have probability either 0 or 1.*

**Proof.** For each $n\in\mathbb{N}$, $\mathcal{T}\subset\sigma(X_{n+1},X_{n+2},\dots)$ which is independent of $\mathcal{B}_n:=\sigma(X_1,\dots,X_n)$. Therefore $\mathcal{T}$ is independent of $\cup\mathcal{B}_n$ which is a multiplicative system. Therefore $\mathcal{T}$ and is independent of $\mathcal{B}_\infty=\sigma(\cup\mathcal{B}_n)=\vee_{n=1}^\infty\mathcal{B}_n$. As $\mathcal{T}\subset\mathcal{B}_\infty$ it follows that $\mathcal{T}$ is independent of itself, i.e. $\mathcal{T}$ is almost trivial. ■

**Corollary 12.54.** *Keeping the assumptions in Proposition 12.53 and let $\{b_n\}_{n=1}^\infty\subset(0,\infty)$ such that $b_n\uparrow\infty$. Then $\limsup_{n\to\infty}X_n$, $\liminf_{n\to\infty}X_n$, $\limsup_{n\to\infty}\frac{S_n}{b_n}$, and $\liminf_{n\to\infty}\frac{S_n}{b_n}$ are all constant almost surely. In particular, either $P\left(\left\{\lim_{n\to\infty}\frac{S_n}{b_n}\text{ exists}\right\}\right)=0$ or $P\left(\left\{\lim_{n\to\infty}\frac{S_n}{b_n}\text{ exists}\right\}\right)=1$ and in the latter case $\lim_{n\to\infty}\frac{S_n}{b_n}=c$ a.s for some $c\in\bar{\mathbb{R}}$.*

*Example 12.55.* Suppose that $\{A_n\}_{n=1}^\infty$ are independent sets and let $X_n:=1_{A_n}$ for all $n$ and $\mathcal{T}=\cap_{n\ge1}\sigma(X_n,X_{n+1},\dots)$. Then $\{A_n\text{ i.o.}\}\in\mathcal{T}$ and therefore by the Kolmogorov 0-1 law, $P(\{A_n\text{ i.o.}\})=0$ or 1. Of course, in this case the Borel zero - one law (Lemma 12.44) tells when $P(\{A_n\text{ i.o.}\})$ is 0 and when it is 1 depending on whether $\sum_{n=1}^\infty P(A_n)$ is finite or infinite respectively.

### 12.8.1 Hewitt-Savage Zero-One Law

In this subsection, let $\Omega:=\mathbb{R}^\infty=\mathbb{R}^\mathbb{N}$ and $X_n(\omega)=\omega_n$ for all $\omega\in\Omega$ and $n\in\mathbb{N}$, and $\mathcal{B}:=\sigma(X_1,X_2,\dots)$ be the product $\sigma$ – algebra on $\Omega$. We say a permutation (i.e. a bijective map on $\mathbb{N}$), $\pi:\mathbb{N}\to\mathbb{N}$ is finite if $\pi(n)=n$ for a.a. $n$. Define $T_\pi:\Omega\to\Omega$ by $T_\pi(\omega)=(\omega_{\pi1},\omega_{\pi2},\dots)$. Since $X_i\circ T_\pi(\omega)=\omega_{\pi i}=X_{\pi i}(\omega)$ for all $i$, it follows that $T_\pi$ is $\mathcal{B}/\mathcal{B}$ – measurable.

Let us further suppose that $\mu$ is a probability measure on $(\mathbb{R},\mathcal{B}_\mathbb{R})$ and let $P=\otimes_{n=1}^\infty\mu$ be the infinite product measure on $(\Omega=\mathbb{R}^\mathbb{N},\mathcal{B})$. Then $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with $\mathrm{Law}_P(X_n)=\mu$ for all $n$. If $\pi:\mathbb{N}\to\mathbb{N}$ is a finite permutation and $A_i\in\mathcal{B}_\mathbb{R}$ for all $i$, then

$$T_\pi^{-1}(A_1\times A_2\times A_3\times\dots)=A_{\pi^{-1}1}\times A_{\pi^{-1}2}\times\dots.$$

Since sets of the form, $A_1\times A_2\times A_3\times\dots$, form a $\pi$ – system generating $\mathcal{B}$ and

$$\begin{aligned}P\circ T_\pi^{-1}(A_1\times A_2\times A_3\times\dots)&=\prod_{i=1}^\infty\mu(A_{\pi^{-1}i})\\&=\prod_{i=1}^\infty\mu(A_i)=P(A_1\times A_2\times A_3\times\dots),\end{aligned}$$

we may conclude that $P\circ T_\pi^{-1}=P$.

**Definition 12.56.** *The* ***permutation invariant*** *$\sigma$ – field, $\mathcal{S} \subset \mathcal{B}$, is the collection of sets, $A \in \mathcal{B}$ such that $T_\pi^{-1}(A) = A$ for all finite permutations $\pi$. (You should check that $\mathcal{S}$ is a $\sigma$ – field!)*

**Proposition 12.57 (Hewitt-Savage Zero-One Law).** *Let $\mu$ be a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $\left(\Omega = \mathbb{R}^\mathbb{N}, \mathcal{B}\right)$ so that $\{X_n\}_{n=1}^\infty$ (recall that $X_n(\omega) = \omega_n$) is an i.i.d. sequence with $\mathrm{Law}_P(X_n) = \mu$ for all $n$. Then $\mathcal{S}$ is $P$ – almost trivial.*

**Proof.** Let $B \in \mathcal{S}$, $f = 1_B$, and $g = G(X_1, \dots, X_n)$ be a $\sigma(X_1, X_2, \dots, X_n)$ – measurable function such that $\sup_{\omega \in \Omega} |g(\omega)| \le 1$. Further let $\pi$ be a finite permutation such that $\{\pi 1, \dots, \pi n\} \cap \{1, 2, \dots, n\} = \emptyset$ – for example we could take $\pi(j) = j + n$, $\pi(j+n) = j$ for $j = 1, 2, \dots, n$, and $\pi(j + 2n) = j + 2n$ for all $j \in \mathbb{N}$. Then $g \circ T_\pi = G(X_{\pi 1}, \dots, X_{\pi n})$ is independent of $g$ and therefore,

$$(\mathbb{E}g)^2 = \mathbb{E}g \cdot \mathbb{E}[g \circ T_\pi] = \mathbb{E}[g \cdot g \circ T_\pi].$$

Since $f \circ T_\pi = 1_{T_\pi^{-1}(B)} = 1_B = f$, it follows that $\mathbb{E}f = \mathbb{E}f^2 = \mathbb{E}[f \cdot f \circ T_\pi]$ and therefore,

$$\begin{aligned}\left|\mathbb{E}f - (\mathbb{E}g)^2\right| &= |\mathbb{E}[f \cdot f \circ T_\pi - g \cdot g \circ T_\pi]| \\ &\le \mathbb{E}|[f - g] f \circ T_\pi| + \mathbb{E}|g[f \circ T_\pi - g \circ T_\pi]| \\ &\le \mathbb{E}|f - g| + \mathbb{E}|f \circ T_\pi - g \circ T_\pi| = 2\mathbb{E}|f - g|. \qquad (12.32)\end{aligned}$$

According to Corollary 10.16 (or see Corollary 6.87 or Theorem 6.24 or Exercise 10.6)), we may choose $g = g_k$ as above with $\mathbb{E}|f - g_k| \to 0$ as $n \to \infty$ and so passing to the limit in Eq. (12.32) with $g = g_k$, we may conclude,

$$\left|P(B) - [P(B)]^2\right| = \left|\mathbb{E}f - (\mathbb{E}f)^2\right| \le 0.$$

That is $P(B) \in \{0, 1\}$ for all $B \in \mathcal{S}$. ■

In a nutshell, here is the crux of the above proof. First off we know that for $B \in \mathcal{S} \subset \mathcal{B}$, there exists $g$ which is $\sigma(X_1, \dots, X_n)$ – measurable such that $f := 1_B \cong g$. Since $P \circ T_\pi^{-1} = P$ it also follows that $f = f \circ T_\pi \cong g \circ T_\pi$. For judiciously chosen $\pi$, we know that $g$ and $g \circ T_\pi$ are independent. Therefore

$$\mathbb{E}f^2 = \mathbb{E}[f \cdot f \circ T_\pi] \cong \mathbb{E}[g \cdot g \circ T_\pi] = \mathbb{E}[g] \cdot \mathbb{E}[g \circ T_\pi] = (\mathbb{E}g)^2 \cong (\mathbb{E}f)^2.$$

As the approximation $f$ by $g$ may be made as accurate as we please, it follows that $P(B) = \mathbb{E}f^2 = (\mathbb{E}f)^2 = [P(B)]^2$ for all $B \in \mathcal{S}$.

*Example 12.58 (Some Random Walk $0-1$ Law Results).* Continue the notation in Proposition 12.57.

1. As above, if $S_n = X_1 + \dots + X_n$, then $P(S_n \in B \text{ i.o.}) \in \{0, 1\}$ for all $B \in \mathcal{B}_\mathbb{R}$. Indeed, if $\pi$ is a finite permutation,
$$T_\pi^{-1}(\{S_n \in B \text{ i.o.}\}) = \{S_n \circ T_\pi \in B \text{ i.o.}\} = \{S_n \in B \text{ i.o.}\}.$$
Hence $\{S_n \in B \text{ i.o.}\}$ is in the permutation invariant $\sigma$ – field, $\mathcal{S}$. The same goes for $\{S_n \in B \text{ a.a.}\}$
2. If $P(X_1 \ne 0) > 0$, then $\limsup_{n\to\infty} S_n = \infty$ a.s. or $\limsup_{n\to\infty} S_n = -\infty$ a.s. Indeed,
$$T_\pi^{-1}\left\{\limsup_{n\to\infty} S_n \le x\right\} = \left\{\limsup_{n\to\infty} S_n \circ T_\pi \le x\right\} = \left\{\limsup_{n\to\infty} S_n \le x\right\}$$
which shows that $\limsup_{n\to\infty} S_n$ is $\mathcal{S}$ – measurable. Therefore, $\limsup_{n\to\infty} S_n = c$ a.s. for some $c \in \bar{\mathbb{R}}$. Since $(X_2, X_3, \dots) \overset{d}{=} (X_1, X_2, \dots)$ it follows (see Corollary 8.47 and Exercise 8.10) that
$$\begin{aligned}c = \limsup_{n\to\infty} S_n &\overset{d}{=} \limsup_{n\to\infty} (X_2 + X_3 + \dots + X_{n+1}) \\ &= \limsup_{n\to\infty} (S_{n+1} - X_1) = \limsup_{n\to\infty} S_{n+1} - X_1 = c - X_1.\end{aligned}$$
By Exercise 12.9 below we may now conclude that $c = c - X_1$ a.s. which is possible iff $c \in \{\pm\infty\}$ or $X_1 = 0$ a.s. Since the latter is not allowed, $\limsup_{n\to\infty} S_n = \infty$ or $\limsup_{n\to\infty} S_n = -\infty$ a.s.
3. Now assume that $P(X_1 \ne 0) > 0$ and $X_1 \overset{\mathrm{d}}{=} -X_1$, i.e. $P(X_1 \in A) = P(-X_1 \in A)$ for all $A \in \mathcal{B}_\mathbb{R}$. By 2. we know $\limsup_{n\to\infty} S_n = c$ a.s. with $c \in \{\pm\infty\}$. Since $\{X_n\}_{n=1}^\infty$ and $\{-X_n\}_{n=1}^\infty$ are i.i.d. and $-X_n \overset{\mathrm{d}}{=} X_n$, it follows that $\{X_n\}_{n=1}^\infty \overset{\mathrm{d}}{=} \{-X_n\}_{n=1}^\infty$ .The results of Exercises 8.10 and 12.9 then imply that $c \overset{\mathrm{d}}{=} \limsup_{n\to\infty} S_n \overset{\mathrm{d}}{=} \limsup_{n\to\infty} (-S_n)$ and in particular
$$c \overset{\text{a.s.}}{=} \limsup_{n\to\infty} (-S_n) = -\liminf_{n\to\infty} S_n \ge -\limsup_{n\to\infty} S_n = -c.$$
Since the $c = -\infty$ does not satisfy, $c \ge -c$, we must $c = \infty$. Hence in this symmetric case we have shown,
$$\limsup_{n\to\infty} S_n = \infty \text{ and } \liminf_{n\to\infty} S_n = -\infty \text{ a.s.}$$

**Exercise 12.9.** Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $Y : \Omega \to \bar{\mathbb{R}}$ is a random variable and $c \in \bar{\mathbb{R}}$ is a constant. Then $Y = c$ a.s. iff $Y \overset{d}{=} c$.

## 12.9 Another Construction of Independent Random Variables*

This section may be skipped as the results are a special case of those given above. The arguments given here avoid the use of Kolmogorov's existence theorem for product measures.

*Example 12.59.* Suppose that $\Omega = \Lambda^n$ where $\Lambda$ is a finite set, $\mathcal{B} = 2^\Omega$, $P(\{\omega\}) = \prod_{j=1}^n q_j(\omega_j)$ where $q_j : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda \in \Lambda} q_j(\lambda) = 1$. Let $\mathcal{C}_i := \{\Lambda^{i-1} \times A \times \Lambda^{n-i} : A \subset \Lambda\}$. Then $\{\mathcal{C}_i\}_{i=1}^n$ are independent. Indeed, if $B_i := \Lambda^{i-1} \times A_i \times \Lambda^{n-i}$, then

$$\cap B_i = A_1 \times A_2 \times \cdots \times A_n$$

and we have

$$P(\cap B_i) = \sum_{\omega \in A_1 \times A_2 \times \cdots \times A_n} \prod_{i=1}^n q_i(\omega_i) = \prod_{i=1}^n \sum_{\lambda \in A_i} q_i(\lambda)$$

while

$$P(B_i) = \sum_{\omega \in \Lambda^{i-1} \times A_i \times \Lambda^{n-i}} \prod_{i=1}^n q_i(\omega_i) = \sum_{\lambda \in A_i} q_i(\lambda).$$

*Example 12.60.* Continue the notation of Example 12.59 and further assume that $\Lambda \subset \mathbb{R}$ and let $X_i : \Omega \to \Lambda$ be defined by, $X_i(\omega) = \omega_i$. Then $\{X_i\}_{i=1}^n$ are independent random variables. Indeed, $\sigma(X_i) = \mathcal{C}_i$ with $\mathcal{C}_i$ as in Example 12.59.

Alternatively, from Exercise **??**, we know that

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)]$$

for all $f_i : \Lambda \to \mathbb{R}$. Taking $A_i \subset \Lambda$ and $f_i := 1_{A_i}$ in the above identity shows that

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{E}_P\left[\prod_{i=1}^n 1_{A_i}(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[1_{A_i}(X_i)] = \prod_{i=1}^n P(X_i \in A_i)$$

as desired.

**Theorem 12.61 (Existence of i.i.d simple R.V.'s).** *Suppose that $\{q_i\}_{i=0}^n$ is a sequence of positive numbers such that $\sum_{i=0}^n q_i = 1$. Then there exists a sequence $\{X_k\}_{k=1}^\infty$ of simple random variables taking values in $\Lambda = \{0,1,2\dots,n\}$ on $((0,1], \mathcal{B}, m)$ such that*

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}$$

*for all $i_1, i_2, \dots, i_k \in \{0,1,2,\dots,n\}$ and all $k \in \mathbb{N}$. (See Example 12.16 above and Theorem 12.65 below for the general case of this theorem.)*

**Proof.** For $i = 0, 1, \dots, n$, let $\sigma_{-1} = 0$ and $\sigma_j := \sum_{i=0}^j q_i$ and for any interval, $(a,b]$, let

$$T_i((a,b]) := (a + \sigma_{i-1}(b-a), a + \sigma_i(b-a)].$$

Given $i_1, i_2, \dots, i_k \in \{0,1,2,\dots,n\}$, let

$$J_{i_1,i_2,\dots,i_k} := T_{i_k}\left(T_{i_{k-1}}(\dots T_{i_1}((0,1]))\right)$$

and define $\{X_k\}_{k=1}^\infty$ on $(0,1]$ by

$$X_k := \sum_{i_1,i_2,\dots,i_k \in \{0,1,2,\dots,n\}} i_k 1_{J_{i_1,i_2,\dots,i_k}},$$

see Figure 12.3. Repeated applications of Corollary 8.27 shows the functions, $X_k : (0,1] \to \mathbb{R}$ are measurable.

Observe that

$$m(T_i((a,b])) = q_i(b-a) = q_i m((a,b]), \tag{12.33}$$

and so by induction,

$$m(J_{i_1,i_2,\dots,i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}.$$

The reader should convince herself/himself that

$$\{X_1 = i_1, \dots X_k = i_i\} = J_{i_1,i_2,\dots,i_k}$$

and therefore, we have

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = m(J_{i_1,i_2,\dots,i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}$$

as desired. ■

**Corollary 12.62 (Independent variables on product spaces).** *Suppose $\Lambda = \{0,1,2\dots,n\}$, $q_i > 0$ with $\sum_{i=0}^n q_i = 1$, $\Omega = \Lambda^\infty = \Lambda^{\mathbb{N}}$, and for $i \in \mathbb{N}$, let $Y_i : \Omega \to \mathbb{R}$ be defined by $Y_i(\omega) = \omega_i$ for all $\omega \in \Omega$. Further let $\mathcal{B} := \sigma(Y_1, Y_2, \dots, Y_n, \dots)$. Then there exists a unique probability measure, $P : \mathcal{B} \to [0,1]$ such that*

$$P(\{Y_1 = i_1, \dots, Y_k = i_i\}) = q_{i_1} \dots q_{i_k}.$$

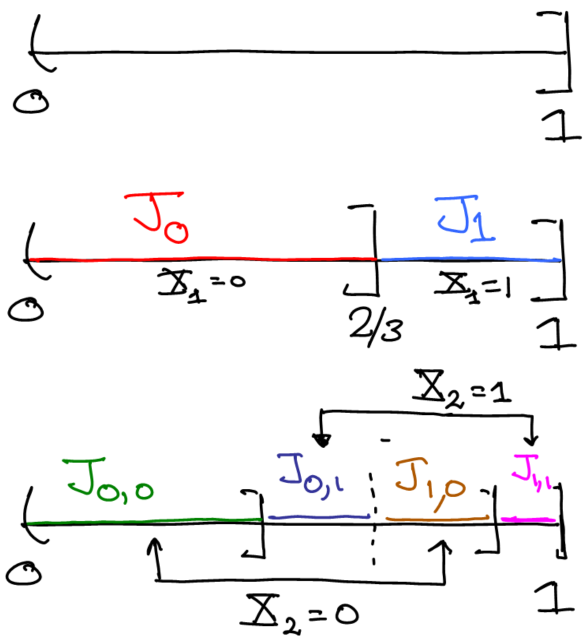



**Fig. 12.3.** Here we suppose that $p_0 = 2/3$ and $p_1 = 1/3$ and then we construct $J_l$ and $J_{l,k}$ for $l, k \in \{0, 1\}$.

**Proof.** Let $\{X_i\}_{i=1}^n$ be as in Theorem 12.61 and define $T : (0, 1] \to \Omega$ by

$$T(x) = (X_1(x), X_2(x), \dots, X_k(x), \dots).$$

Observe that $T$ is measurable since $Y_i \circ T = X_i$ is measurable for all $i$. We now define, $P := T_* m$. Then we have

$$\begin{aligned} P(\{Y_1 = i_1, \dots, Y_k = i_i\}) &= m\left(T^{-1}(\{Y_1 = i_1, \dots, Y_k = i_i\})\right) \\ &= m(\{Y_1 \circ T = i_1, \dots, Y_k \circ T = i_i\}) \\ &= m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}. \end{aligned}$$

■

**Theorem 12.63.** *Given a finite subset, $\Lambda \subset \mathbb{R}$ and a function $q : \Lambda \to [0, 1]$ such that $\sum_{\lambda \in \Lambda} q(\lambda) = 1$, there exists a probability space, $(\Omega, \mathcal{B}, P)$ and an independent sequence of random variables, $\{X_n\}_{n=1}^\infty$ such that $P(X_n = \lambda) = q(\lambda)$ for all $\lambda \in \Lambda$.*

**Proof.** Use Corollary 12.10 to shows that random variables constructed in Example 6.74 or Theorem 12.61 fit the bill. ■

**Proposition 12.64.** *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables with distribution, $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$. If we let $U := \sum_{n=1}^\infty 2^{-n} X_n$, then $P(U \le x) = (0 \vee x) \wedge 1$, i.e. $U$ has the uniform distribution on $[0, 1]$.*

**Proof.** Let us recall that $P(X_n = 0 \text{ a.a.}) = 0 = P(X_n = 1 \text{ a.a.})$. Hence we may, by shrinking $\Omega$ if necessary, assume that $\{X_n = 0 \text{ a.a.}\} = \emptyset = \{X_n = 1 \text{ a.a.}\}$. With this simplification, we have

$$\begin{aligned} \left\{U < \frac{1}{2}\right\} &= \{X_1 = 0\}, \\ \left\{U < \frac{1}{4}\right\} &= \{X_1 = 0, X_2 = 0\} \text{ and} \\ \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} &= \{X_1 = 1, X_2 = 0\} \end{aligned}$$

and hence that

$$\begin{aligned} \left\{U < \frac{3}{4}\right\} &= \left\{U < \frac{1}{2}\right\} \cup \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} \\ &= \{X_1 = 0\} \cup \{X_1 = 1, X_2 = 0\}. \end{aligned}$$

From these identities, it follows that

$$P(U < 0) = 0,\ P\left(U < \frac{1}{4}\right) = \frac{1}{4},\ P\left(U < \frac{1}{2}\right) = \frac{1}{2}, \text{ and } P\left(U < \frac{3}{4}\right) = \frac{3}{4}.$$

More generally, we claim that if $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ with $\varepsilon_j \in \{0, 1\}$, then

$$P(U < x) = x. \tag{12.34}$$

The proof is by induction on $n$. Indeed, we have already verified (12.34) when $n = 1, 2$. Suppose we have verified (12.34) up to some $n \in \mathbb{N}$ and let $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ and consider

$$\begin{aligned} P\left(U < x + 2^{-(n+1)}\right) &= P(U < x) + P\left(x \le U < x + 2^{-(n+1)}\right) \\ &= x + P\left(x \le U < x + 2^{-(n+1)}\right). \end{aligned}$$

Since

$$\left\{x \le U < x + 2^{-(n+1)}\right\} = \left[\cap_{j=1}^n \{X_j = \varepsilon_j\}\right] \cap \{X_{n+1} = 0\}$$

we see that

$$P\left(x \le U < x + 2^{-(n+1)}\right) = 2^{-(n+1)}$$

and hence

$$P\left(U < x + 2^{-(n+1)}\right) = x + 2^{-(n+1)}$$

which completes the induction argument.

Since $x \to P(U < x)$ is left continuous we may now conclude that $P(U < x) = x$ for all $x \in (0,1)$ and since $x \to x$ is continuous we may also deduce that $P(U \le x) = x$ for all $x \in (0,1)$. Hence we may conclude that

$$P(U \le x) = (0 \vee x) \wedge 1.$$

■

We may now show the existence of independent random variables with arbitrary distributions.

**Theorem 12.65.** *Suppose that $\{\mu_n\}_{n=1}^{\infty}$ are a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then there exists a probability space, $(\Omega, \mathcal{B}, P)$ and a sequence $\{Y_n\}_{n=1}^{\infty}$ independent random variables with Law $(Y_n) := P \circ Y_n^{-1} = \mu_n$ for all $n$.*

**Proof.** By Theorem 12.63, there exists a sequence of i.i.d. random variables, $\{Z_n\}_{n=1}^{\infty}$, such that $P(Z_n = 1) = P(Z_n = 0) = \frac{1}{2}$. These random variables may be put into a two dimensional array, $\{X_{i,j} : i, j \in \mathbb{N}\}$, see the proof of Lemma 4.9. For each $i$, let $U_i := \sum_{j=1}^{\infty} 2^{-i} X_{i,j}$ – $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable random variable. According to Proposition 12.64, $U_i$ is uniformly distributed on $[0,1]$. Moreover by the grouping Lemma 12.17, $\left\{\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)\right\}_{i=1}^{\infty}$ are independent $\sigma$ – algebras and hence $\{U_i\}_{i=1}^{\infty}$ is a sequence of i.i.d.. random variables with the uniform distribution.

Finally, let $F_i(x) := \mu((-\infty, x])$ for all $x \in \mathbb{R}$ and let $G_i(y) = \inf\{x : F_i(x) \ge y\}$. Then according to Theorem 8.48, $Y_i := G_i(U_i)$ has $\mu_i$ as its distribution. Moreover each $Y_i$ is $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable and therefore the $\{Y_i\}_{i=1}^{\infty}$ are independent random variables. ■

# 13

# The Standard Poisson Process

## 13.1 Poisson Random Variables

Recall from Exercise 9.5 that a Random variable, $X$, is Poisson distributed with intensity, $a$, if

$$P(X=k)=\frac{a^k}{k!}e^{-a} \text{ for all } k\in\mathbb{N}_0.$$

We will abbreviate this in the future by writing $X\overset{d}{=}\mathrm{Poi}(a)$. Let us also recall that

$$\mathbb{E}\left[z^X\right]=\sum_{k=0}^{\infty}z^k\frac{a^k}{k!}e^{-a}=e^{az}e^{-a}=e^{a(z-1)}$$

and as in Exercise 9.5 we have $\mathbb{E}X=a=\mathrm{Var}(X)$.

**Lemma 13.1.** *If $X=\mathrm{Poi}(a)$ and $Y=\mathrm{Poi}(b)$ and $X$ and $Y$ are independent, then $X+Y=\mathrm{Poi}(a+b)$.*

**Proof.** For $k\in\mathbb{N}_0$,

$$\begin{aligned}P(X+Y=k)&=\sum_{l=0}^{k}P(X=l,Y=k-l)=\sum_{l=0}^{k}P(X=l)\,P(Y=k-l)\\&=\sum_{l=0}^{k}e^{-a}\frac{a^l}{l!}e^{-b}\frac{b^{k-l}}{(k-l)!}=\frac{e^{-(a+b)}}{k!}\sum_{l=0}^{k}\binom{k}{l}a^lb^{k-l}\\&=\frac{e^{-(a+b)}}{k!}(a+b)^k.\end{aligned}$$

**Alternative Proof.** Notice that

$$\mathbb{E}\left[z^{X+Y}\right]=\mathbb{E}\left[z^X\right]\mathbb{E}\left[z^Y\right]=e^{a(z-1)}e^{b(z-1)}=\exp\left((a+b)(z-1)\right).$$

This suffices to complete the proof. ■

**Lemma 13.2.** *Suppose that $\{N_i\}_{i=1}^\infty$ are independent Poisson random variables with parameters, $\{\lambda_i\}_{i=1}^\infty$ such that $\lambda:=\sum_{i=1}^\infty\lambda_i<\infty$. Then $N:=\sum_{i=1}^\infty N_i$ is Poisson with parameter $\lambda$.*

**Proof.** By Lemma 13.1, for each $n\in\mathbb{N}$, $\sum_{i=1}^n N_i\overset{d}{=}Pois\left(\sum_{i=1}^n\lambda_i\right)$. Since for each $k\in\mathbb{N}_0$, $\left\{\sum_{i=1}^n N_i=k\right\}\downarrow\{N=k\}$ as $n\uparrow\infty$ we have

$$\begin{aligned}P(N=k)&=\lim_{n\to\infty}P\left(\sum_{i=1}^n N_i=k\right)=\lim_{n\to\infty}\frac{\left(\sum_{i=1}^n\lambda_i\right)^k}{k!}\exp\left(-\sum_{i=1}^n\lambda_i\right)\\&=\frac{\lambda^k}{k!}e^{-\lambda}\end{aligned}$$

which shows $N\overset{d}{=}Pois(\lambda)$. ■

**Lemma 13.3.** *Suppose that $\{N_i\}_{i=1}^\infty$ are independent Poisson random variables with parameters, $\{\lambda_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty\lambda_i=\infty$. Then $\sum_{i=1}^\infty N_i=\infty$ a.s.*

**Proof.** From Figure 13.1 we see that $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$ for all $\lambda\geq0$. Therefore,

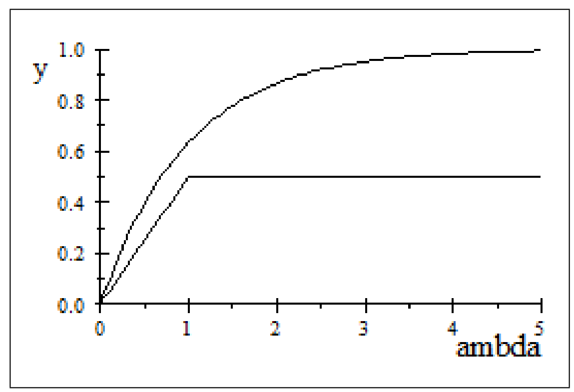



**Fig. 13.1.** This plot shows, $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$.

$$\sum_{i=1}^\infty P(N_i\geq1)=\sum_{i=1}^\infty\left(1-P(N_i=0)\right)=\sum_{i=1}^\infty\left(1-e^{-\lambda_i}\right)\geq\frac{1}{2}\sum_{i=1}^\infty\lambda_i\wedge1=\infty$$

and so by the second Borel Cantelli Lemma, $P(\{N_i \geq 1 \text{ i.o.}\}) = 1$. From this it certainly follows that $\sum_{i=1}^{\infty} N_i = \infty$ a.s.

**Alternatively**, let $\Lambda_n = \lambda_1 + \cdots + \lambda_n$, then

$$P\left(\sum_{i=1}^{\infty} N_i \geq k\right) \geq P\left(\sum_{i=1}^{n} N_i \geq k\right) = 1 - e^{-\Lambda_n} \sum_{l=0}^{k-1} \frac{\Lambda_n^l}{l!} \to 1 \text{ as } n \to \infty.$$

Therefore $P(\sum_{i=1}^{\infty} N_i \geq k) = 1$ for all $k \in \mathbb{N}$ and hence,

$$P\left(\sum_{i=1}^{\infty} N_i \geq \infty\right) = P\left(\cap_{k=1}^{\infty} \left\{\sum_{i=1}^{\infty} N_i \geq k\right\}\right) = 1.$$

■

## 13.2 Exponential Random Variables

Recall from Definition 9.67 that $T \stackrel{d}{=} E(\lambda)$ is an exponential random variable with parameter $\lambda \in [0, \infty)$ provided, $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$. We have seen that

$$\mathbb{E}\left[e^{aT}\right] = \frac{1}{1 - a\lambda^{-1}} \text{ for } a < \lambda. \tag{13.1}$$

$\mathbb{E}T = \lambda^{-1}$ and $\operatorname{Var}(T) = \lambda^{-2}$, and (see Theorem 9.68) that $T$ being exponential is characterized by the following memoryless property;

$$P(T > s + t | T > s) = P(T > t) \text{ for all } s, t \geq 0.$$

**Theorem 13.4.** *Let $\{T_j\}_{j=1}^{\infty}$ be independent random variables such that $T_j \stackrel{d}{=} E(\lambda_j)$ with $0 < \lambda_j < \infty$ for all $j$. Then:*

1. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ then $P(\sum_{n=1}^{\infty} T_n = \infty) = 0$ (i.e. $P(\sum_{n=1}^{\infty} T_n < \infty) = 1$).*
2. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $P(\sum_{n=1}^{\infty} T_n = \infty) = 1$.*

*(By Kolmogorov's zero-one law (see Proposition 12.53) it follows that $P(\sum_{n=1}^{\infty} T_n = \infty)$ is always either 0 or 1. We are showing here that $P(\sum_{n=1}^{\infty} T_n = \infty) = 1$ iff $\mathbb{E}[\sum_{n=1}^{\infty} T_n] = \infty$.)*

**Proof.** 1. Since

$$\mathbb{E}\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[T_n] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} T_n < \infty$ a.s., i.e. $P(\sum_{n=1}^{\infty} T_n = \infty) = 0$.

2. By the DCT, independence, and Eq. (13.1) with $a = -1$,

$$\begin{aligned}\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] &= \lim_{N\to\infty} \mathbb{E}\left[e^{-\sum_{n=1}^{N} T_n}\right] = \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{-T_n}\right] \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \left(\frac{1}{1+\lambda_n^{-1}}\right) = \prod_{n=1}^{\infty} (1 - a_n)\end{aligned}$$

where

$$a_n = 1 - \frac{1}{1+\lambda_n^{-1}} = \frac{1}{1+\lambda_n}.$$

Hence by Exercise 12.8, $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $\infty = \sum_{n=1}^{\infty} a_n$ which happens iff $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ as you should verify. This completes the proof since $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $e^{-\sum_{n=1}^{\infty} T_n} = 0$ a.s. or equivalently $\sum_{n=1}^{\infty} T_n = \infty$ a.s.

■

### 13.2.1 Appendix: More properties of Exponential random Variables*

**Theorem 13.5.** *Let $I$ be a countable set and let $\{T_k\}_{k\in I}$ be independent random variables such that $T_k \sim E(q_k)$ with $q := \sum_{k\in I} q_k \in (0, \infty)$. Let $T := \inf_k T_k$ and let $K = k$ on the set where $T_j > T_k$ for all $j \neq k$. On the complement of all these sets, define $K = *$ where $*$ is some point not in $I$. Then $P(K = *) = 0$, $K$ and $T$ are independent, $T \sim E(q)$, and $P(K = k) = q_k/q$.*

**Proof.** Let $k \in I$ and $t \in \mathbb{R}_+$ and $\Lambda_n \subset_f I$ such that $\Lambda_n \uparrow I \setminus \{k\}$, then

$$\begin{aligned}P(K = k, T > t) &= P(\cap_{j\neq k} \{T_j > T_k\}, T_k > t) = \lim_{n\to\infty} P(\cap_{j\in\Lambda_n} \{T_j > T_k\}, T_k > t) \\ &= \lim_{n\to\infty} \int_{[0,\infty)^{\Lambda_n \cup \{k\}}} \prod_{j\in\Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n \left(\{t_j\}_{j\in\Lambda_n}\right) q_k e^{-q_k t_k} dt_k\end{aligned}$$

where $\mu_n$ is the joint distribution of $\{T_j\}_{j\in\Lambda_n}$. So by Fubini's theorem,

$$
\begin{aligned}
P(K=k,T>t) &= \lim_{n\to\infty}\int_t^\infty q_k e^{-q_k t_k}dt_k \int_{[0,\infty)^{\Lambda_n}}\prod_{j\in\Lambda_n} 1_{t_j>t_k}\cdot 1_{t_k>t}d\mu_n\left(\{t_j\}_{j\in\Lambda_n}\right)\\
&= \lim_{n\to\infty}\int_t^\infty P\left(\cap_{j\in\Lambda_n}\{T_j>t_k\}\right)q_k e^{-q_k t_k}dt_k\\
&= \int_t^\infty P\left(\cap_{j\neq k}\{T_j>\tau\}\right)q_k e^{-q_k\tau}d\tau\\
&= \int_t^\infty \prod_{j\neq k} e^{-q_j\tau}q_k e^{-q_k\tau}d\tau = \int_t^\infty \prod_{j\in I} e^{-q_j\tau}q_k d\tau\\
&= \int_t^\infty e^{-\sum_{j=1}^\infty q_j\tau}q_k d\tau = \int_t^\infty e^{-q\tau}q_k d\tau = \frac{q_k}{q}e^{-qt}. \qquad (13.2)
\end{aligned}
$$

Taking $t=0$ shows that $P(K=k)=\frac{q_k}{q}$ and summing this on $k$ shows $P(K\in I)=1$ so that $P(K=*)=0$. Moreover summing Eq. (13.2) on $k$ now shows that $P(T>t)=e^{-qt}$ so that $T$ is exponential. Moreover we have shown that

$$P(K=k,T>t)=P(K=k)\,P(T>t)$$

proving the desired independence. ■

**Theorem 13.6.** *Suppose that $S\sim E(\lambda)$ and $R\sim E(\mu)$ are independent. Then for $t\geq 0$ we have*

$$\mu P(S\leq t<S+R)=\lambda P(R\leq t<R+S).$$

**Proof.** We have

$$
\begin{aligned}
\mu P(S\leq t<S+R) &= \mu\int_0^t \lambda e^{-\lambda s}P(t<s+R)\,ds\\
&= \mu\lambda\int_0^t e^{-\lambda s}e^{-\mu(t-s)}ds\\
&= \mu\lambda e^{-\mu t}\int_0^t e^{-(\lambda-\mu)s}ds = \mu\lambda e^{-\mu t}\cdot\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\
&= \mu\lambda\cdot\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}
\end{aligned}
$$

which is symmetric in the interchanged of $\mu$ and $\lambda$.**Alternatively:**

$$
\begin{aligned}
P(S\leq t<S+R) &= \lambda\mu\int_{\mathbb{R}_+^2} 1_{s\leq t<s+r}e^{-\lambda s}e^{-\mu r}dsdr\\
&= \lambda\mu\int_0^t ds\int_{t-s}^\infty dr e^{-\lambda s}e^{-\mu r}\\
&= \lambda\int_0^t ds e^{-\lambda s}e^{-\mu(t-s)}\\
&= \lambda e^{-\mu t}\int_0^t ds e^{-(\lambda-\mu)s}\\
&= \lambda e^{-\mu t}\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\
&= \lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.
\end{aligned}
$$

Therefore,

$$\mu P(S\leq t<S+R)=\mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}$$

which is symmetric in the interchanged of $\mu$ and $\lambda$ and hence

$$\lambda P(R\leq t<S+R)=\mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.$$

■

*Example 13.7.* Suppose $T$ is a positive random variable such that $P(T\geq t+s|T\geq s)=P(T\geq t)$ for all $s,t\geq 0$, or equivalently

$$P(T\geq t+s)=P(T\geq t)\,P(T\geq s)\ \text{ for all } s,t\geq 0,$$

then $P(T\geq t)=e^{-at}$ for some $a>0$. (Such exponential random variables are often used to model "waiting times.") The distribution function for $T$ is $F_T(t):=P(T\leq t)=1-e^{-a(t\vee 0)}$. Since $F_T(t)$ is piecewise differentiable, the law of $T$, $\mu:=P\circ T^{-1}$, has a density,

$$d\mu(t)=F_T'(t)\,dt=ae^{-at}1_{t\geq 0}dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right]=\int_0^\infty ae^{-at}e^{i\lambda t}dt=\frac{a}{a-i\lambda}=\hat{\mu}(\lambda).$$

Since

$$\hat{\mu}'(\lambda)=i\frac{a}{(a-i\lambda)^2}\ \text{ and }\ \hat{\mu}''(\lambda)=-2\frac{a}{(a-i\lambda)^3}$$

it follows that

$$\mathbb{E}T=\frac{\hat{\mu}'(0)}{i}=a^{-1}\ \text{ and }\ \mathbb{E}T^2=\frac{\hat{\mu}''(0)}{i^2}=\frac{2}{a^2}$$

and hence $\mathrm{Var}(T)=\frac{2}{a^2}-\left(\frac{1}{a}\right)^2=a^{-2}$.

## 13.3 The Standard Poisson Process

Let $\{T_k\}_{k=1}^{\infty}$ be an i.i.d. sequence of random exponential times with parameter $\lambda$, i.e. $P(T_k \in [t, t+dt]) = \lambda e^{-\lambda t} dt$. For each $n \in \mathbb{N}$ let $W_n := T_1 + \cdots + T_n$ be the "**waiting time**" for the $n^{\text{th}}$ event to occur. Because of Theorem 13.4 we know that $\lim_{n\to\infty} W_n = \infty$ a.s.

**Definition 13.8 (Poisson Process I).** *For any subset $A \subset \mathbb{R}_+$ let $N(A) := \sum_{n=1}^{\infty} 1_A(W_n)$ count the number of waiting times which occurred in $A$. When $A = (0,t]$ we will write, $N_t := N((0,t])$ for all $t \geq 0$ and refer to $\{N_t\}_{t\geq 0}$ as the **Poisson Process with intensity** $\lambda$. (Observe that $\{N_t = n\} = W_n \leq t < W_{n+1}$.)*

The next few results summarize a number of the basic properties of this Poisson process. Many of the proofs will be left as exercises to the reader. We will use the following notation below; for each $n \in \mathbb{N}$ and $T \geq 0$ let

$$\Delta_n(T) := \{(w_1, \dots, w_n) \in \mathbb{R}^n : 0 < w_1 < w_2 < \cdots < w_n < T\}$$

and let

$$\Delta_n := \cup_{T>0} \Delta_n(T) = \{(w_1, \dots, w_n) \in \mathbb{R}^n : 0 < w_1 < w_2 < \cdots < w_n < \infty\}.$$

(We equip each of these spaces with their Borel $\sigma -$ algebras.)

**Exercise 13.1.** Show $m_n(\Delta_n(T)) = T^n/n!$ where $m_n$ is Lebesgue measure on $\mathcal{B}_{\mathbb{R}^n}$.

**Exercise 13.2.** If $n \in \mathbb{N}$ and $g : \Delta_n \to \mathbb{R}$ bounded (non-negative) measurable, then

$$\mathbb{E}[g(W_1, \dots, W_n)] = \int_{\Delta_n} g(w_1, w_2, \dots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n. \tag{13.3}$$

As a simple corollary we have the following direct proof of Example 12.31.

**Corollary 13.9.** *If $n \in \mathbb{N}$, then $W_n \stackrel{d}{=} Gamma(n, \lambda^{-1})$.*

**Proof.** Taking $g(w_1, w_2, \dots, w_n) = f(w_n)$ in Eq. (13.3) we find with the aid of Exercise 13.1 that

$$\mathbb{E}[f(W_n)] = \int_{\Delta_n} f(w_n) \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n$$
$$= \int_0^{\infty} f(w) \lambda^n \frac{w^{n-1}}{(n-1)!} e^{-\lambda w} dw$$

which shows that $W_n \stackrel{d}{=} \text{Gamma}(n, \lambda^{-1})$. ∎

**Corollary 13.10.** *If $t \in \mathbb{R}_+$ and $f : \Delta_n(t) \to \mathbb{R}$ is a bounded (or non-negative) measurable function, then*

$$\mathbb{E}[f(W_1, \dots, W_n) : N_t = n] = \lambda^n e^{-\lambda t} \int_{\Delta_n(t)} f(w_1, w_2, \dots, w_n) dw_1 \dots dw_n. \tag{13.4}$$

**Proof.** Making use of the observation that $\{N_t = n\} = \{W_n \leq t < W_{n+1}\}$, we may apply Eq. (13.3) at level $n+1$ with

$$g(w_1, w_2, \dots, w_{n+1}) = f(w_1, w_2, \dots, w_n) 1_{w_n \leq t < w_{n+1}}$$

to learn

$$\mathbb{E}[f(W_1, \dots, W_n) : N_t = n]$$
$$= \int_{0<w_1<\cdots<w_n<t<w_{n+1}} f(w_1, w_2, \dots, w_n) \lambda^{n+1} e^{-\lambda w_{n+1}} dw_1 \dots dw_n dw_{n+1}$$
$$= \int_{\Delta_n(t)} f(w_1, w_2, \dots, w_n) \lambda^n e^{-\lambda t} dw_1 \dots dw_n.$$

∎

**Exercise 13.3.** Show $N_t \stackrel{d}{=} \text{Poi}(\lambda t)$ for all $t > 0$.

**Definition 13.11 (Order Statistics).** *Suppose that $X_1, \dots, X_n$ are non-negative random variables such that $P(X_i = X_j) = 0$ for all $i \neq j$. The **order statistics** of $X_1, \dots, X_n$ are the random variables, $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ defined by*

$$\tilde{X}_k = \min_{\#(\Lambda)=k} \max\{X_i : i \in \Lambda\} \tag{13.5}$$

*where $\Lambda$ always denotes a subset of $\{1, 2, \dots, n\}$ in Eq. (13.5).*

The reader should verify that $\tilde{X}_1 \leq \tilde{X}_2 \leq \cdots \leq \tilde{X}_n$, $\{X_1, \dots, X_n\} = \left\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\right\}$ with repetitions, and that $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n$ if $X_i \neq X_j$ for all $i \neq j$. In particular if $P(X_i = X_j) = 0$ for all $i \neq j$ then $P(\cup_{i\neq j}\{X_i = X_j\}) = 0$ and $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n$ a.s.

**Exercise 13.4.** Suppose that $X_1, \dots, X_n$ are non-negative[1] random variables such that $P(X_i = X_j) = 0$ for all $i \neq j$. Show;

[1] The non-negativity of the $X_i$ are not really necessary here but this is all we need to consider.

1. If $f : \Delta_n \to \mathbb{R}$ is bounded (non-negative) measurable, then
$$\mathbb{E}\left[f\left(\tilde{X}_1,\dots,\tilde{X}_n\right)\right] = \sum_{\sigma\in S_n} \mathbb{E}\left[f\left(X_{\sigma 1},\dots,X_{\sigma n}\right) : X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\right], \tag{13.6}$$
where $S_n$ is the permutation group on $\{1,2,\dots,n\}$.
2. If we further assume that $\{X_1,\dots,X_n\}$ are i.i.d. random variables, then
$$\mathbb{E}\left[f\left(\tilde{X}_1,\dots,\tilde{X}_n\right)\right] = n!\cdot\mathbb{E}\left[f\left(X_1,\dots,X_n\right) : X_1 < X_2 < \dots < X_n\right]. \tag{13.7}$$
(It is not important that $f\left(\tilde{X}_1,\dots,\tilde{X}_n\right)$ is not defined on the null set, $\cup_{i\neq j}\{X_i = X_j\}$.)
3. $f : \mathbb{R}_+^n \to \mathbb{R}$ is a bounded (non-negative) measurable symmetric function (i.e. $f(w_{\sigma 1},\dots,w_{\sigma n}) = f(w_1,\dots,w_n)$ for all $\sigma\in S_n$ and $(w_1,\dots,w_n)\in\mathbb{R}_+^n$) then
$$\mathbb{E}\left[f\left(\tilde{X}_1,\dots,\tilde{X}_n\right)\right] = \mathbb{E}\left[f\left(X_1,\dots,X_n\right)\right].$$
4. Suppose that $Y_1,\dots,Y_n$ is another collection of non-negative random variables such that $P(Y_i = Y_j) = 0$ for all $i\neq j$ such that
$$\mathbb{E}\left[f\left(X_1,\dots,X_n\right)\right] = \mathbb{E}\left[f\left(Y_1,\dots,Y_n\right)\right]$$
for all bounded (non-negative) measurable symmetric functions from $\mathbb{R}_+^n \to \mathbb{R}$. Show that $\left(\tilde{X}_1,\dots,\tilde{X}_n\right) \overset{d}{=} \left(\tilde{Y}_1,\dots,\tilde{Y}_n\right)$.
**Hint:** if $g : \Delta_n \to \mathbb{R}$ is a bounded measurable function, define $f : \mathbb{R}_+^n \to \mathbb{R}$ by;
$$f(y_1,\dots,y_n) = \sum_{\sigma\in S_n} 1_{y_{\sigma 1}<y_{\sigma 2}<\dots<y_{\sigma n}}\, g\left(y_{\sigma 1},y_{\sigma 2},\dots,y_{\sigma n}\right)$$
and then show $f$ is symmetric.

**Exercise 13.5.** Let $t\in\mathbb{R}_+$ and $\{U_i\}_{i=1}^n$ be i.i.d. uniformly distributed random variables on $[0,t]$. Show that the order statistics, $\left(\tilde{U}_1,\dots,\tilde{U}_n\right)$, of $(U_1,\dots,U_n)$ has the same distribution as $(W_1,\dots,W_n)$ given $N_t = n$. (Thus, given $N_t = n$, the collection of points, $\{W_1,\dots,W_n\}$, has the same distribution as the collection of points, $\{U_1,\dots,U_n\}$, in $[0,t]$.)

**Theorem 13.12 (Joint Distributions).** *If $\{A_i\}_{i=1}^k \subset \mathcal{B}_{[0,t]}$ is a partition of $[0,t]$, then $\{N(A_i)\}_{i=1}^k$ are independent random variables and $N(A) \overset{d}{=} \operatorname{Poi}(\lambda m(A))$ for all $A\in\mathcal{B}_{[0,t]}$ with $m(A)<\infty$. In particular, if $0<t_1<t_2<\dots<t_n$, then $\left\{N_{t_i}-N_{t_{i-1}}\right\}_{i=1}^n$ are independent random variables and $N_t - N_s \overset{d}{=} \operatorname{Poi}(\lambda(t-s))$ for all $0\le s<t<\infty$. (We say that $\{N_t\}_{t\ge 0}$ is a stochastic process with **independent increments**.)*

**Proof.** If $z\in\mathbb{C}$ and $A\in\mathcal{B}_{[0,t]}$, then
$$z^{N(A)} = z^{\sum_{i=1}^n 1_A(W_i)} \text{ on } \{N_t = n\}.$$
Let $n\in\mathbb{N}$, $z_i\in\mathbb{C}$, and define
$$f(w_1,\dots,w_n) = z_1^{\sum_{i=1}^n 1_{A_1}(w_i)}\dots z_k^{\sum_{i=1}^n 1_{A_k}(w_i)}$$
which is a symmetric function. On $N_t = n$ we have,
$$z_1^{N(A_1)}\dots z_k^{N(A_k)} = f(W_1,\dots,W_n)$$
and therefore,
$$\begin{aligned}\mathbb{E}\left[z_1^{N(A_1)}\dots z_k^{N(A_k)}|N_t = n\right] &= \mathbb{E}\left[f(W_1,\dots,W_n)|N_t = n\right]\\ &= \mathbb{E}\left[f(U_1,\dots,U_n)\right]\\ &= \mathbb{E}\left[z_1^{\sum_{i=1}^n 1_{A_1}(U_i)}\dots z_k^{\sum_{i=1}^n 1_{A_k}(U_i)}\right]\\ &= \prod_{i=1}^n \mathbb{E}\left[\left(z_1^{1_{A_1}(U_i)}\dots z_k^{1_{A_k}(U_i)}\right)\right]\\ &= \left(\mathbb{E}\left[\left(z_1^{1_{A_1}(U_1)}\dots z_k^{1_{A_k}(U_1)}\right)\right]\right)^n\\ &= \left(\frac{1}{t}\sum_{i=1}^k m(A_i)\cdot z_i\right)^n,\end{aligned}$$
wherein we have made use of the fact that $\{A_i\}_{i=1}^n$ is a partition of $[0,t]$ so that
$$z_1^{1_{A_1}(U_1)}\dots z_k^{1_{A_k}(U_1)} = \sum_{i=1}^k z_i 1_{A_i}(U_i).$$
Thus it follows that

$$\begin{aligned}\mathbb{E}\left[z_1^{N(A_1)}\dots z_k^{N(A_k)}\right] &= \sum_{n=0}^{\infty}\mathbb{E}\left[z_1^{N(A_1)}\dots z_k^{N(A_k)}|N_t=n\right]P\left(N_t=n\right)\\ &= \sum_{n=0}^{\infty}\left(\frac{1}{t}\sum_{i=1}^{k}m\left(A_i\right)\cdot z_i\right)^n\frac{(\lambda t)^n}{n!}e^{-\lambda t}\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\left(\lambda\sum_{i=1}^{k}m\left(A_i\right)\cdot z_i\right)^n e^{-\lambda t}\\ &= \exp\left(\lambda\left[\sum_{i=1}^{k}m\left(A_i\right)z_i-t\right]\right)\\ &= \exp\left(\lambda\left[\sum_{i=1}^{k}m\left(A_i\right)\left(z_i-1\right)\right]\right).\end{aligned}$$

From this result it follows that $\{N(A_i)\}_{i=1}^n$ are independent random variables and $N(A)=\mathrm{Poi}(\lambda m(A))$ for all $A\in\mathcal{B}_{\mathbb{R}}$ with $m(A)<\infty$.

**Alternatively;** suppose that $a_i\in\mathbb{N}_0$ and $n:=a_1+\dots+a_k$, then

$$\begin{aligned}P\left[N(A_1)=a_1,\dots,N(A_k)=a_k|N_t=n\right] &= P\left[\sum_{i=1}^{n}1_{A_l}\left(U_i\right)=a_l \text{ for } 1\le l\le k\right]\\ &= \frac{n!}{a_1!\dots a_k!}\prod_{l=1}^{k}\left[\frac{m\left(A_l\right)}{t}\right]^{a_l}\\ &= \frac{n!}{t^n}\cdot\prod_{l=1}^{k}\frac{\left[m\left(A_l\right)\right]^{a_l}}{a_l!}\end{aligned}$$

and therefore,

$$\begin{aligned}&P\left[N(A_1)=a_1,\dots,N(A_k)=a_k\right]\\ &=P\left[N(A_1)=a_1,\dots,N(A_k)=a_k|N_t=n\right]\cdot P\left(N_t=n\right)\\ &=\frac{n!}{t^n}\cdot\prod_{l=1}^{k}\frac{\left[m\left(A_l\right)\right]^{a_l}}{a_l!}\cdot e^{-\lambda t}\frac{(\lambda t)^n}{n!}\\ &=\prod_{l=1}^{k}\frac{\left[m\left(A_l\right)\right]^{a_l}}{a_l!}\cdot e^{-\lambda t}\lambda^n\\ &=\prod_{l=1}^{k}\frac{\left[m\left(A_l\right)\lambda\right]^{a_l}}{a_l!}e^{-\lambda a_l}\end{aligned}$$

which shows that $\{N(A_l)\}_{l=1}^k$ are independent and that $N(A_l)\overset{d}{=}\mathrm{Poi}(\lambda m(A_l))$ for each $l$. ∎

*Remark 13.13.* If $A\in\mathcal{B}_{[0,\infty)}$ with $m(A)=\infty$, then $N(A)=\infty$ a.s. To prove this observe that $N(A)=\uparrow\lim_{n\to\infty}N(A\cap[0,n])$. Therefore for any $k\in\mathbb{N}$, we have

$$\begin{aligned}P\left(N(A)\ge k\right)&\ge P\left(N(A\cap[0,n])\ge k\right)\\ &=1-e^{-\lambda m(A\cap[0,n])}\sum_{0\le l<k}\frac{\left(\lambda m\left(A\cap[0,n]\right)\right)^l}{l!}\to 1 \text{ as } n\to\infty.\end{aligned}$$

This shows that $N(A)\ge k$ a.s. for all $k\in\mathbb{N}$, i.e. $N(A)=\infty$ a.s.

**Exercise 13.6 (A Generalized Poisson Process I).** Suppose that $(S,\mathcal{B}_S,\mu)$ is a finite measure space with $\mu(S)<\infty$. Define $\Omega=\sum_{n=0}^{\infty}S^n$ where $S^0=\{*\}$, were $*$ is some arbitrary point. Define $\mathcal{B}_\Omega$ to be those sets, $B=\sum_{n=0}^{\infty}B_n$ where $B_n\in\mathcal{B}_{S^n}:=\mathcal{B}_S^{\otimes n}$ – the product $\sigma$ – algebra on $S^n$. Now define a probability measure, $P$, on $(\Omega,\mathcal{B}_\Omega)$ by

$$P(B):=e^{-\mu(S)}\sum_{n=0}^{\infty}\frac{1}{n!}\mu^{\otimes n}\left(B_n\right)$$

where $\mu^{\otimes 0}(\{*\})=1$ by definition. (We denote $P$ schematically by $P:=e^{-\mu(S)}e^{\mu\otimes}$.) Finally for ever $\omega\in\Omega$, let $N_\omega$, be the point measure on $(S,\mathcal{B}_S)$ defined by; $N_*=0$ and

$$N_\omega=\sum_{i=1}^{n}\delta_{s_i} \text{ if } \omega=(s_1,\dots,s_n)\in S^n \text{ for } n\ge 1.$$

So for $A\in\mathcal{B}_S$, we have $N_*(A)=0$ and $N_\omega(A)=\sum_{i=1}^n 1_A(s_i)$. Show;

1. For each $A\in\mathcal{B}_S$, $\omega\to N_\omega(A)$ is a Poisson random variable with intensity $\mu(A)$, i.e. $N(A)=\mathrm{Poi}(\mu(A))$.
2. If $\{A_k\}_{k=1}^m\subset\mathcal{B}_S$ are disjoint sets, the $\{\omega\to N_\omega(A_k)\}_{k=1}^m$ are independent random variables.

An integer valued random measure on $(S,\mathcal{B}_S)$ ($\Omega\ni\omega\to N_\omega$) satisfying properties 1. and 2. of Exercise 13.6 is called a **Poisson process on** $(S,\mathcal{B}_S)$ **with intensity measure** $\mu$. For more motivation as to why Poisson processes are important see Proposition 23.11 and/or Remark 23.12 below.

**Exercise 13.7 (A Generalized Poisson Process II).** Let $(S,\mathcal{B}_S,\mu)$ be as in Exercise 13.6, $\{Y_i\}_{i=1}^\infty$ be i.i.d. $S$ – valued Random variables with $\mathrm{Law}_P(Y_i)=\mu(\cdot)/\mu(S)$ and $\nu$ be a $\mathrm{Poi}(\mu(S))$ – random variable which is independent of $\{Y_i\}$. Show $N:=\sum_{i=1}^{\nu}\delta_{Y_i}$ is a Poisson process on $(S,\mathcal{B}_S)$ with intensity measure, $\mu$. **Hints:**

1. Assume that $\{A_k\}_{k=1}^m \subset \mathcal{B}_S$ is a measurable partition of $S$ and show $\{N(A_k)\}_{k=1}^m$ are i.i.d. with $N(A_k) = \mathrm{Poi}(\mu(A_k))$ for each $k$.
2. Model your proof of item 1. on either of the proofs of Theorem 13.12.

**Exercise 13.8 (A Generalized Poisson Process III).** Suppose now that $(S, \mathcal{B}_S, \mu)$ is a $\sigma$ – finite measure space and $S = \sum_{l=1}^\infty S_l$ is a partition of $S$ such that $0 < \mu(S_l) < \infty$ for all $l$. For each $l \in \mathbb{N}$, using either of the construction above we may construct a Poisson point process, $N_l$, on $(S, \mathcal{B}_S)$ with intensity measure, $\mu_l$ where $\mu_l(A) := \mu(A \cap S_l)$ for all $A \in \mathcal{B}_S$. We do this in such a what that $\{N_l\}_{l=1}^\infty$ are all **independent**. Show that $N := \sum_{l=1}^\infty N_l$ is a Poisson point process on $(S, \mathcal{B}_S)$ with intensity measure, $\mu$. To be more precise observe that $N$ is a random measure on $(S, \mathcal{B}_S)$ which satisfies (as you should show);

1. For each $A \in \mathcal{B}_S$ with $\mu(A) < \infty$, show $N(A) \overset{d}{=} \mathrm{Poi}(\mu(A))$. Also show $N(A) = \infty$ a.s. if $\mu(A) = \infty$.
2. If $\{A_k\}_{k=1}^m \subset \mathcal{B}_S$ are disjoint sets with $\mu(A_k) < \infty$, show $\{N(A_k)\}_{k=1}^m$ are independent random variables.

## 13.4 Poission Process Extras*

(This subsection still needs work!) In Definition 13.8 we really gave a construction of a Poisson process as defined in Definition 13.14. The goal of this section is to show that the Poisson process, $\{N_t\}_{t\ge0}$, as defined in Definition 13.14 is uniquely determined and is essentially equivalent to what we have already done above.

**Definition 13.14 (Poisson Process II).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $N_t : \Omega \to \mathbb{N}_0$ be a random variable for each $t \ge 0$. We say that $\{N_t\}_{t\ge0}$ is a Poisson process with intensity $\lambda$ if; 1) $N_0 = 0$, 2) $N_t - N_s \overset{d}{=} \mathrm{Poi}(\lambda(t-s))$ for all $0 \le s < t < \infty$, 3) $\{N_t\}_{t\ge0}$ has independent increments, and 4) $t \to N_t(\omega)$ is right continuous and non-decreasing for all $\omega \in \Omega$.*

Let $N_\infty(\omega) :=\uparrow \lim_{t\uparrow\infty} N_t(\omega)$ and observe that $N_\infty = \sum_{k=0}^\infty (N_k - N_{k-1}) = \infty$ a.s. by Lemma 13.3. Therefore, we may and do assume that $N_\infty(\omega) = \infty$ for all $\omega \in \Omega$.

**Lemma 13.15.** *There is zero probability that $\{N_t\}_{t\ge0}$ makes a jump greater than or equal to 2.*

**Proof.** Suppose that $T \in (0, \infty)$ is fixed and $\omega \in \Omega$ is sample point where $t \to N_t(\omega)$ makes a jump of 2 or more for $t \in [0, T]$. Then for all $n \in \mathbb{N}$ we must have $\omega \in \cup_{k=1}^n \left\{N_{\frac{k}{n}T} - N_{\frac{k-1}{n}T} \ge 2\right\}$. Therefore,

$$
\begin{aligned}
&P^*(\{\omega : [0, T] \ni t \to N_t(\omega) \text{ has jump } \ge 2\}) \\
&\quad\le \sum_{k=1}^n P\left(N_{\frac{k}{n}T} - N_{\frac{k-1}{n}T} \ge 2\right) = \sum_{k=1}^n O\left(T^2/n^2\right) = O(1/n) \to 0
\end{aligned}
$$

as $n \to \infty$. I am leaving open the possibility that the set of $\omega$ where a jump size 2 or larger is not measurable. ■

**Theorem 13.16.** *Suppose that $\{N_t\}_{t\ge0}$ is a Poisson process with intensity $\lambda$ as in Definition 13.14,*

$$W_n := \inf\{t : N_t = n\} \text{ for all } n \in \mathbb{N}_0$$

*be the first time $N_t$ reaches $n$. (The $\{W_n\}_{n=0}^\infty$ are well defined off a set of measure zero and $W_n < W_{n+1}$ for all $n$ by the right continuity of $\{N_t\}_{t\ge0}$.) Then the $\{T_n := W_n - W_{n-1}\}_{n=1}^\infty$ are i.i.d. $E(\lambda)$ – random variables. Thus the two descriptions of a Poisson process given in Definitions 13.8 and 13.14 are equivalent.*

**Proof.** Suppose that $J_i = (a_i, b_i]$ with $b_i \le a_{i+1} < \infty$ for all $i$. We will begin by showing

$$
\begin{aligned}
P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) &= \lambda^n \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} e^{-\lambda w_n} dw_n &(13.8)\\
&= \lambda^n \int_{J_1\times J_2\times\dots\times J_n} e^{-\lambda w_n} dw_1 \dots dw_n. &(13.9)
\end{aligned}
$$

To show this let $K_i := (b_{i-1}, a_i]$ where $b_0 = 0$. Then

$$\cap_{i=1}^n \{W_i \in J_i\} = \cap_{i=1}^n \{N(K_i) = 0\} \cap \cap_{i=1}^{n-1} \{N(J_i) = 0\} \cap \{N(J_n) \ge 2\}$$

and therefore,

$$
\begin{aligned}
P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) &= \prod_{i=1}^n e^{-\lambda m(K_i)} \cdot \prod_{i=1}^{n-1} e^{-\lambda m(J_i)} \lambda m(J_i) \cdot \left(1 - e^{-\lambda m(J_n)}\right) \\
&= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \left[e^{-\lambda a_n} - e^{-\lambda b_n}\right] \\
&= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} \lambda e^{-\lambda w_n} dw_n.
\end{aligned}
$$

We may now apply a $\pi$ – $\lambda$ – argument, using $\sigma(\{J_1 \times \dots \times J_n\}) = \mathcal{B}_{\Delta_n}$, to show

$$\mathbb{E}\left[g\left(W_1,\ldots,W_n\right)\right]=\int_{\Delta_n} g\left(w_1,\ldots,w_n\right)\lambda^n e^{-\lambda w_n}dw_1\ldots dw_n$$

holds for all bounded $\mathcal{B}_{\Delta_n}/\mathcal{B}_{\mathbb{R}}$ measurable functions, $g:\Delta_n\to\mathbb{R}$. Undoing the change of variables you made in Exercise 13.2 allows us to conclude that $\{T_n\}_{n=1}^{\infty}$ are i.i.d. $E\left(\lambda\right)$ – distributed random variables. ■

# 14

# $L^p$ – spaces

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and for $0 < p < \infty$ and a measurable function $f : \Omega \to \mathbb{C}$ let

$$\|f\|_p := \left( \int_\Omega |f|^p \, d\mu \right)^{1/p} \tag{14.1}$$

and when $p = \infty$, let

$$\|f\|_\infty = \inf \{a \geq 0 : \mu(|f| > a) = 0\} \tag{14.2}$$

For $0 < p \leq \infty$, let

$$L^p(\Omega, \mathcal{B}, \mu) = \{f : \Omega \to \mathbb{C} : f \text{ is measurable and } \|f\|_p < \infty\} / \sim$$

where $f \sim g$ iff $f = g$ a.e. Notice that $\|f - g\|_p = 0$ iff $f \sim g$ and if $f \sim g$ then $\|f\|_p = \|g\|_p$. In general we will (by abuse of notation) use $f$ to denote both the function $f$ and the equivalence class containing $f$.

*Remark 14.1.* Suppose that $\|f\|_\infty \leq M$, then for all $a > M$, $\mu(|f| > a) = 0$ and therefore $\mu(|f| > M) = \lim_{n\to\infty} \mu(|f| > M + 1/n) = 0$, i.e. $|f(\omega)| \leq M$ for $\mu$ - a.e. $\omega$. Conversely, if $|f| \leq M$ a.e. and $a > M$ then $\mu(|f| > a) = 0$ and hence $\|f\|_\infty \leq M$. This leads to the identity:

$$\|f\|_\infty = \inf \{a \geq 0 : |f(\omega)| \leq a \text{ for } \mu - \text{a.e. } \omega\}.$$

## 14.1 Modes of Convergence

Let $\{f_n\}_{n=1}^\infty \cup \{f\}$ be a collection of complex valued measurable functions on $\Omega$. We have the following notions of convergence and Cauchy sequences.

**Definition 14.2.** *1. $f_n \to f$ a.e. if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} 1_{E^c} f_n = 1_{E^c} f$.*
*2. $f_n \to f$ in $\mu$ – measure if $\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0$ for all $\varepsilon > 0$. We will abbreviate this by saying $f_n \to f$ in $L^0$ or by $f_n \xrightarrow{\mu} f$.*
*3. $f_n \to f$ in $L^p$ iff $f \in L^p$ and $f_n \in L^p$ for all $n$, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.*

**Definition 14.3.** *1. $\{f_n\}$ is a.e. Cauchy if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $\{1_{E^c} f_n\}$ is a pointwise Cauchy sequences.*
*2. $\{f_n\}$ is Cauchy in $\mu$ – measure (or $L^0$ – Cauchy) if $\lim_{m,n\to\infty} \mu(|f_n - f_m| > \varepsilon) = 0$ for all $\varepsilon > 0$.*
*3. $\{f_n\}$ is Cauchy in $L^p$ if $\lim_{m,n\to\infty} \|f_n - f_m\|_p = 0$.*

When $\mu$ is a probability measure, we describe, $f_n \xrightarrow{\mu} f$ as $f_n$ **converging to $f$ in probability.** If a sequence $\{f_n\}_{n=1}^\infty$ is $L^p$ – convergent, then it is $L^p$ – Cauchy. For example, when $p \in [1, \infty]$ and $f_n \to f$ in $L^p$, we have (using Minkowski's inequality of Theorem 14.22 below)

$$\|f_n - f_m\|_p \leq \|f_n - f\|_p + \|f - f_m\|_p \to 0 \text{ as } m, n \to \infty.$$

The case where $p = 0$ will be handled in Theorem 14.8 below.

**Lemma 14.4 ($L^p$ – convergence implies convergence in probability).** *Let $p \in [1, \infty)$. If $\{f_n\} \subset L^p$ is $L^p$ – convergent (Cauchy) then $\{f_n\}$ is also convergent (Cauchy) in measure.*

**Proof.** By Chebyshev's inequality (9.2),

$$\mu(|f| \geq \varepsilon) = \mu(|f|^p \geq \varepsilon^p) \leq \frac{1}{\varepsilon^p} \int_\Omega |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_p^p$$

and therefore if $\{f_n\}$ is $L^p$ – Cauchy, then

$$\mu(|f_n - f_m| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \|f_n - f_m\|_p^p \to 0 \text{ as } m, n \to \infty$$

showing $\{f_n\}$ is $L^0$ – Cauchy. A similar argument holds for the $L^p$ – convergent case. ∎

*Example 14.5.* Let us consider a number of examples here to get a feeling for these different notions of convergence. In each of these examples we will work in the measure space, $\left(\mathbb{R}_+, \mathcal{B} = \mathcal{B}_{\mathbb{R}_+}, m\right)$.

1. Let $f_n = \frac{1}{n} 1_{[0,n]}$ as in Figure 14.1. In this case $f_n \nrightarrow 0$ in $L^1$ but $f_n \to 0$ a.e.,$f_n \to 0$ in $L^p$ for all $p > 1$ and $f_n \xrightarrow{m} 0$.

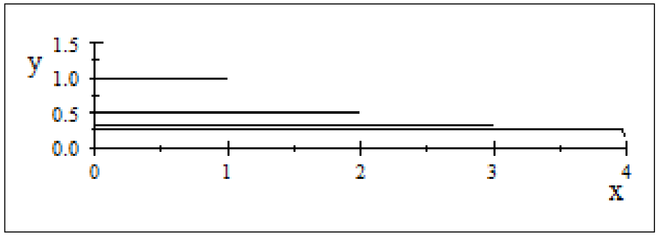
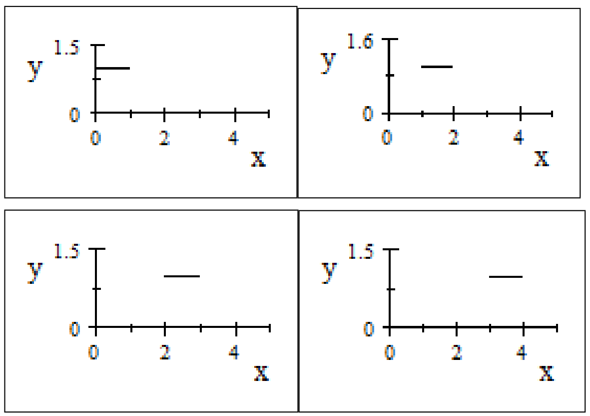



**Fig. 14.1.** Graphs of $f_n = \frac{1}{n} 1_{[0,n]}$ for $n = 1, 2, 3, 4$.

2. Let $f_n = 1_{[n-1,n]}$ as in the figure below. Then $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in any $L^p$ –space or in measure.



3. Now suppose that $f_n = n \cdot 1_{[0,1/n]}$ as in Figure 14.2. In this case $f_n \to 0$ a.e., $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ in $L^1$ or in any $L^p$ for $p \geq 1$. Observe that $\|f_n\|_p = n^{1-1/p}$ for all $p \geq 1$.

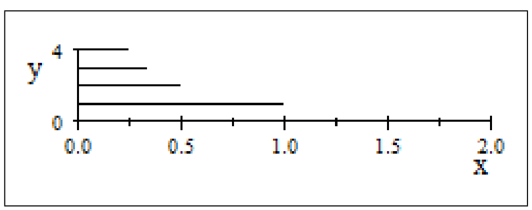
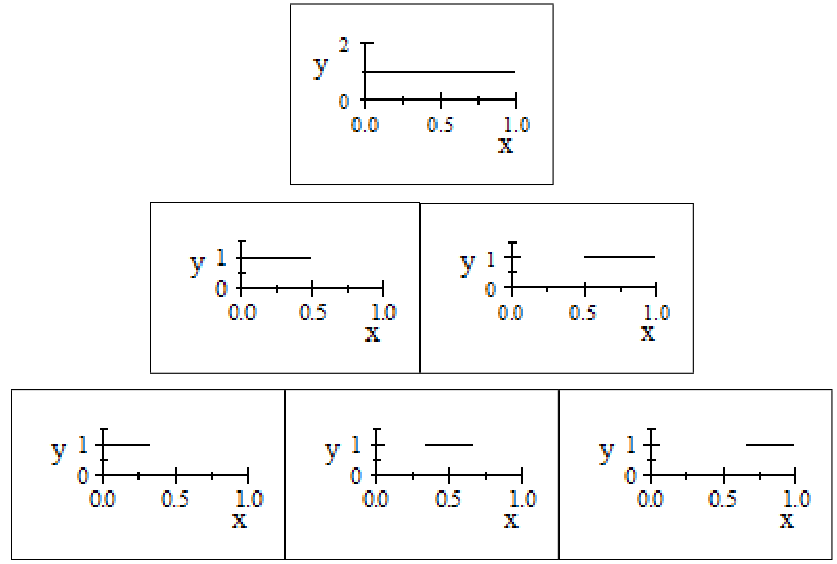



**Fig. 14.2.** Graphs of $f_n = n \cdot 1_{[0,n]}$ for $n = 1, 2, 3, 4$.

4. For $n \in \mathbb{N}$ and $1 \leq k \leq n$, let $g_{n,k} := 1_{\left(\frac{k-1}{n}, \frac{k}{n}\right]}$. Then define $\{f_n\}$ as

$$(f_1, f_2, f_3, \dots) = (g_{1,1}, g_{2,1}, g_{2,2}, g_{3,1}, g_{3,2}, g_{3,3}, g_{4,1}, g_{4,2}, g_{4,3}, g_{4,4}, \dots)$$

as depicted in the figures below.



For this sequence of functions we have $f_n \to 0$ in $L^p$ for all $1 \leq p < \infty$ and $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ a.e. and $f_n \nrightarrow 0$ in $L^\infty$. In this case, $\|g_{n,k}\|_p = \left(\frac{1}{n}\right)^{1/p}$ for $1 \leq p < \infty$ while $\|g_{n,k}\|_\infty = 1$ for all $n, k$.

## 14.2 Almost Everywhere and Measure Convergence

**Theorem 14.6 (Egorov: *a.s.* $\implies$ convergence in probability).** *Suppose $\mu(\Omega) = 1$ and $f_n \to f$ a.s. Then for all $\varepsilon > 0$ there exists $E = E_\varepsilon \in \mathcal{B}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on $E^c$. In particular $f_n \xrightarrow{\mu} f$ as $n \to \infty$.*

**Proof.** Let $f_n \to f$ a.e. Then for all $\varepsilon > 0$,

$$\begin{aligned} 0 &= \mu(\{|f_n - f| > \varepsilon \text{ i.o. } n\}) \\ &= \lim_{N\to\infty} \mu\left(\bigcup_{n \geq N} \{|f_n - f| > \varepsilon\}\right) \qquad (14.3) \\ &\geq \limsup_{N\to\infty} \mu\left(\{|f_N - f| > \varepsilon\}\right) \end{aligned}$$

from which it follows that $f_n \xrightarrow{\mu} f$ as $n \to \infty$.

We now prove that the convergence is uniform off a small exceptional set. By Eq. (14.3), there exists an increasing sequence $\{N_k\}_{k=1}^\infty$, such that $\mu(E_k) < \varepsilon 2^{-k}$, where

$$E_k := \bigcup_{n \geq N_k} \left\{ |f_n - f| > \frac{1}{k} \right\}.$$

If we now set $E := \cup_{k=1}^{\infty} E_k$, then $\mu(E) < \sum_k \varepsilon 2^{-k} = \varepsilon$ and for $\omega \notin E$ we have $|f_n(\omega) - f(\omega)| \leq \frac{1}{k}$ for all $n \geq N_k$ and $k \in \mathbb{N}$. That is $f_n \to f$ uniformly on $E^c$. ■

**Lemma 14.7.** *Suppose $a_n \in \mathbb{C}$ and $|a_{n+1} - a_n| \leq \varepsilon_n$ and $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in \mathbb{C}$ exists and $|a - a_n| \leq \delta_n := \sum_{k=n}^{\infty} \varepsilon_k$.*

**Proof.** Let $m > n$ then

$$|a_m - a_n| = \left| \sum_{k=n}^{m-1} (a_{k+1} - a_k) \right| \leq \sum_{k=n}^{m-1} |a_{k+1} - a_k| \leq \sum_{k=n}^{\infty} \varepsilon_k := \delta_n. \tag{14.4}$$

So $|a_m - a_n| \leq \delta_{\min(m,n)} \to 0$ as $, m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (14.4) to find $|a - a_n| \leq \delta_n$. ■

**Theorem 14.8.** *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions on $\Omega$.*

1. *If $f$ and $g$ are measurable functions and $f_n \xrightarrow{\mu} f$ and $f_n \xrightarrow{\mu} g$ then $f = g$ a.e.*
2. *If $f_n \xrightarrow{\mu} f$ and $g_n \xrightarrow{\mu} g$ then $\lambda f_n \to \lambda f$ for all $\lambda \in \mathbb{C}$ and $f_n + g_n \xrightarrow{\mu} f + g$.*
3. *If $f_n \xrightarrow{\mu} f$ then $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure.*
4. *If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure, there exists a measurable function, $f$, and a subsequence $g_j = f_{n_j}$ of $\{f_n\}$ such that $\lim_{j\to\infty} g_j := f$ exists a.e.*
5. *(**Completeness of convergence in measure.**) If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure and $f$ is as in item 4. then $f_n \xrightarrow{\mu} f$.*

**Proof.** One of the basic tricks here is to observe that if $\varepsilon > 0$ and $a, b \geq 0$ such that $a + b \geq \varepsilon$, then either $a \geq \varepsilon/2$ or $b \geq \varepsilon/2$.

1. Suppose that $f$ and $g$ are measurable functions such that $f_n \xrightarrow{\mu} g$ and $f_n \xrightarrow{\mu} f$ as $n \to \infty$ and $\varepsilon > 0$ is given. Since
$$|f - g| \leq |f - f_n| + |f_n - g|,$$
if $\varepsilon > 0$ and $|f - g| \geq \varepsilon$, then either $|f - f_n| \geq \varepsilon/2$ or $|f_n - g| \geq \varepsilon/2$. Thus it follows
$$\{|f - g| > \varepsilon\} \subset \{|f - f_n| > \varepsilon/2\} \cup \{|g - f_n| > \varepsilon/2\},$$
and therefore,
$$\mu(|f - g| > \varepsilon) \leq \mu(|f - f_n| > \varepsilon/2) + \mu(|g - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$
Hence
$$\mu(|f - g| > 0) = \mu\left( \cup_{n=1}^{\infty} \left\{ |f - g| > \frac{1}{n} \right\} \right) \leq \sum_{n=1}^{\infty} \mu\left( |f - g| > \frac{1}{n} \right) = 0,$$
i.e. $f = g$ a.e.
2. The first claim is easy and the second follows similarly to the proof of the first item.
3. Suppose $f_n \xrightarrow{\mu} f$, $\varepsilon > 0$ and $m, n \in \mathbb{N}$, then $|f_n - f_m| \leq |f - f_n| + |f_m - f|$. So by the basic trick,
$$\mu(|f_n - f_m| > \varepsilon) \leq \mu(|f_n - f| > \varepsilon/2) + \mu(|f_m - f| > \varepsilon/2) \to 0 \text{ as } m, n \to \infty.$$
4. Suppose $\{f_n\}$ is $L^0(\mu)$ – Cauchy and let $\varepsilon_n > 0$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^{\infty} \varepsilon_k$. Choose $g_j = f_{n_j}$ where $\{n_j\}$ is a subsequence of $\mathbb{N}$ such that
$$\mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \leq \varepsilon_j.$$
Let
$$F_N := \cup_{j \geq N} \{|g_{j+1} - g_j| > \varepsilon_j\} \text{ and}$$
$$E := \cap_{N=1}^{\infty} F_N = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}.$$
Since
$$\mu(F_N) \leq \delta_N < \infty$$
and $F_N \downarrow E$ it follows[1] that $0 = \mu(E) = \lim_{N\to\infty} \mu(F_N)$. For $\omega \notin E$, $|g_{j+1}(\omega) - g_j(\omega)| \leq \varepsilon_j$ for a.a. $j$ and so by Lemma 14.7, $f(\omega) := \lim_{j\to\infty} g_j(\omega)$ exists. For $\omega \in E$ we may define $f(\omega) \equiv 0$.
5. Next we will show $g_N \xrightarrow{\mu} f$ as $N \to \infty$ where $f$ and $g_N$ are as above. If
$$\omega \in F_N^c = \cap_{j \geq N} \{|g_{j+1} - g_j| \leq \varepsilon_j\},$$
then
$$|g_{j+1}(\omega) - g_j(\omega)| \leq \varepsilon_j \text{ for all } j \geq N.$$
Another application of Lemma 14.7 shows $|f(\omega) - g_j(\omega)| \leq \delta_j$ for all $j \geq N$, i.e.

[1] Alternatively, $\mu(E) = 0$ by the first Borel Cantelli lemma and the fact that $\sum_{j=1}^{\infty} \mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \leq \sum_{j=1}^{\infty} \varepsilon_j < \infty$.

$$F_N^c \subset \cap_{j\geq N}\{|f-g_j|\leq\delta_j\}\subset\{|f-g_N|\leq\delta_N\}.$$

Therefore, by taking complements of this equation, $\{|f-g_N|>\delta_N\}\subset F_N$ and hence

$$\mu(|f-g_N|>\delta_N)\leq\mu(F_N)\leq\delta_N\to 0 \text{ as } N\to\infty$$

and in particular, $g_N\overset{\mu}{\to}f$ as $N\to\infty$.

With this in hand, it is straightforward to show $f_n\overset{\mu}{\to}f$. Indeed, by the usual trick, for all $j\in\mathbb{N}$,

$$\mu(\{|f_n-f|>\varepsilon\})\leq\mu(\{|f-g_j|>\varepsilon/2\})+\mu(|g_j-f_n|>\varepsilon/2).$$

Therefore, letting $j\to\infty$ in this inequality gives,

$$\mu(\{|f_n-f|>\varepsilon\})\leq\limsup_{j\to\infty}\mu(|g_j-f_n|>\varepsilon/2)\to 0 \text{ as } n\to\infty,$$

wherein we have used $\{f_n\}_{n=1}^\infty$ is Cauchy in measure and $g_j\overset{\mu}{\to}f$. ■

**Corollary 14.9 (Dominated Convergence Theorem).** *Let $(\Omega,\mathcal{B},\mu)$ be a measure space. Suppose $\{f_n\}$, $\{g_n\}$, and $g$ are in $L^1$ and $f\in L^0$ are functions such that*

$$|f_n|\leq g_n \text{ a.e.},\ f_n\overset{\mu}{\longrightarrow}f,\ g_n\overset{\mu}{\longrightarrow}g,\ \text{and}\int g_n\to\int g \text{ as } n\to\infty.$$

*Then $f\in L^1$ and $\lim_{n\to\infty}\|f-f_n\|_1=0$, i.e. $f_n\to f$ in $L^1$. In particular $\lim_{n\to\infty}\int f_n=\int f$.*

**Proof.** First notice that $|f|\leq g$ a.e. and hence $f\in L^1$ since $g\in L^1$. To see that $|f|\leq g$, use item 4. of Theorem 14.8 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f|=\lim_{k\to\infty}|f_{n_k}|\leq\lim_{k\to\infty}g_{n_k}=g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty}\|f-f_n\|_1\neq 0$ there exists $\varepsilon>0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int|f-f_{n_k}|\geq\varepsilon \text{ for all } k. \tag{14.5}$$

Using item 4. of Theorem 14.8 again, we may assume (by passing to a further subsequences if necessary) that $f_{n_k}\to f$ and $g_{n_k}\to g$ almost everywhere. Noting, $|f-f_{n_k}|\leq g+g_{n_k}\to 2g$ and $\int(g+g_{n_k})\to\int 2g$, an application of the dominated convergence Theorem 9.27 implies $\lim_{k\to\infty}\int|f-f_{n_k}|=0$ which contradicts Eq. (14.5). ■

**Exercise 14.1 (Fatou's Lemma).** Let $(\Omega,\mathcal{B},\mu)$ be a measure space. If $f_n\geq 0$ and $f_n\overset{\mu}{\to}f$, then $\int_\Omega fd\mu\leq\liminf_{n\to\infty}\int_\Omega f_nd\mu$.

**Lemma 14.10.** *Suppose $1\leq p<\infty$, $\{f_n\}_{n=1}^\infty\subset L^p(\mu)$, and $f_n\overset{\mu}{\to}f$, then $\|f\|_p\leq\liminf_{n\to\infty}\|f_n\|_p$. Moreover if $\{f_n\}_{n=1}^\infty\cup\{f\}\subset L^p(\mu)$, then $\|f-f_n\|_p\to 0$ as $n\to\infty$ iff $\lim_{n\to\infty}\|f_n\|_p=\|f\|_p<\infty$ and $f_n\overset{\mu}{\longrightarrow}f$.*

**Proof.** Choose a subsequence, $g_k=f_{n_k}$, such that $\liminf_{n\to\infty}\|f_n\|_p=\lim_{k\to\infty}\|g_k\|_p$. By passing to a further subsequence if necessary, we may further assume that $g_k\to f$ a.e. Therefore, by Fatou's lemma,

$$\|f\|_p^p=\int_\Omega|f|^p\,d\mu=\int_\Omega\lim_{k\to\infty}|g_k|^p\,d\mu\leq\liminf_{k\to\infty}\int_\Omega|g_k|^p\,d\mu=\liminf_{n\to\infty}\|f_n\|_p^p$$

which proves the first assertion.

If $\|f-f_n\|_p\to 0$ as $n\to\infty$, then by the triangle inequality, $\left|\|f\|_p-\|f_n\|_p\right|\leq\|f-f_n\|_p$ which shows $\int|f_n|^p\to\int|f|^p$ if $f_n\to f$ in $L^p$. Chebyschev's inequality implies $f_n\overset{\mu}{\longrightarrow}f$ if $f_n\to f$ in $L^p$.

Conversely if $\lim_{n\to\infty}\|f_n\|_p=\|f\|_p<\infty$ and $f_n\overset{\mu}{\longrightarrow}f$, let $F_n:=|f-f_n|^p$ and $G_n:=2^{p-1}[|f|^p+|f_n|^p]$. Then $F_n\overset{\mu}{\longrightarrow}0$[2], $F_n\leq G_n\in L^1$, and $\int G_n\to\int G$ where $G:=2^p|f|^p\in L^1$. Therefore, by Corollary 14.9, $\int|f-f_n|^p=\int F_n\to\int 0=0$. ■

**Exercise 14.2.** Let $(\Omega,\mathcal{B},\mu)$ be a measure space, $p\in[1,\infty)$, and suppose that $0\leq f\in L^1(\mu)$, $0\leq f_n\in L^1(\mu)$ for all $n$, $f_n\overset{\mu}{\longrightarrow}f$, and $\int f_nd\mu\to\int fd\mu$. Then $f_n\to f$ in $L^1(\mu)$. In particular if $f,f_n\in L^p(\mu)$ and $f_n\to f$ in $L^p(\mu)$, then $|f_n|^p\to|f|^p$ in $L^1(\mu)$.

**Proposition 14.11.** *Suppose $(\Omega,\mathcal{B},\mu)$ is a probability space and $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions on $\Omega$. Then $\{f_n\}_{n=1}^\infty$ converges to $f$ in probability iff every subsequence, $\{f_n'\}_{n=1}^\infty$ of $\{f_n\}_{n=1}^\infty$ has a further subsequence, $\{f_n''\}_{n=1}^\infty$, which is almost surely convergent to $f$.*

**Proof.** If $\{f_n\}_{n=1}^\infty$ is convergent and hence Cauchy in probability then any subsequence, $\{f_n'\}_{n=1}^\infty$ is also Cauchy in probability. Hence by item 4. of Theorem 14.8 there is a further subsequence, $\{f_n''\}_{n=1}^\infty$ of $\{f_n'\}_{n=1}^\infty$ which is convergent almost surely.

Conversely if $\{f_n\}_{n=1}^\infty$ does not converge to $f$ in probability, then there exists an $\varepsilon>0$ and a subsequence, $\{n_k\}$ such that $\inf_k\mu(|f-f_{n_k}|\geq\varepsilon)>0$. Any subsequence of $\{f_{n_k}\}$ would have the same property and hence can not be almost surely convergent because of Egorov's Theorem 14.6. ■

[2] This is becuase $|F_n|\geq\varepsilon$ iff $|f-f_n|\geq\varepsilon^{1/p}$.

**Corollary 14.12.** *Suppose $(\Omega, \mathcal{B}, \mu)$ is a probability space, $f_n \xrightarrow{\mu} f$ and $g_n \xrightarrow{\mu} g$ and $\varphi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions. Then*

1. $\varphi(f_n) \xrightarrow{\mu} \varphi(f)$,
2. $\psi(f_n, g_n) \xrightarrow{\mu} \psi(f, g)$, *and*
3. $f_n \cdot g_n \xrightarrow{\mu} f \cdot g$.

**Proof.** Item 1. and 3. follow from item 2. by taking $\psi(x, y) = \varphi(x)$ and $\psi(x, y) = x \cdot y$ respectively. So it suffices to prove item 2. To do this we will make repeated use of Theorem 14.8.

Given any subsequence, $\{n_k\}$, of $\mathbb{N}$ there is a subsequence, $\{n_k'\}$ of $\{n_k\}$ such that $f_{n_k'} \to f$ a.s. and yet a further subsequence $\{n_k''\}$ of $\{n_k'\}$ such that $g_{n_k''} \to g$ a.s. Hence, by the continuity of $\psi$, it now follows that

$$\lim_{k\to\infty} \psi\left(f_{n_k''}, g_{n_k''}\right) = \psi(f, g) \text{ a.s.}$$

which completes the proof. ∎

*Example 14.13.* It is not possible to drop the assumption that $\mu(\Omega) < \infty$ in Corollary 14.12. For example, let $\Omega = \mathbb{R}$, $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$, $\mu = m$ be Lebesgue measure, $f_n(x) = \frac{1}{n}$ and $g_n(x) = x^2 = g(x)$. Then $f_n \overset{\mu}{\to} 0$, $g_n \overset{\mu}{\to} g$ while $f_n g_n$ does not converge to $0 = 0 \cdot g$ in measure. Also if we let $\varphi(y) = y^2$, $f_n(x) = x + 1/n$ and $f(x) = x$ for all $x \in \mathbb{R}$, then $f_n \overset{\mu}{\to} f$ while

$$[\varphi(f_n) - \varphi(f)](x) = (x + 1/n)^2 - x^2 = \frac{2}{n}x + \frac{1}{n^2}$$

does not go to 0 in measure as $n \to \infty$.

## 14.3 Jensen's, Hölder's and Minkowski's Inequalities

**Theorem 14.14 (Jensen's Inequality).** *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a probability space, i.e. $\mu$ is a positive measure and $\mu(\Omega) = 1$. Also suppose that $f \in L^1(\mu)$, $f : \Omega \to (a, b)$, and $\varphi : (a, b) \to \mathbb{R}$ is a convex function, (i.e. $\varphi''(x) \geq 0$ on $(a, b)$.) Then*

$$\varphi\left(\int_\Omega f d\mu\right) \leq \int_\Omega \varphi(f) d\mu$$

*where if $\varphi \circ f \notin L^1(\mu)$, then $\varphi \circ f$ is integrable in the extended sense and $\int_\Omega \varphi(f) d\mu = \infty$.*

**Proof.** Let $t = \int_\Omega f d\mu \in (a, b)$ and let $\beta \in \mathbb{R}$ ($\beta = \dot{\varphi}(t)$ when $\dot{\varphi}(t)$ exists), be such that $\varphi(s) - \varphi(t) \geq \beta(s - t)$ for all $s \in (a, b)$. (See Lemma 14.60) and Figure 14.5 when $\varphi$ is $C^1$ and Theorem 14.63 below for the existence of such a $\beta$ in the general case.) Then integrating the inequality, $\varphi(f) - \varphi(t) \geq \beta(f - t)$, implies that

$$0 \leq \int_\Omega \varphi(f) d\mu - \varphi(t) = \int_\Omega \varphi(f) d\mu - \varphi(\int_\Omega f d\mu).$$

Moreover, if $\varphi(f)$ is not integrable, then $\varphi(f) \geq \varphi(t) + \beta(f - t)$ which shows that negative part of $\varphi(f)$ is integrable. Therefore, $\int_\Omega \varphi(f) d\mu = \infty$ in this case. ∎

*Example 14.15.* Since $e^x$ for $x \in \mathbb{R}$, $-\ln x$ for $x > 0$, and $x^p$ for $x \geq 0$ and $p \geq 1$ are all convex functions, we have the following inequalities

$$\exp\left(\int_\Omega f d\mu\right) \leq \int_\Omega e^f d\mu, \tag{14.6}$$

$$\int_\Omega \log(|f|) d\mu \leq \log\left(\int_\Omega |f| \, d\mu\right)$$

and for $p \geq 1$,

$$\left|\int_\Omega f d\mu\right|^p \leq \left(\int_\Omega |f| \, d\mu\right)^p \leq \int_\Omega |f|^p \, d\mu.$$

*Example 14.16.* As a special case of Eq. (14.6), if $p_i, s_i > 0$ for $i = 1, 2, \dots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \dots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}. \tag{14.7}$$

Indeed, we have applied Eq. (14.6) with $\Omega = \{1, 2, \dots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i^{p_i}$. As a special case of Eq. (14.7), suppose that $s, t, p, q \in (1, \infty)$ with $q = \frac{p}{p-1}$ (i.e. $\frac{1}{p} + \frac{1}{q} = 1$) then

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q. \tag{14.8}$$

(When $p = q = 1/2$, the inequality in Eq. (14.8) follows from the inequality, $0 \leq (s - t)^2$.)

As another special case of Eq. (14.7), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \dots a_n} \leq \frac{1}{n} \sum_{i=1}^n a_i. \tag{14.9}$$

*Example 14.17.* Let $(\Omega,\mathcal{B},\mu)$ be a probability space, $0<p<q<\infty$, and $f:\Omega\to\mathbb{C}$ be a measurable function. Then by Jensen's inequality,

$$\left(\int_\Omega |f|^p\,d\mu\right)^{q/p}\le\int_\Omega (|f|^p)^{q/p}\,d\mu=\int_\Omega |f|^q\,d\mu$$

from which it follows that $\|f\|_p\le\|f\|_q$. In particular, $L^p(\mu)\subset L^q(\mu)$ for all $0<p<q<\infty$. See Corollary 14.31 for an alternative proof.

**Theorem 14.18 (Hölder's inequality).** *Suppose that $1\le p\le\infty$ and $q:=\frac{p}{p-1}$, or equivalently $p^{-1}+q^{-1}=1$. If $f$ and $g$ are measurable functions then*

$$\|fg\|_1\le\|f\|_p\cdot\|g\|_q. \tag{14.10}$$

*Assuming $p\in(1,\infty)$ and $\|f\|_p\cdot\|g\|_q<\infty$, equality holds in Eq. (14.10) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of $L^1$ which happens iff*

$$|g|^q\|f\|_p^p=\|g\|_q^q\,|f|^p\ \ a.e. \tag{14.11}$$

**Proof.** The cases $p=1$ and $q=\infty$ or $p=\infty$ and $q=1$ are easy to deal with and will be left to the reader. So we now assume that $p,q\in(1,\infty)$. If $\|f\|_q=0$ or $\infty$ or $\|g\|_p=0$ or $\infty$, Eq. (14.10) is again easily verified. So we will now assume that $0<\|f\|_q,\|g\|_p<\infty$. Taking $s=|f|/\|f\|_p$ and $t=|g|/\|g\|_q$ in Eq. (14.8) gives,

$$\frac{|fg|}{\|f\|_p\|g\|_q}\le\frac{1}{p}\ \frac{|f|^p}{\|f\|_p}+\frac{1}{q}\ \frac{|g|^q}{\|g\|^q} \tag{14.12}$$

with equality iff $|g/\|g\|_q|=|f|^{p-1}/\|f\|_p^{(p-1)}=|f|^{p/q}/\|f\|_p^{p/q}$, i.e. $|g|^q\|f\|_p^p=\|g\|_q^q\,|f|^p$. Integrating Eq. (14.12) implies

$$\frac{\|fg\|_1}{\|f\|_p\|g\|_q}\le\frac{1}{p}+\frac{1}{q}=1$$

with equality iff Eq. (14.11) holds. The proof is finished since it is easily checked that equality holds in Eq. (14.10) when $|f|^p=c\,|g|^q$ of $|g|^q=c\,|f|^p$ for some constant $c$. ■

*Example 14.19.* Suppose that $a_k\in\mathbb{C}$ for $k=1,2,\dots,n$ and $p\in[1,\infty)$, then

$$\left|\sum_{k=1}^n a_k\right|^p\le n^{p-1}\sum_{k=1}^n|a_k|^p. \tag{14.13}$$

Indeed, by Hölder's inequality applied using the measure space, $\{1,2,\dots,n\}$ equipped with counting measure, we have

$$\left|\sum_{k=1}^n a_k\right|=\left|\sum_{k=1}^n a_k\cdot 1\right|\le\left(\sum_{k=1}^n|a_k|^p\right)^{1/p}\left(\sum_{k=1}^n 1^q\right)^{1/q}=n^{1/q}\left(\sum_{k=1}^n|a_k|^p\right)^{1/p}$$

where $q=\frac{p}{p-1}$. Taking the $p^{\text{th}}$ – power of this inequality then gives, Eq. (14.14).

**Theorem 14.20 (Generalized Hölder's inequality).** *Suppose that $f_i:\Omega\to\mathbb{C}$ are measurable functions for $i=1,\dots,n$ and $p_1,\dots,p_n$ and $r$ are positive numbers such that $\sum_{i=1}^n p_i^{-1}=r^{-1}$, then*

$$\left\|\prod_{i=1}^n f_i\right\|_r\le\prod_{i=1}^n\|f_i\|_{p_i}. \tag{14.14}$$

**Proof.** One may prove this theorem by induction based on Hölder's Theorem 14.18 above. Alternatively we may give a proof along the lines of the proof of Theorem 14.18 which is what we will do here.

Since Eq. (14.14) is easily seen to hold if $\|f_i\|_{p_i}=0$ for some $i$, we will assume that $\|f_i\|_{p_i}>0$ for all $i$. By assumption, $\sum_{i=1}^n\frac{r_i}{p_i}=1$, hence we may replace $s_i$ by $s_i^r$ and $p_i$ by $p_i/r$ for each $i$ in Eq. (14.7) to find

$$s_1^r\dots s_n^r\le\sum_{i=1}^n\frac{(s_i^r)^{p_i/r}}{p_i/r}=r\sum_{i=1}^n\frac{s_i^{p_i}}{p_i}.$$

Now replace $s_i$ by $|f_i|/\|f_i\|_{p_i}$ in the previous inequality and integrate the result to find

$$\frac{1}{\prod_{i=1}^n\|f_i\|_{p_i}}\left\|\prod_{i=1}^n f_i\right\|_r^r\le r\sum_{i=1}^n\frac{1}{p_i}\frac{1}{\|f_i\|_{p_i}^{p_i}}\int_\Omega|f_i|^{p_i}\,d\mu=\sum_{i=1}^n\frac{r}{p_i}=1.$$

■

**Definition 14.21.** *A* ***norm*** *on a vector space $Z$ is a function $\|\cdot\|:Z\to[0,\infty)$ such that*

1. *(Homogeneity) $\|\lambda f\|=|\lambda|\,\|f\|$ for all $\lambda\in\mathbb{F}$ and $f\in Z$.*
2. *(Triangle inequality) $\|f+g\|\le\|f\|+\|g\|$ for all $f,g\in Z$.*
3. *(Positive definite) $\|f\|=0$ implies $f=0$.*

*A pair $(Z,\|\cdot\|)$ where $Z$ is a vector space and $\|\cdot\|$ is a norm on $Z$ is called a* ***normed vector space.***

**Theorem 14.22 (Minkowski's Inequality).** *If $1\le p\le\infty$ and $f,g\in L^p(\mu)$ then*

$$\|f+g\|_p\le\|f\|_p+\|g\|_p. \tag{14.15}$$

*In particular, $\left(L^p(\mu),\|\cdot\|_p\right)$ is a normed vector space for all $1\le p\le\infty$.*

**Proof.** When $p=\infty$, $|f| \le \|f\|_\infty$ a.e. and $|g| \le \|g\|_\infty$ a.e. so that $|f+g| \le |f|+|g| \le \|f\|_\infty + \|g\|_\infty$ a.e. and therefore

$$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty .$$

When $p<\infty$,

$$|f+g|^p \le (2\max(|f|,|g|))^p = 2^p \max(|f|^p,|g|^p) \le 2^p (|f|^p + |g|^p),$$

which implies[3] $f+g \in L^p$ since

$$\|f+g\|_p^p \le 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

Furthermore, when $p=1$ we have

$$\|f+g\|_1 = \int_\Omega |f+g| d\mu \le \int_\Omega |f|\, d\mu + \int_\Omega |g| d\mu = \|f\|_1 + \|g\|_1.$$

We now consider $p \in (1,\infty)$. We may assume $\|f+g\|_p$, $\|f\|_p$ and $\|g\|_p$ are all positive since otherwise the theorem is easily verified. Integrating

$$|f+g|^p = |f+g||f+g|^{p-1} \le (|f|+|g|)|f+g|^{p-1}$$

and then applying Holder's inequality with $q=p/(p-1)$ gives

$$\begin{aligned}\int_\Omega |f+g|^p d\mu &\le \int_\Omega |f|\ |f+g|^{p-1} d\mu + \int_\Omega |g|\ |f+g|^{p-1} d\mu \\ &\le (\|f\|_p + \|g\|_p)\, \|\,|f+g|^{p-1}\,\|_q, \end{aligned} \tag{14.16}$$

where

$$\||f+g|^{p-1}\|_q^q = \int_\Omega (|f+g|^{p-1})^q d\mu = \int_\Omega |f+g|^p d\mu = \|f+g\|_p^p. \tag{14.17}$$

Combining Eqs. (14.16) and (14.17) implies

$$\|f+g\|_p^p \le \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} \tag{14.18}$$

Solving this inequality for $\|f+g\|_p$ gives Eq. (14.15). ∎

[3] In light of Example 14.19, the last $2^p$ in the above inequality may be replaced by $2^{p-1}$.

## 14.4 Completeness of $L^p$ – spaces

**Definition 14.23 (Banach space).** *A normed vector space $(Z, \|\cdot\|)$ is a **Banach space** if is is complete, i.e. all Cauchy sequences are convergent. To be more precise we are assuming that if $\{x_n\}_{n=1}^\infty \subset Z$ satisfies, $\lim_{m,n\to\infty} \|x_n - x_m\| = 0$, then there exists an $x \in Z$ such that $\lim_{n\to\infty} \|x - x_n\| = 0$.*

**Theorem 14.24.** *Let $\|\cdot\|_\infty$ be as defined in Eq. (14.2), then $(L^\infty(\Omega,\mathcal{B},\mu), \|\cdot\|_\infty)$ is a Banach space. A sequence $\{f_n\}_{n=1}^\infty \subset L^\infty$ converges to $f \in L^\infty$ iff there exists $E \in \mathcal{B}$ such that $\mu(E)=0$ and $f_n \to f$ uniformly on $E^c$. Moreover, bounded simple functions are dense in $L^\infty$.*

**Proof.** By Minkowski's Theorem 14.22, $\|\cdot\|_\infty$ satisfies the triangle inequality. The reader may easily check the remaining conditions that ensure $\|\cdot\|_\infty$ is a norm. Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such $f_n \to f \in L^\infty$, i.e. $\|f-f_n\|_\infty \to 0$ as $n\to\infty$. Then for all $k \in \mathbb{N}$, there exists $N_k < \infty$ such that

$$\mu\left(|f-f_n| > k^{-1}\right) = 0 \text{ for all } n \ge N_k.$$

Let

$$E = \cup_{k=1}^\infty \cup_{n \ge N_k} \left\{|f-f_n| > k^{-1}\right\}.$$

Then $\mu(E)=0$ and for $x \in E^c$, $|f(x)-f_n(x)| \le k^{-1}$ for all $n \ge N_k$. This shows that $f_n \to f$ uniformly on $E^c$. Conversely, if there exists $E \in \mathcal{B}$ such that $\mu(E)=0$ and $f_n \to f$ uniformly on $E^c$, then for any $\varepsilon > 0$,

$$\mu(|f-f_n| \ge \varepsilon) = \mu(\{|f-f_n| \ge \varepsilon\} \cap E^c) = 0$$

for all $n$ sufficiently large. That is to say $\limsup_{j\to\infty} \|f-f_n\|_\infty \le \varepsilon$ for all $\varepsilon > 0$. The density of simple functions follows from the approximation Theorem 8.39. So the last item to prove is the completeness of $L^\infty$.

Suppose $\varepsilon_{m,n} := \|f_m - f_n\|_\infty \to 0$ as $m,n \to \infty$. Let $E_{m,n} = \{|f_n - f_m| > \varepsilon_{m,n}\}$ and $E := \cup E_{m,n}$, then $\mu(E)=0$ and

$$\sup_{x\in E^c} |f_m(x) - f_n(x)| \le \varepsilon_{m,n} \to 0 \text{ as } m,n \to \infty.$$

Therefore, $f := \lim_{n\to\infty} f_n$ exists on $E^c$ and the limit is uniform on $E^c$. Letting $f = \lim_{n\to\infty} 1_{E^c} f_n$, it then follows that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. ∎

**Theorem 14.25 (Completeness of $L^p(\mu)$).** *For $1 \le p \le \infty$, $L^p(\mu)$ equipped with the $L^p$ – norm, $\|\cdot\|_p$ (see Eq. (14.1)), is a Banach space.*

**Proof.** By Minkowski's Theorem 14.22, $\|\cdot\|_p$ satisfies the triangle inequality. As above the reader may easily check the remaining conditions that ensure $\|\cdot\|_p$ is a norm. So we are left to prove the completeness of $L^p(\mu)$ for $1 \le p < \infty$, the case $p = \infty$ being done in Theorem 14.24.

Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 14.4), $\{f_n\}$ is $L^0$-Cauchy (i.e. Cauchy in measure) and by Theorem 14.8 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou's Lemma,

$$\begin{aligned}\|g_j - f\|_p^p &= \int \lim_{k\to\infty} \inf |g_j - g_k|^p d\mu \le \lim_{k\to\infty} \inf \int |g_j - g_k|^p d\mu \\ &= \lim_{k\to\infty} \inf \|g_j - g_k\|_p^p \to 0 \text{ as } j \to \infty.\end{aligned}$$

In particular, $\|f\|_p \le \|g_j - f\|_p + \|g_j\|_p < \infty$ so the $f \in L^p$ and $g_j \xrightarrow{L^p} f$. The proof is finished because,

$$\|f_n - f\|_p \le \|f_n - g_j\|_p + \|g_j - f\|_p \to 0 \text{ as } j, n \to \infty.$$

**Alternative method of constructing** $\{g_j\}$ . Choose $\{g_j = f_{n_j}\}_{j=1}^\infty$ where $\{n_j\}_{j=1}^\infty$ is an increasing subsequence so that $\sum_{j=1}^\infty \|g_{j+1} - g_j\|_p < \infty$ and let $U := \sum_{j=0}^\infty |g_{j+1} - g_j|$ where $g_0 \equiv 0$. Then

$$\begin{aligned}\|U\|_p &\overset{\text{MCT}}{=} \lim_{N\to\infty} \left\| \sum_{j=0}^N |g_{j+1} - g_j| \right\|_p \\ &\le \lim_{N\to\infty} \sum_{j=0}^N \|g_{j+1} - g_j\|_p = \sum_{j=0}^\infty \|g_{j+1} - g_j\|_p < \infty\end{aligned}$$

and therefore $U < \infty$ a.e. But on the set $\{U < \infty\}$ of full measure, the sum $\sum_{j=0}^\infty (g_{j+1} - g_j)$ is absolutely convergent and therefore

$$f := \lim_{N\to\infty} \sum_{j=0}^N (g_{j+1} - g_j) = \lim_{N\to\infty} g_{N+1} \text{ exists a.e.}$$

■

See Definition 16.2 for a very important example of where completeness is used. To end this section we are going to record a few results we will need later regarding subspace of $L^p(\mu)$ which are induced by sub – $\sigma$ – algebras, $\mathcal{B}_0 \subset \mathcal{B}$.

**Lemma 14.26.** *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{B}_0$ be a sub – $\sigma$ – algebra of $\mathcal{B}$. Then for $1 \le p < \infty$, the map $i : L^p(\Omega, \mathcal{B}_0, \mu) \to L^p(\Omega, \mathcal{B}, \mu)$ defined by $i([f]_0) = [f]$ is a well defined linear isometry. Here we are writing,*

$$\begin{aligned}[f]_0 &= \{g \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu) : g = f \text{ a.e.}\} \text{ and} \\ [f] &= \{g \in \mathrm{L}^p(\Omega, \mathcal{B}, \mu) : g = f \text{ a.e.}\}.\end{aligned}$$

*Moreover the image of $i$, $i(L^p(\Omega, \mathcal{B}_0, \mu))$, is a closed subspace of $L^p(\Omega, \mathcal{B}, \mu)$.*

**Proof.** This is proof is routine and most of it will be left to the reader. Let us just check that $i(L^p(\Omega, \mathcal{B}_0, \mu))$, is a closed subspace of $L^p(\Omega, \mathcal{B}, \mu)$. To this end, suppose that $i([f_n]_0) = [f_n]$ is a convergent sequence in $\mathrm{L}^p(\Omega, \mathcal{B}, \mu)$. Because, $i$, is an isometry it follows that $\{[f_n]_0\}_{n=1}^\infty$ is a Cauchy and hence convergent sequence in $L^p(\Omega, \mathcal{B}_0, \mu)$. Letting $f \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$ such that $\|f - f_n\|_{L^p(\mu)} \to 0$, we will have, since $i$ is isometric, that $[f_n] \to [f] = i([f]_0) \in i(L^p(\Omega, \mathcal{B}_0, \mu))$ as desired. ■

**Exercise 14.3.** Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{B}_0$ be a sub – $\sigma$ – algebra of $\mathcal{B}$. Further suppose that to every $B \in \mathcal{B}$ there exists $A \in \mathcal{B}_0$ such that $\mu(B \Delta A) = 0$. Show for all $1 \le p < \infty$ that $i(L^p(\Omega, \mathcal{B}_0, \mu)) = L^p(\Omega, \mathcal{B}, \mu)$, i.e. to each $f \in \mathrm{L}^p(\Omega, \mathcal{B}, \mu)$ there exists a $g \in \mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$ such that $f = g$ a.e. **Hints:** 1. verify the last assertion for simple functions in $\mathrm{L}^p(\Omega, \mathcal{B}_0, \mu)$. 2. then make use of Theorem 8.39 and Exercise 8.4.

**Exercise 14.4.** Suppose that $1 \le p < \infty$, $(\Omega, \mathcal{B}, \mu)$ is a $\sigma$ – finite measure space and $\mathcal{B}_0$ is a sub – $\sigma$ – algebra of $\mathcal{B}$. Show that $i(L^p(\Omega, \mathcal{B}_0, \mu)) = L^p(\Omega, \mathcal{B}, \mu)$ implies; to every $B \in \mathcal{B}$ there exists $A \in \mathcal{B}_0$ such that $\mu(B \Delta A) = 0$.

**Convention:** From now on we will drop the cumbersome notation and simply identify $[f]$ with $f$ and $L^p(\Omega, \mathcal{B}_0, \mu)$ with its image, $i(L^p(\Omega, \mathcal{B}_0, \mu))$, in $L^p(\Omega, \mathcal{B}, \mu)$.

## 14.5 Density Results

**Theorem 14.27 (Density Theorem).** *Let $p \in [1, \infty)$, $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that*

1. *$\mathbb{M} \subset L^p(\mu, \mathbb{R})$ and $\sigma(\mathbb{M}) = \mathcal{B}$.*
2. *There exists $\psi_k \in \mathbb{M}$ such that $\psi_k \to 1$ boundedly.*

*Then to every function $f \in L^p(\mu, \mathbb{R})$, there exist $\varphi_n \in \mathbb{M}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. $\mathbb{M}$ is dense in $L^p(\mu, \mathbb{R})$.*

**Proof.** Fix $k \in \mathbb{N}$ for the moment and let $\mathbb{H}$ denote those bounded $\mathcal{B}$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{M}$ such that $\lim_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded measurable $\mathbb{R}$ – valued functions on $\Omega$, $1 \in \mathbb{H}$, $\mathbb{M} \subset \mathbb{H}$ and $\mathbb{H}$

is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} = 0.$[4] (Take the dominating function to be $g = [2C|\psi_k|]^p$ where $C$ is a constant bounding all of the $\{|f_n|\}_{n=1}^\infty$.) We may now choose $\varphi_n \in \mathbb{M}$ such that $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \le \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} \le \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} + \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \qquad (14.19)$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 10.20, now shows $\mathbb{H}$ contains all bounded measurable functions on $\Omega$. Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f|\le k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C|f|]^p$ where $C$ is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\varphi_k \in \mathbb{M}$ such that

$$\left\|\psi_k 1_{\{|f|\le k\}} f - \varphi_k\right\|_{L^p(\mu)} \le \frac{1}{k}.$$

The same line of reasoning used in Eq. (14.19) now implies $\lim_{k\to\infty} \|f - \varphi_k\|_{L^p(\mu)} = 0$. ∎

*Example 14.28.* Let $\mu$ be a measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu([-M, M]) < \infty$ for all $M < \infty$. Then, $C_c(\mathbb{R}, \mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ with compact support) is dense in $L^p(\mu)$ for all $1 \le p < \infty$. To see this, apply Theorem 14.27 with $\mathbb{M} = C_c(\mathbb{R}, \mathbb{R})$ and $\psi_k$ being the function which is 1 on $[-k, k]$, 0 on $\mathbb{R} \setminus (-(k+1), k+1)$, interpolates linearly between 0 and 1 on $[-(k+1), -k]$ and on $[k, k+1]$.

**Theorem 14.29.** *Suppose $p \in [1, \infty)$, $\mathcal{A} \subset \mathcal{B} \subset 2^\Omega$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and $\mu$ is $\sigma$ – finite on $\mathcal{A}$. Let $\mathbb{S}(\mathcal{A}, \mu)$ denote the measurable simple functions, $\varphi : \Omega \to \mathbb{R}$ such $\{\varphi = y\} \in \mathcal{A}$ for all $y \in \mathbb{R}$ and $\mu(\{\varphi \ne 0\}) < \infty$. Then $\mathbb{S}(\mathcal{A}, \mu)$ is dense subspace of $L^p(\mu)$.*

**Proof.** Let $\mathbb{M} := \mathbb{S}(\mathcal{A}, \mu)$. By assumption there exists $\Omega_k \in \mathcal{A}$ such that $\mu(\Omega_k) < \infty$ and $\Omega_k \uparrow \Omega$ as $k \to \infty$. If $A \in \mathcal{A}$, then $\Omega_k \cap A \in \mathcal{A}$ and $\mu(\Omega_k \cap A) < \infty$ so that $1_{\Omega_k \cap A} \in \mathbb{M}$. Therefore $1_A = \lim_{k\to\infty} 1_{\Omega_k \cap A}$ is $\sigma(\mathbb{M})$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(\mathbb{M}) \subset \mathcal{B}$ and therefore $\mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathbb{M}) \subset \mathcal{B}$, i.e. $\sigma(\mathbb{M}) = \mathcal{B}$. The theorem now follows from Theorem 14.27 after observing $\psi_k := 1_{\Omega_k} \in \mathbb{M}$ and $\psi_k \to 1$ boundedly. ∎

[4] It is at this point that the proof would break down if $p = \infty$.

**Theorem 14.30 (Separability of $L^p$ – Spaces).** *Suppose, $p \in [1, \infty)$, $\mathcal{A} \subset \mathcal{B}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and $\mu$ is $\sigma$ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

*is a countable dense subset.*

**Proof.** It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}(\mathcal{A}, \mu)$ relative to the $L^p(\mu)$ – norm. Once this is done, the proof is then complete since $\mathbb{S}(\mathcal{A}, \mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 14.29. ∎

## 14.6 Relationships between different $L^p$ – spaces

The $L^p(\mu)$ – norm controls two types of behaviors of $f$, namely the "behavior at infinity" and the behavior of "local singularities." So in particular, if $f$ blows up at a point $x_0 \in \Omega$, then locally near $x_0$ it is harder for $f$ to be in $L^p(\mu)$ as $p$ increases. On the other hand a function $f \in L^p(\mu)$ is allowed to decay at "infinity" slower and slower as $p$ increases. With these insights in mind, we should not in general expect $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$. However, there are two notable exceptions. (1) If $\mu(\Omega) < \infty$, then there is no behavior at infinity to worry about and $L^q(\mu) \subset L^p(\mu)$ for all $q \ge p$ as is shown in Corollary 14.31 below. (2) If $\mu$ is counting measure, i.e. $\mu(A) = \#(A)$, then all functions in $L^p(\mu)$ for any $p$ can not blow up on a set of positive measure, so there are no local singularities. In this case $L^p(\mu) \subset L^q(\mu)$ for all $q \ge p$, see Corollary 14.36 below.

**Corollary 14.31 (Example 14.17 revisited).** *If $\mu(\Omega) < \infty$ and $0 < p \le q \le \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \le [\mu(\Omega)]^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q .$$

**Proof.** Take $a \in [1, \infty]$ such that

$$\frac{1}{p} = \frac{1}{a} + \frac{1}{q}, \text{ i.e. } a = \frac{pq}{q-p}.$$

Then by Theorem 14.20,

$$\|f\|_p = \|f \cdot 1\|_p \le \|f\|_q \cdot \|1\|_a = \mu(\Omega)^{1/a} \|f\|_q = \mu(\Omega)^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

The reader may easily check this final formula is correct even when $q = \infty$ provided we interpret $1/p - 1/\infty$ to be $1/p$. ∎

## The rest of this section may be skipped.

*Example 14.32 (Power Inequalities).* Let $a := (a_1, \dots, a_n)$ with $a_i > 0$ for $i = 1, 2, \dots, n$ and for $p \in \mathbb{R} \setminus \{0\}$, let

$$\|a\|_p := \left(\frac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p}.$$

Then by Corollary 14.31, $p \to \|a\|_p$ is increasing in $p$ for $p > 0$. For $p = -q < 0$, we have

$$\|a\|_p := \left(\frac{1}{n}\sum_{i=1}^n a_i^{-q}\right)^{-1/q} = \left(\frac{1}{\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{a_i}\right)^q}\right)^{1/q} = \left\|\frac{1}{a}\right\|_q^{-1}$$

where $\frac{1}{a} := (1/a_1, \dots, 1/a_n)$. So for $p < 0$, as $p$ increases, $q = -p$ decreases, so that $\left\|\frac{1}{a}\right\|_q$ is decreasing and hence $\left\|\frac{1}{a}\right\|_q^{-1}$ is increasing. Hence we have shown that $p \to \|a\|_p$ is increasing for $p \in \mathbb{R} \setminus \{0\}$.

We now claim that $\lim_{p\to 0}\|a\|_p = \sqrt[n]{a_1 \dots a_n}$. To prove this, write $a_i^p = e^{p \ln a_i} = 1 + p \ln a_i + O(p^2)$ for $p$ near zero. Therefore,

$$\frac{1}{n}\sum_{i=1}^n a_i^p = 1 + p\frac{1}{n}\sum_{i=1}^n \ln a_i + O(p^2).$$

Hence it follows that

$$\begin{aligned}\lim_{p\to 0}\|a\|_p &= \lim_{p\to 0}\left(\frac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p} = \lim_{p\to 0}\left(1 + p\frac{1}{n}\sum_{i=1}^n \ln a_i + O(p^2)\right)^{1/p} \\ &= e^{\frac{1}{n}\sum_{i=1}^n \ln a_i} = \sqrt[n]{a_1 \dots a_n}.\end{aligned}$$

So if we now define $\|a\|_0 := \sqrt[n]{a_1 \dots a_n}$, the map $p \in \mathbb{R} \to \|a\|_p \in (0, \infty)$ is continuous and increasing in $p$.

We will now show that $\lim_{p\to\infty}\|a\|_p = \max_i a_i =: M$ and $\lim_{p\to-\infty}\|a\|_p = \min_i a_i =: m$. Indeed, for $p > 0$,

$$\frac{1}{n}M^p \le \frac{1}{n}\sum_{i=1}^n a_i^p \le M^p$$

and therefore,

$$\left(\frac{1}{n}\right)^{1/p} M \le \|a\|_p \le M.$$

Since $\left(\frac{1}{n}\right)^{1/p} \to 1$ as $p \to \infty$, it follows that $\lim_{p\to\infty}\|a\|_p = M$. For $p = -q < 0$, we have

$$\lim_{p\to-\infty}\|a\|_p = \lim_{q\to\infty}\left(\frac{1}{\left\|\frac{1}{a}\right\|_q}\right) = \frac{1}{\max_i(1/a_i)} = \frac{1}{1/m} = m = \min_i a_i.$$

**Conclusion**. If we extend the definition of $\|a\|_p$ to $p = \infty$ and $p = -\infty$ by $\|a\|_\infty = \max_i a_i$ and $\|a\|_{-\infty} = \min_i a_i$, then $\bar{\mathbb{R}} \ni p \to \|a\|_p \in (0, \infty)$ is a continuous non-decreasing function of $p$.

**Proposition 14.33.** *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined by*

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1} \tag{14.20}$$

*with the interpretation that $\lambda/p_1 = 0$ if $p_1 = \infty$.*[5] *Then $L^{p_\lambda} \subset L^{p_0} + L^{p_1}$, i.e. every function $f \in L^{p_\lambda}$ may be written as $f = g + h$ with $g \in L^{p_0}$ and $h \in L^{p_1}$. For $1 \le p_0 < p_1 \le \infty$ and $f \in L^{p_0} + L^{p_1}$ let*

$$\|f\| := \inf\left\{\|g\|_{p_0} + \|h\|_{p_1} : f = g + h\right\}.$$

*Then $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map from $L^{p_\lambda}$ to $L^{p_0} + L^{p_1}$ is bounded; in fact $\|f\| \le 2\|f\|_{p_\lambda}$ for all $f \in L^{p_\lambda}$.*

**Proof.** Let $M > 0$, then the local singularities of $f$ are contained in the set $E := \{|f| > M\}$ and the behavior of $f$ at "infinity" is solely determined by $f$ on $E^c$. Hence let $g = f1_E$ and $h = f1_{E^c}$ so that $f = g + h$. By our earlier discussion we expect that $g \in L^{p_0}$ and $h \in L^{p_1}$ and this is the case since,

$$\begin{aligned}\|g\|_{p_0}^{p_0} &= \int |f|^{p_0} 1_{|f|>M} = M^{p_0}\int \left|\frac{f}{M}\right|^{p_0} 1_{|f|>M} \\ &\le M^{p_0}\int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \le M^{p_0 - p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty\end{aligned}$$

and

$$\begin{aligned}\|h\|_{p_1}^{p_1} &= \left\|f1_{|f|\le M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\le M} = M^{p_1}\int \left|\frac{f}{M}\right|^{p_1} 1_{|f|\le M} \\ &\le M^{p_1}\int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\le M} \le M^{p_1 - p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty.\end{aligned}$$

[5] A little algebra shows that $\lambda$ may be computed in terms of $p_0$, $p_\lambda$ and $p_1$ by

$$\lambda = \frac{p_0}{p_\lambda} \cdot \frac{p_1 - p_\lambda}{p_1 - p_0}.$$

Moreover this shows

$$\|f\| \le M^{1-p_\lambda/p_0} \|f\|_{p_\lambda}^{p_\lambda/p_0} + M^{1-p_\lambda/p_1} \|f\|_{p_\lambda}^{p_\lambda/p_1}.$$

Taking $M = \lambda \|f\|_{p_\lambda}$ then gives

$$\|f\| \le \left(\lambda^{1-p_\lambda/p_0} + \lambda^{1-p_\lambda/p_1}\right) \|f\|_{p_\lambda}$$

and then taking $\lambda = 1$ shows $\|f\| \le 2\|f\|_{p_\lambda}$. The proof that $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space is left as Exercise 14.11 to the reader. ■

**Corollary 14.34 (Interpolation of $L^p$ – norms).** *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined as in Eq. (14.20), then $L^{p_0} \cap L^{p_1} \subset L^{p_\lambda}$ and*

$$\|f\|_{p_\lambda} \le \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}. \tag{14.21}$$

*Further assume $1 \le p_0 < p_\lambda < p_1 \le \infty$, and for $f \in L^{p_0} \cap L^{p_1}$ let*

$$\|f\| := \|f\|_{p_0} + \|f\|_{p_1}.$$

*Then $(L^{p_0} \cap L^{p_1}, \|\cdot\|)$ is a Banach space and the inclusion map of $L^{p_0} \cap L^{p_1}$ into $L^{p_\lambda}$ is bounded, in fact*

$$\|f\|_{p_\lambda} \le \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{14.22}$$

The heuristic explanation of this corollary is that if $f \in L^{p_0} \cap L^{p_1}$, then $f$ has local singularities no worse than an $L^{p_1}$ function and behavior at infinity no worse than an $L^{p_0}$ function. Hence $f \in L^{p_\lambda}$ for any $p_\lambda$ between $p_0$ and $p_1$.

**Proof.** Let $\lambda$ be determined as above, $a = p_0/\lambda$ and $b = p_1/(1-\lambda)$, then by Theorem 14.20,

$$\|f\|_{p_\lambda} = \left\| |f|^\lambda |f|^{1-\lambda} \right\|_{p_\lambda} \le \left\| |f|^\lambda \right\|_a \left\| |f|^{1-\lambda} \right\|_b = \|f\|_{p_0}^{\lambda} \|f\|_{p_1}^{1-\lambda}.$$

It is easily checked that $\|\cdot\|$ is a norm on $L^{p_0} \cap L^{p_1}$. To show this space is complete, suppose that $\{f_n\} \subset L^{p_0} \cap L^{p_1}$ is a $\|\cdot\|$ – Cauchy sequence. Then $\{f_n\}$ is both $L^{p_0}$ and $L^{p_1}$ – Cauchy. Hence there exist $f \in L^{p_0}$ and $g \in L^{p_1}$ such that $\lim_{n\to\infty} \|f - f_n\|_{p_0} = 0$ and $\lim_{n\to\infty} \|g - f_n\|_{p_\lambda} = 0$. By Chebyshev's inequality (Lemma 14.4) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 14.8, $f = g$ a.e. It now is clear that $\lim_{n\to\infty} \|f - f_n\| = 0$. The estimate in Eq. (14.22) is left as Exercise 14.10 to the reader. ■

*Remark 14.35.* Combining Proposition 14.33 and Corollary 14.34 gives

$$L^{p_0} \cap L^{p_1} \subset L^{p_\lambda} \subset L^{p_0} + L^{p_1}$$

for $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ as in Eq. (14.20).

**Corollary 14.36.** *Suppose now that $\mu$ is counting measure on $\Omega$. Then $L^p(\mu) \subset L^q(\mu)$ for all $0 < p < q \le \infty$ and $\|f\|_q \le \|f\|_p$.*

**Proof.** Suppose that $0 < p < q = \infty$, then

$$\|f\|_\infty^p = \sup\{|f(x)|^p : x \in \Omega\} \le \sum_{x\in\Omega} |f(x)|^p = \|f\|_p^p,$$

i.e. $\|f\|_\infty \le \|f\|_p$ for all $0 < p < \infty$. For $0 < p \le q \le \infty$, apply Corollary 14.34 with $p_0 = p$ and $p_1 = \infty$ to find

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q} \|f\|_p^{1-p/q} = \|f\|_p.$$

■

### 14.6.1 Summary:

1. $L^{p_0} \cap L^{p_1} \subset L^q \subset L^{p_0} + L^{p_1}$ for any $q \in (p_0, p_1)$.
2. If $p \le q$, then $\ell^p \subset \ell^q$ and $\|f\|_q \le \|f\|_p$.
3. Since $\mu(|f| > \varepsilon) \le \varepsilon^{-p} \|f\|_p^p$, $L^p$ – convergence implies $L^0$ – convergence.
4. $L^0$ – convergence implies almost everywhere convergence for some subsequence.
5. If $\mu(\Omega) < \infty$ then almost everywhere convergence implies uniform convergence off certain sets of small measure and in particular we have $L^0$ – convergence.
6. If $\mu(\Omega) < \infty$, then $L^q \subset L^p$ for all $p \le q$ and $L^q$ – convergence implies $L^p$ – convergence.

## 14.7 Uniform Integrability

Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, $1 \le p < \infty$, and $\{f\} \cup \{f_n\}_{n=1}^\infty$ be a collection of random variables. The goal of this section is to find necessary and sufficient conditions on $\{f_n\}_{n=1}^\infty$ such that $f_n \to f$ in $L^p(\mu)$.

**Notation 14.37** *For $f \in L^1(\mu)$ and $E \in \mathcal{B}$, let*

$$\mu(f : E) := \int_E f d\mu.$$

*and more generally if $A, B \in \mathcal{B}$ let*

$$\mu(f : A, B) := \int_{A\cap B} f d\mu.$$

*When $\mu$ is a probability measure, we will often write $\mathbb{E}[f : E]$ for $\mu(f : E)$ and $\mathbb{E}[f : A, B]$ for $\mu(f : A, B)$.*

**Definition 14.38.** *A collection of functions, $\Lambda \subset L^1(\mu)$ is said to be **uniformly integrable** if,*

$$\lim_{a\to\infty} \sup_{f\in\Lambda} \mu(|f| : |f| \geq a) = 0. \tag{14.23}$$

*In words, $\Lambda \subset L^1(\mu)$ is uniformly integrable if "tail expectations" can be made uniformly small.*

*Example 14.39.* If $\Lambda \subset L^1(\mu)$ and there exists a dominating function, $g \in L^1(\mu)$, such that $|f| \leq g$ for all $f \in \Lambda$, then $\Lambda$ is uniformly integrable. Indeed,

$$\sup_{f\in\Lambda} \mu(|f| : |f| \geq a) \leq \mu(g : g \geq a) \to 0 \text{ as } a \uparrow \infty$$

by the dominated convergence theorem.

*Example 14.40.* Suppose that $\Lambda := \{X_n\}_{n=1}^\infty \subset L^1(\mu)$ is a sequence of random variables which all have the same law, then $\Lambda$ is uniformly integrable. This is because, $\mu(|X_n| : |X_n| \geq a) = \mu(|X_1| : |X_1| \geq a)$, and so

$$\sup_n \mu(|X_n| : |X_n| \geq a) = \mu(|X_1| : |X_1| \geq a) \to 0 \text{ as } a \uparrow \infty.$$

This example illustrates the fact that uniform integrability is really is a condition on the collection of measures, $\{\mathrm{Law}_\mu(X) : X \in \Lambda\}$, on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$.

*Example 14.41.* Suppose that $\Lambda := \{X_n\}_{n=1}^\infty \subset L^1(\mu)$ is a sequence of i.i.d. random variables such that $\mu(|X_1| \geq a) > 0$ for all $a \in (0,\infty)$, i.e. $X_n$ is not essentially bounded. The smallest dominating function for all of the $|X_n|$ is $Y := \sup_n |X_n|$. However, since

$$\sum_{n=1}^\infty \mu(|X_n| \geq a) = \sum_{n=1}^\infty \mu(|X_1| \geq a) = \infty,$$

the second Borel Cantelli lemma implies that $\mu(\{|X_n| \geq a \text{ i.o.}\}) = 1$ from which it follows that $Y \geq a$ a.s. Since $a > 0$ was arbitrary we conclude that $Y = \infty$ a.s. Thus we can not use Example 14.39 to show $\Lambda$ is uniformly integrable. Of course we do know by Example 14.40 that $\Lambda$ is uniformly integrable.

**Exercise 14.5.** Suppose $A$ is an index set, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two collections of random variables and $C \in (0,\infty)$. If $\{g_\alpha\}_{\alpha\in A}$ is uniformly integrable and $|f_\alpha| \leq C|g_\alpha|$ for all $\alpha \in A$, show $\{f_\alpha\}_{\alpha\in A}$ is uniformly integrable as well. [An an example which occurs in the dominated convergence theorem is when $g_\alpha = g \in L^1(\mu)$ for all $\alpha \in A$.]

**Lemma 14.42.** *If $\Lambda \subset L^1(\mu)$ is uniformly integrable, then $\sup_{f\in\Lambda} \|f\|_1 < \infty$.*[6]

**Proof.** Choose $a \in (0,\infty)$ sufficiently large so that $\sup_{f\in\Lambda} \mu(|f| : |f| \geq a) \leq 1$. Then for $f \in \Lambda$,

$$\|f\|_1 = \mu(|f| : |f| \geq a) + \mu(|f| : |f| < a) \leq 1 + a\mu(\Omega).$$

■

**Lemma 14.43.** *If Let $\{f_n\}_{n=1}^\infty$ is a collection of random variables such that $f_n \xrightarrow{\mu} 0$. Then for every $a \in (0,\infty)$, $g_n := f_n 1_{|f_n| \leq a} \to 0$ in $L^p(\mu)$ for all $1 \leq p < \infty$.*

**Proof.** As $|g_n| \leq |f_n|$ it follows that $g_n \xrightarrow{\mu} 0$ as $n \to \infty$. Since $|g_n| \leq a$, we now apply the dominated convergence theorem (Corollary 14.9) to conclude $\lim_{n\to\infty} \|g_n\|_p = 0$. ■

**Proposition 14.44.** *If $\{f_n\}_{n=1}^\infty$ is a sequence of random variables, then $\lim_{n\to\infty} \|f_n\|_1 = 0$ iff $f_n \xrightarrow{\mu} 0$ and $\{f_n\}_{n=1}^\infty$ is uniformly integrable.*

**Proof.** ($\Longrightarrow$) Suppose that $\lim_{n\to\infty} \|f_n\|_1 = 0$. By Chebyschev's inequality,

$$\mu(|f_n| \geq \varepsilon) \leq \frac{1}{\varepsilon} \|f_n\|_1 \to 0 \text{ as } n \to \infty,$$

i.e. $f_n \xrightarrow{\mu} 0$ as $n \to \infty$. If $N \in \mathbb{N}$, then using Example 14.39 it follows that

$$\sup_n \mu(|f_n| : |f_n| \geq a) \leq \left[\sup_{n<N} \mu(|f_n| : |f_n| \geq a)\right] \vee \sup_{n\geq N} \|f_n\|_1 \to \sup_{n\geq N} \|f_n\|_1 \text{ as } a \uparrow \infty$$

and so

$$\lim_{a\uparrow\infty} \left[\sup_n \mu(|f_n| : |f_n| \geq a)\right] \leq \sup_{n\geq N} \|f_n\|_1 \to 0 \text{ as } N \to \infty.$$

This shows $\{f_n\}_{n=1}^\infty$ is uniformly integrable.

($\Longleftarrow$) Now suppose that $f_n \xrightarrow{\mu} 0$ and $\{f_n\}_{n=1}^\infty$ is uniformly integrable. Then given any $a \in (0,\infty)$,

$$\begin{aligned}\|f_n\|_1 = \mu(|f_n|) &= \mu\left(|f_n| 1_{|f_n|<a}\right) + \mu(|f_n| : |f_n| \geq a) \\ &\leq \mu\left(|f_n| 1_{|f_n|<a}\right) + \sup_k \mu(|f_k| : |f_k| \geq a).\end{aligned}$$

[6] This is not necessarily the case if $\mu(\Omega) = \infty$. Indeed, if $\Omega = \mathbb{R}$ and $\mu = m$ is Lebesgue measure, the sequences of functions, $\left\{f_n := 1_{[-n,n]}\right\}_{n=1}^\infty$ are uniformly integrable but not bounded in $L^1(m)$.

From Lemma 14.43, $\lim_{n\to\infty}\mu\left(|f_n|\,1_{|f_n|<a}\right)=0$ and therefore,

$$\limsup_{n\to\infty}\|f_n\|_1\le\sup_k\mu\left(|f_k|:|f_k|\ge a\right)\to 0 \text{ as } a\uparrow\infty.$$

■

**Definition 14.45.** *A collection of functions, $\Lambda\subset L^1(\mu)$ is said to be **uniformly absolutely continuous** if for all $\varepsilon>0$ there exists $\delta>0$ such that*

$$\sup_{f\in\Lambda}\mu\left(|f|:E\right)<\varepsilon \text{ whenever } \mu(E)<\delta. \tag{14.24}$$

*Equivalently put,*

$$\lim_{\delta\downarrow 0}\sup\left\{\mu\left(|f|:E\right):f\in\Lambda \text{ and } \mu(E)<\delta\right\}=0. \tag{14.25}$$

*Example 14.46 (Optional).* I claim that $\Lambda=\{f\}$ with $f\in L^1(\mu)$ is uniformly absolutely continuous. If not there would exist and $\varepsilon>0$ and $E_n\in\mathcal{B}$ such that $\mu\left(|f|:E_n\right)\ge\varepsilon$ while $\lim_{n\to\infty}\mu(E_n)=0$. But this is not possible since $1_{E_n}f\overset{\mu}{\to}0$ as $n\to\infty$ and $|1_{E_n}f|\le|f|\in L^1(\mu)$ and so by dominated convergence theorem (Corollary 14.9),

$$\varepsilon\le\lim_{n\to\infty}\mu\left(|f|:E_n\right)=\lim_{n\to\infty}\mu\left(|1_{E_n}f|\right)=0.$$

*Remark 14.47 (Optional).* It is not in general true that if $\{f_n\}\subset L^1(\mu)$ is uniformly absolutely continuous implies $\sup_n\|f_n\|_1<\infty$. For example take $\Omega=\{*\}$ and $\mu(\{*\})=1$. Let $f_n(*)=n$. Since for $\delta<1$ a set $E\subset\Omega$ such that $\mu(E)<\delta$ is in fact the empty set and hence $\{f_n\}_{n=1}^\infty$ is uniformly absolutely continuous. However, for finite measure spaces without "atoms", for every $\delta>0$ we may find a finite partition of $\Omega$ by sets $\{E_\ell\}_{\ell=1}^k$ with $\mu(E_\ell)<\delta$, see Lemma 14.55 below. If Eq. (14.24) holds with $\varepsilon=1$, then

$$\mu(|f_n|)=\sum_{\ell=1}^k\mu(|f_n|:E_\ell)\le k$$

showing that $\mu(|f_n|)\le k$ for all $n$.

**Proposition 14.48.** *A subset $\Lambda\subset L^1(\mu)$ is uniformly integrable iff $\Lambda\subset L^1(\mu)$ is bounded and uniformly absolutely continuous.*

**Proof.** ($\Longrightarrow$) We have already seen that uniformly integrable subsets, $\Lambda$, are bounded in $L^1(\mu)$. Moreover, for $f\in\Lambda$, and $E\in\mathcal{B}$,

$$\begin{aligned}\mu(|f|:E)&=\mu(|f|:|f|\ge M,E)+\mu(|f|:|f|<M,E)\\&\le\mu(|f|:|f|\ge M)+M\mu(E).\end{aligned}$$

Therefore,

$$\lim_{\delta\downarrow 0}\sup\left\{\mu\left(|f|:E\right):f\in\Lambda \text{ and } \mu(E)<\delta\right\}\le\sup_{f\in\Lambda}\mu(|f|:|f|\ge M)\to 0 \text{ as } M\to\infty$$

which verifies that $\Lambda$ is uniformly absolutely continuous.

($\Longleftarrow$) Let $K:=\sup_{f\in\Lambda}\|f\|_1<\infty$. Then for $f\in\Lambda$, we have

$$\mu\left(|f|\ge a\right)\le\|f\|_1/a\le K/a \text{ for all } a>0.$$

Hence given $\varepsilon>0$ and $\delta>0$ as in the definition of uniform absolute continuity, we may choose $a=K/\delta$ in which case

$$\sup_{f\in\Lambda}\mu\left(|f|:|f|\ge a\right)<\varepsilon.$$

Since $\varepsilon>0$ was arbitrary, it follows that $\lim_{a\to\infty}\sup_{f\in\Lambda}\mu\left(|f|:|f|\ge a\right)=0$ as desired. ■

**Corollary 14.49.** *Suppose $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two uniformly integrable collections of functions, then $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ is also uniformly integrable.*

**Proof.** By Proposition 14.48, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are both bounded in $L^1(\mu)$ and are both uniformly absolutely continuous. Since $\|f_\alpha+g_\alpha\|_1\le\|f_\alpha\|_1+\|g_\alpha\|_1$ it follows that $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ is bounded in $L^1(\mu)$ as well. Moreover, for $\varepsilon>0$ we may choose $\delta>0$ such that $\mu\left(|f_\alpha|:E\right)<\varepsilon$ and $\mu\left(|g_\alpha|:E\right)<\varepsilon$ whenever $\mu(E)<\delta$. For this choice of $\varepsilon$ and $\delta$, we then have

$$\mu\left(|f_\alpha+g_\alpha|:E\right)\le\mu\left(|f_\alpha|+|g_\alpha|:E\right)<2\varepsilon \text{ whenever } \mu(E)<\delta,$$

showing $\{f_\alpha+g_\alpha\}_{\alpha\in A}$ uniformly absolutely continuous. Another application of Proposition 14.48 completes the proof. ■

**Corollary 14.50.** *If $1\le p<\infty$ and $\{f_n\}_{n=1}^\infty\cup\{f\}\subset L^p(\mu)$, then $\{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable iff $\{|f_n-f|^p\}_{n=1}^\infty$ is uniformly integrable.*

**Proof.** ($\Longrightarrow$) By Hölder's inequality,

$$|f_n-f|^p\le 2^{\frac{p}{q}}\left[|f_n|^p+|f|^p\right]$$

and so by Corollary 14.49 and Exercise 14.5 it follows that $\{|f_n-f|^p\}_{n=1}^\infty$ is uniformly integrable.

($\Longleftarrow$) The proof here is similar but now based on

$$|f_n|^p\le\left[|f_n-f|+|f|\right]^p\le 2^{\frac{p}{q}}\left[|f_n-f|^p+|f|^p\right].$$

■

**Exercise 14.6 (Problem 5 on p. 196 of Resnick.).** Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of integrable and i.i.d random variables. Then $\left\{\frac{S_n}{n}\right\}_{n=1}^\infty$ is uniformly integrable. [Actually there is no reason to assume independence here – identically distributed is enough.]

**Theorem 14.51 (Vitali Convergence Theorem).** *Let $1 \le p < \infty$, $(\Omega, \mathcal{B}, \mu)$ be a finite measure space, $\Lambda := \{f_n\}_{n=1}^\infty$ be a sequence of functions in $L^p(\mu)$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then $f \in L^p(\mu)$ and $\|f - f_n\|_p \to 0$ as $n \to \infty$ iff $f_n \xrightarrow{\mu} f$ and $\{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable.*

**Proof.** ($\Longrightarrow$) If $f_n \to f$ in $L^p(\mu)$ then by Chebyschev's inequality $f_n \xrightarrow{\mu} f$, Lemma 14.4. By assumption $|f - f_n|^p \to 0$ in $L^1(\mu)$ and so, by Proposition 14.48, $\{|f - f_n|^p\}_{n=1}^\infty$ is uniformly integrable. It then follows by Corollary 14.50 that $\{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable.

($\Longleftarrow$) Assume $f_n \xrightarrow{\mu} f$ as $n \to \infty$ and $\{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable. Then[7] by Fatou's Lemma (Exercise 14.1) and Lemma 14.42,

$$\int_\Omega |f|^p \, d\mu \le \liminf_{n\to\infty} \int_\Omega |f_n|^p \, d\mu \le \sup_n \int_\Omega |f_n|^p \, d\mu < \infty,$$

i.e. $f \in L^p(\mu)$. From Corollary 14.50 we now know that $\{|f - f_n|^p\}_{n=1}^\infty$ is uniformly integrable. As $f_n \xrightarrow{\mu} f$ is equivalent to $|f_n - f|^p \xrightarrow{\mu} 0$ we may apply Proposition 14.48 to show $|f - f_n|^p \to 0$ in $L^1(\mu)$, i.e. $f_n \to f$ in $L^p(\mu)$. ■

*Example 14.52.* Let $\Omega = [0,1]$, $\mathcal{B} = \mathcal{B}_{[0,1]}$ and $P = m$ be Lebesgue measure on $\mathcal{B}$. Then the collection of functions, $f_\varepsilon := \frac{1}{\varepsilon} 1_{[0,\varepsilon]}$ for $\varepsilon \in (0,1)$ is bounded in $L^1(P)$, $f_\varepsilon \to 0$ a.e. as $\varepsilon \downarrow 0$ but

$$0 = \int_\Omega \lim_{\varepsilon\downarrow 0} f_\varepsilon dP \ne \lim_{\varepsilon\downarrow 0} \int_\Omega f_\varepsilon dP = 1.$$

This is a typical example of a bounded and pointwise convergent sequence in $L^1$ which is **not** uniformly integrable. This is easy to check directly as well since,

$$\sup_{\varepsilon\in(0,1)} m(|f_\varepsilon| : |f_\varepsilon| \ge a) = 1 \text{ for all } a > 0.$$

*Example 14.53.* Let $\Omega = [0,1]$, $P$ be Lebesgue measure on $\mathcal{B} = \mathcal{B}_{[0,1]}$, and for $\varepsilon \in (0,1)$ let $a_\varepsilon > 0$ with $\lim_{\varepsilon\downarrow 0} a_\varepsilon = \infty$ and let $f_\varepsilon := a_\varepsilon 1_{[0,\varepsilon]}$. Then $\mathbb{E} f_\varepsilon = \varepsilon a_\varepsilon$ and so $\sup_{\varepsilon>0} \|f_\varepsilon\|_1 =: K < \infty$ iff $\varepsilon a_\varepsilon \le K$ for all $\varepsilon$. Since

$$\sup_\varepsilon \mathbb{E}[f_\varepsilon : f_\varepsilon \ge M] = \sup_\varepsilon [\varepsilon a_\varepsilon \cdot 1_{a_\varepsilon \ge M}],$$

if $\{f_\varepsilon\}$ is uniformly integrable and $\delta > 0$ is given, for large $M$ we have $\varepsilon a_\varepsilon \le \delta$ for $\varepsilon$ small enough so that $a_\varepsilon \ge M$. From this we conclude that $\limsup_{\varepsilon\downarrow 0} (\varepsilon a_\varepsilon) \le \delta$ and since $\delta > 0$ was arbitrary, $\lim_{\varepsilon\downarrow 0} \varepsilon a_\varepsilon = 0$ if $\{f_\varepsilon\}$ is uniformly integrable. By reversing these steps one sees the converse is also true.

**Alternatively.** No matter how $a_\varepsilon > 0$ is chosen, $\lim_{\varepsilon\downarrow 0} f_\varepsilon = 0$ a.s.. So from Theorem 14.51, if $\{f_\varepsilon\}$ is uniformly integrable we would have to have

$$\lim_{\varepsilon\downarrow 0} (\varepsilon a_\varepsilon) = \lim_{\varepsilon\downarrow 0} \mathbb{E} f_\varepsilon = \mathbb{E} 0 = 0.$$

The following Lemma gives a concrete necessary and sufficient conditions for verifying a sequence of functions is uniformly integrable.

**Lemma 14.54.** *Suppose that $\mu(\Omega) < \infty$, and $\Lambda \subset L^0(\Omega)$ is a collection of functions.*

1. *If there exists a measurable function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} \varphi(x)/x = \infty$ and*
$$K := \sup_{f\in\Lambda} \mu(\varphi(|f|)) < \infty, \tag{14.26}$$
*then $\Lambda$ is uniformly integrable. (A typical example for $\varphi$ in item 1. is $\varphi(x) = x^p$ for some $p > 1$.)*
2. **(Skip this if you like.) Conversely if $\Lambda$ is uniformly integrable, there exists a non-decreasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(0) = 0$, $\lim_{x\to\infty} \varphi(x)/x = \infty$ and Eq. (14.26) is valid.*

**Proof. 1.** Let $\varphi$ be as in item 1. above and set $\varepsilon_a := \sup_{x\ge a} \frac{x}{\varphi(x)} \to 0$ as $a \to \infty$ by assumption. Then for $f \in \Lambda$

$$\begin{aligned}\mu(|f| : |f| \ge a) &= \mu\left(\frac{|f|}{\varphi(|f|)} \varphi(|f|) : |f| \ge a\right) \le \mu(\varphi(|f|) : |f| \ge a)\varepsilon_a \\ &\le \mu(\varphi(|f|))\varepsilon_a \le K\varepsilon_a\end{aligned}$$

and hence

$$\lim_{a\to\infty} \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\ge a}\right) \le \lim_{a\to\infty} K\varepsilon_a = 0.$$

**2.** *(Skip this if you like.) By assumption, $\varepsilon_a := \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\ge a}\right) \to 0$ as $a \to \infty$. Therefore we may choose $a_n \uparrow \infty$ such that

$$\sum_{n=0}^\infty (n+1)\, \varepsilon_{a_n} < \infty$$

where by convention $a_0 := 0$. Now define $\varphi$ so that $\varphi(0) = 0$ and

[7] We are actually reproving Lemma 14.10 here.

$$\varphi'(x) = \sum_{n=0}^{\infty} (n+1)\, 1_{(a_n, a_{n+1}]}(x),$$

i.e.

$$\varphi(x) = \int_0^x \varphi'(y)dy = \sum_{n=0}^{\infty} (n+1)\,(x \wedge a_{n+1} - x \wedge a_n)\,.$$

By construction $\varphi$ is continuous, $\varphi(0) = 0$, $\varphi'(x)$ is increasing (so $\varphi$ is convex) and $\varphi'(x) \geq (n+1)$ for $x \geq a_n$. In particular

$$\frac{\varphi(x)}{x} \geq \frac{\varphi(a_n) + (n+1)x}{x} \geq n+1 \text{ for } x \geq a_n$$

from which we conclude $\lim_{x\to\infty} \varphi(x)/x = \infty$. We also have $\varphi'(x) \leq (n+1)$ on $[0, a_{n+1}]$ and therefore

$$\varphi(x) \leq (n+1)x \text{ for } x \leq a_{n+1}.$$

So for $f \in \Lambda$,

$$\begin{aligned}\mu\left(\varphi(|f|)\right) &= \sum_{n=0}^{\infty} \mu\left(\varphi(|f|)1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\,1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\,1_{|f|\geq a_n}\right) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n}\end{aligned}$$

and hence

$$\sup_{f\in\Lambda} \mu\left(\varphi(|f|)\right) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty.$$

■

**Exercise 14.7.** Show directly that if $\mu(\Omega) < \infty$, $\varphi$ is as in Lemma 14.54, and $\{f_n\} \subset L^1(\Omega)$ such that $f_n \overset{\mu}{\to} f$ and $K := \sup_n \mathbb{E}\left[\varphi(|f_n|)\right] < \infty$, then $\|f - f_n\|_1 \to 0$ as $n \to \infty$.

### 14.7.1 Atoms

**Lemma 14.55 (Saks' Lemma [10, Lemma 7 on p. 308]).** *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space such that $P$ has no atoms. (Recall that $A \in \mathcal{B}$ is an atom if $P(A) > 0$ and for any $B \subset A$ with $B \in \mathcal{B}$ we have either $P(B) = 0$ or $P(B) = P(A)$.) Then for every $\delta > 0$ there exists a partition $\{E_\ell\}_{\ell=1}^n$ of $\Omega$ with $\mu(E_\ell) < \delta$. (For related results along this line also see [9, 13, 20, 24] to name a few.)*

**Proof.** For any $A \in \mathcal{B}$ let

$$\beta(A) := \sup\left\{P(B) : B \subset A \text{ and } P(B) \leq \delta\right\}.$$

We begin by showing if $\mu(A) > 0$ then $\beta(A) > 0$. As there are no atoms there exists $A_1 \subset A$ such that $0 < P(A_1) < P(A)$. Similarly there exists $A_2 \subset A \setminus A_1$ such that $0 < P(A_2) < P(A \setminus A_1)$ and continuing inductively we find $\{A_n\}_{n=1}^{\infty}$ disjoint subsets of $A$ such that $A_n \subset A \setminus (A_1 \cup \dots \cup A_{n-1})$ and

$$0 < P(A_n) < P(A \setminus (A_1 \cup \dots \cup A_{n-1}))\,.$$

As $\sum_{n=1}^{\infty} A_n \subset A$ we must have $\sum_{n=1}^{\infty} P(A_n) \leq P(A) < \infty$ and therefore $\lim_{n\to\infty} P(A_n) = 0$. Thus for sufficiently large $n$ we have $0 < P(A_n) \leq \delta$ and therefore $\beta(A) \geq P(A_n) > 0$.

Now to construct the desired partition. Choose $A_1 \subset \Omega$ such that $\delta \geq P(A_1) \geq \frac{1}{2}\beta(\Omega)$. If $P(\Omega \setminus A_1) > 0$ we may then choose $A_2 \subset \Omega \setminus A_1$ such that $\delta \geq P(A_2) \geq \frac{1}{2}\beta(\Omega \setminus [A_1 \cup A_2])$. We may continue on this way inductively to find disjoint subsets $\{A_k\}_{k=1}^n$ of $\Omega$

$$\delta \geq P(A_k) \geq \frac{1}{2}\beta\left(\Omega \setminus [A_1 \cup \dots \cup A_{k-1}]\right)$$

with either $P(\Omega \setminus [A_1 \cup \dots \cup A_{n-1}]) > 0$. If it happens that $P(\Omega \setminus [A_1 \cup \dots \cup A_n]) = 0$ it is easy to see we are done. So we may assume that process can be carried on indefinitely. We then let $F := \Omega \setminus \cup_{k=1}^{\infty} A_k$ and observe that

$$\beta(F) \leq \beta\left(\Omega \setminus [A_1 \cup \dots \cup A_{n-1}]\right) \leq 2P(A_n) \to 0 \text{ as } n \to \infty$$

as

$$\sum_{n=1}^{\infty} P(A_n) \leq P(\Omega) < \infty.$$

But by the first paragraph this implies that $P(F) = 0$. Hence there exists $n < \infty$ such that $P\left(\Omega \setminus \cup_{k=1}^{n-1} A_k\right) \leq \delta$. We may then define $E_k = A_k$ for $1 \leq k \leq n-1$ and $E_n = \Omega \setminus \cup_{k=1}^{n-1} A_k$ in order to construct the desired partition. ■

**Corollary 14.56.** *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space such that $P$ has no atoms. Then for any $\alpha \in (0,1)$ there exists $A \in \mathcal{B}$ with $P(A) = \alpha$.*

**Proof.** We may assume the $\alpha \in (0, 1/2)$. By dividing $\Omega$ into a partition $\{E_l\}_{l=1}^N$ with $P(E_l) \leq \alpha/2$ we may let $A_1 := \cup_{l=1}^k E_l$ with $k$ chosen so that $P(A_1) \leq \alpha$ but

$$\alpha < P(A_1 \cup E_{k+1}) \leq \frac{3}{2}\alpha.$$

Notice that $\alpha/2 \leq P(A_1) \leq \alpha$. Apply this procedure to $\Omega \setminus A_1$ in order to find $A_2 \supset A_1$ such that $\alpha/4 \leq P(A_2) \leq \alpha$. Continue this way inductively to find $A_n \uparrow A$ such that $P(A_n) \uparrow \alpha = P(A)$. (BRUCE: clean this proof up.) ■

## 14.8 Exercises

**Exercise 14.8.** Let $f \in L^p \cap L^\infty$ for some $p < \infty$. Show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$. If we further assume $\mu(X) < \infty$, show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$ for all measurable functions $f : X \to \mathbb{C}$. In particular, $f \in L^\infty$ iff $\lim_{q\to\infty} \|f\|_q < \infty$. **Hints:** Use Corollary 14.34 to show $\limsup_{q\to\infty} \|f\|_q \le \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \ge \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

**Exercise 14.9.** Let $\infty > a, b > 1$ with $a^{-1} + b^{-1} = 1$. Give a calculus proof of the inequality

$$st \le \frac{s^a}{a} + \frac{t^b}{b} \text{ for all } s, t \ge 0.$$

**Hint**: by taking $s = xt^{b/a}$, show that it suffices to prove

$$x \le \frac{x^a}{a} + \frac{1}{b} \text{ for all } x \ge 0.$$

and then maximize the function $f(x) = x - x^a/a$ for $x \in [0, \infty)$.

**Exercise 14.10.** Prove Eq. (14.22) in Corollary 14.34. (Part of Folland 6.3 on p. 186.) **Hint:** Use the inequality, with $a, b \ge 1$ with $a^{-1} + b^{-1} = 1$ chosen appropriately,

$$st \le \frac{s^a}{a} + \frac{t^b}{b}$$

applied to the right side of Eq. (14.21).

**Exercise 14.11.** Complete the proof of Proposition 14.33 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space.

**Exercise 14.12.** Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. Show directly that for any $g \in L^1(\mu)$, $\Lambda = \{g\}$ is uniformly absolutely continuous. (We already know this is true by combining Example 14.39 with Proposition 14.48.)

**Exercise 14.13.** Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space and $\{X_n\}_{n=1}^\infty$ is a sequence of uncorrelated (i.e. $\mathrm{Cov}(X_n, X_m) = 0$ if $m \ne n$) square integrable random variables such that $\mu = \mathbb{E}X_n$ and $\sigma^2 = \mathrm{Var}(X_n)$ for all $n$. Let $S_n := X_1 + \cdots + X_n$. Show $\left\|\frac{S_n}{n} - \mu\right\|_2^2 = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$.

**Exercise 14.14.** Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. integrable random variables and $S_n := X_1 + \cdots + X_n$ and $\mu := \mathbb{E}X_n$. Show, $\frac{S_n}{n} \to \mu$ in $L^1(P)$ as $n \to \infty$. (Incidentally, this shows that $\left\{\frac{S_n}{n}\right\}_{n=1}^\infty$ is U.I. **Hint:** for $M \in (0, \infty)$, let $X_i^M := X_i \cdot 1_{|X_i| \le M}$ and $S_n^M := X_1^M + \cdots + X_n^M$ and use Exercise 14.13 to see that

$$\frac{S_n^M}{n} \to \mathbb{E}X_1^M \text{ in } L^2(P) \subset L^1(P) \text{ for all } M.$$

Using this to show $\lim_{n\to\infty} \left\|\frac{S_n}{n} - \mathbb{E}X_1\right\|_1 = 0$ by getting good control on $\left\|\frac{S_n}{n} - \frac{S_n^M}{n}\right\|_1$ and $\left|\mathbb{E}X_n - \mathbb{E}X_n^M\right|$.

**Exercise 14.15.** Suppose $1 \le p < \infty$, $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables such that $\mathbb{E}|X_n|^p < \infty$, $S_n := X_1 + \cdots + X_n$ and $\mu := \mathbb{E}X_n$. Show, $\frac{S_n}{n} \to \mu$ in $L^p(P)$ as $n \to \infty$. **Hint:** show $\left\{\left|\frac{S_n}{n}\right|^p\right\}_{n=1}^\infty$ is U.I. – this is not meant to be hard!

## 14.9 Appendix: Convex Functions

Reference; see the appendix (page 500) of Revuz and Yor.

**Definition 14.57.** *Given any function, $\varphi : (a, b) \to \mathbb{R}$, we say that $\varphi$ is* ***convex*** *if for all $a < x_0 \le x_1 < b$ and $t \in [0, 1]$,*

$$\varphi(x_t) \le h_t := (1-t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in [0,1], \tag{14.27}$$

*where*

$$x_t := x_0 + t(x_1 - x_0) = (1-t)x_0 + tx_1, \tag{14.28}$$

*see Figure 14.3 below.*

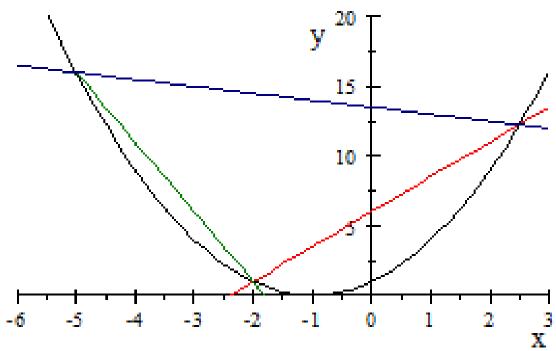



**Fig. 14.3.** A convex function along with three cords corresponding to $x_0 = -5$ and $x_1 = -2$ $x_0 = -2$ and $x_1 = 5/2$, and $x_0 = -5$ and $x_1 = 5/2$ with slopes, $m_1 = -15/3$, $m_2 = 15/6$ and $m_3 = -1/2$ respectively. Notice that $m_1 \le m_3 \le m_2$.

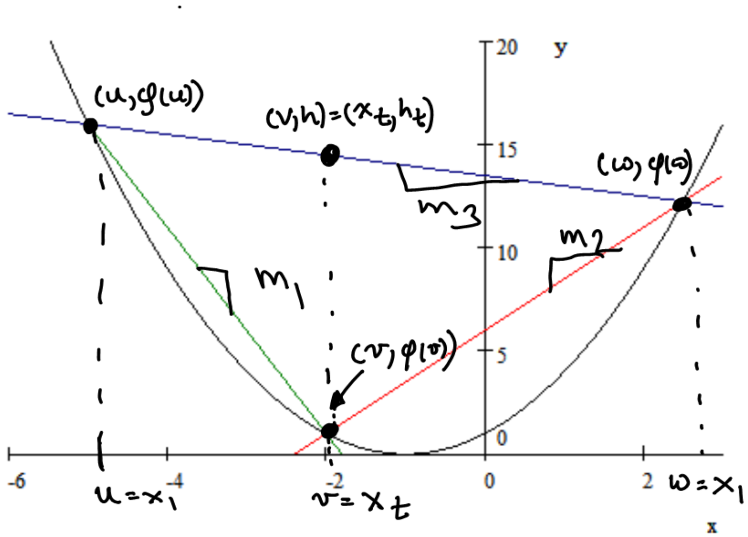



**Fig. 14.4.** A convex function with three cords. Notice the slope relationships; $m_1 \leq m_3 \leq m_2$.

**Lemma 14.58.** *Let $\varphi : (a,b) \to \mathbb{R}$ be a function and*

$$F(x_0, x_1) := \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \text{ for } a < x_0 < x_1 < b.$$

*Then the following are equivalent;*

1. *$\varphi$ is convex,*
2. *$F(x_0, x_1)$ is non-decreasing in $x_0$ for all $a < x_0 < x_1 < b$, and*
3. *$F(x_0, x_1)$ is non-decreasing in $x_1$ for all $a < x_0 < x_1 < b$.*

**Proof.** Let $x_t$ and $h_t$ be as in Eq. (14.27), then $(x_t, h_t)$ is on the line segment joining $(x_0, \varphi(x_0))$ to $(x_1, \varphi(x_1))$ and the statement that $\varphi$ is convex is then equivalent to the assertion that $\varphi(x_t) \leq h_t$ for all $0 \leq t \leq 1$. Since $(x_t, h_t)$ lies on a straight line we always have the following three slopes are equal;

$$\frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t}.$$

In light of this identity, it is now clear that the convexity of $\varphi$ is equivalent to either,

$$F(x_0, x_t) = \frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} \leq \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = F(x_0, x_1)$$

or

$$F(x_0, x_1) = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \leq \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t} = F(x_t, x_1)$$

holding for all $x_0 < x_t < x_1$. ■

**Lemma 14.59 (A generalized FTC).** *If $\varphi \in PC^1((a,b) \to \mathbb{R})$[8], then for all $a < x < y < b$,*

$$\varphi(y) - \varphi(x) = \int_x^y \varphi'(t)\, dt.$$

**Proof.** Let $b_1, \dots, b_{l-1}$ be the points of non-differentiability of $\varphi$ in $(x,y)$ and set $b_0 = x$ and $b_l = y$. Then

$$\begin{aligned} \varphi(y) - \varphi(x) &= \sum_{k=1}^{l} [\varphi(b_k) - \varphi(b_{k-1})] \\ &= \sum_{k=1}^{l} \int_{b_{k-1}}^{b_k} \varphi'(t)\, dt = \int_x^y \varphi'(t)\, dt. \end{aligned}$$

■

Figure 14.5 below serves as motivation for the following elementary lemma on convex functions.

**Lemma 14.60 (Convex Functions).** *Let $\varphi \in PC^1((a,b) \to \mathbb{R})$ and for $x \in (a,b)$, let*

$$\varphi'(x+) := \lim_{h \downarrow 0} \frac{\varphi(x+h) - \varphi(x)}{h} \text{ and}$$
$$\varphi'(x-) := \lim_{h \uparrow 0} \frac{\varphi(x+h) - \varphi(x)}{h}.$$

*(Of course, $\varphi'(x\pm) = \varphi'(x)$ at points $x \in (a,b)$ where $\varphi$ is differentiable.)*

1. *If $\varphi'(x) \leq \varphi'(y)$ for all $a < x < y < b$ with $x$ and $y$ be points where $\varphi$ is differentiable, then for any $x_0 \in (a,b)$, we have $\varphi'(x_0-) \leq \varphi'(x_0+)$ and for $m \in (\varphi'(x_0-), \varphi'(x_0+))$ we have,*

$$\varphi(x_0) + m(x - x_0) \leq \varphi(x) \quad \forall\, x_0, x \in (a,b). \tag{14.29}$$

[8] $PC^1$ denotes the space of piecewise $C^1$ – functions, i.e. $\varphi \in PC^1((a,b) \to \mathbb{R})$ means the $\varphi$ is continuous and there are a finite number of points,

$$\{a = a_0 < a_1 < a_2 < \dots < a_{n-1} < a_n = b\},$$

such that $\varphi|_{[a_{j-1}, a_j] \cap (a,b)}$ is $C^1$ for all $j = 1, 2, \dots, n$.

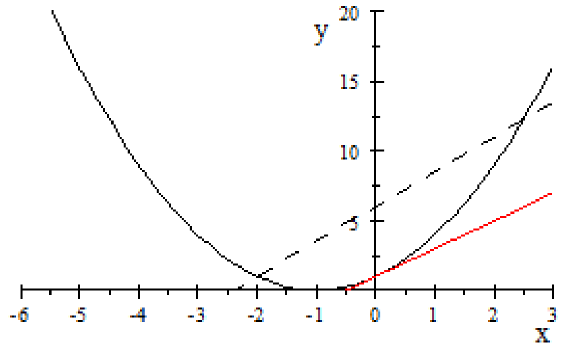



**Fig. 14.5.** A convex function, $\varphi$, along with a cord and a tangent line. Notice that the tangent line is always below $\varphi$ and the cord lies above $\varphi$ between the points of intersection of the cord with the graph of $\varphi$.

2. *If $\varphi \in PC^2((a,b) \to \mathbb{R})$[9] with $\varphi''(x) \geq 0$ for almost all $x \in (a,b)$, then Eq. (14.29) holds with $m = \varphi'(x_0)$.*
3. *If either of the hypothesis in items 1. and 2. above hold then $\varphi$ is convex.*

*(This lemma applies to the functions, $e^{\lambda x}$ for all $\lambda \in \mathbb{R}$, $|x|^\alpha$ for $\alpha > 1$, and $-\ln x$ to name a few examples. See Appendix 14.9 below for much more on convex functions.)*

**Proof. 1.** If $x_0$ is a point where $\varphi$ is not differentiable and $h > 0$ is small, by the mean value theorem, for all $h > 0$ small, there exists $c_+(h) \in (x_0, x_0 + h)$ and $c_-(h) \in (x_0 - h, x_0)$ such that

$$\frac{\varphi(x_0 - h) - \varphi(x_0)}{-h} = \varphi'(c_-(h)) \leq \varphi'(c_+(h)) = \frac{\varphi(x_0 + h) - \varphi(x_0)}{h}.$$

Letting $h \downarrow 0$ in this equation shows $\varphi'(x_0-) \leq \varphi'(x_0+)$. Furthermore if $x < x_0 < y$ with $x$ and $y$ being points of differentiability of $\varphi$, the for small $h > 0$,

$$\varphi'(x) \leq \varphi'(c_-(h)) \leq \varphi'(c_+(h)) \leq \varphi'(y).$$

Letting $h \downarrow 0$ in these inequalities shows,

$$\varphi'(x) \leq \varphi'(x_0-) \leq \varphi'(x_0+) \leq \varphi'(y). \tag{14.30}$$

Now let $m \in (\varphi'(x_0-), \varphi'(x_0+))$. By the fundamental theorem of calculus in Lemma 14.59 and making use of Eq. (14.30), if $x > x_0$ then

$$\varphi(x) - \varphi(x_0) = \int_{x_0}^x \varphi'(t)\,dt \geq \int_{x_0}^x m\ dt = m(x - x_0)$$

and if $x < x_0$, then

$$\varphi(x_0) - \varphi(x) = \int_x^{x_0} \varphi'(t)\,dt \leq \int_x^{x_0} m\ dt = m(x_0 - x).$$

These two equations implies Eq. (14.29) holds.

**2.** Notice that $\varphi' \in PC^1((a,b))$ and therefore,

$$\varphi'(y) - \varphi'(x) = \int_x^y \varphi''(t)\,dt \geq 0 \text{ for all } a < x \leq y < b$$

which shows that item 1. may be used.

**Alternatively;** by Taylor's theorem with integral remainder (see Eq. (9.63) with $F = \varphi$, $a = x_0$, and $b = x$) implies

$$\begin{aligned}\varphi(x) &= \varphi(x_0) + \varphi'(x_0)(x - x_0) + (x - x_0)^2 \int_0^1 \varphi''(x_0 + \tau(x - x_0))(1 - \tau)\,d\tau \\ &\geq \varphi(x_0) + \varphi'(x_0)(x - x_0).\end{aligned}$$

**3.** For any $\xi \in (a,b)$, let $h_\xi(x) := \varphi(x_0) + \varphi'(x_0)(x - x_0)$. By Eq. (14.29) we know that $h_\xi(x) \leq \varphi(x)$ for all $\xi, x \in (a,b)$ with equality when $\xi = x$ and therefore,

$$\varphi(x) = \sup_{\xi \in (a,b)} h_\xi(x).$$

Since $h_\xi$ is an affine function for each $\xi \in (a,b)$, it follows that

$$h_\xi(x_t) = (1 - t)h_\xi(x_0) + th_\xi(x_1) \leq (1 - t)\varphi(x_0) + t\varphi(x_1)$$

for all $t \in [0,1]$. Thus we may conclude that

$$\varphi(x_t) = \sup_{\xi \in (a,b)} h_\xi(x_t) \leq (1 - t)\varphi(x_0) + t\varphi(x_1)$$

as desired.

*For fun, here are three more proofs of Eq. (14.27) under the hypothesis of item 2. Clearly these proofs may be omitted.

**3a.** By Lemma 14.58 below it suffices to show either

[9] $PC^2$ denotes the space of piecewise $C^2$ – functions, i.e. $\varphi \in PC^2((a,b) \to \mathbb{R})$ means the $\varphi$ is $C^1$ and there are a finite number of points,

$$\{a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b\},$$

such that $\varphi|_{[a_{j-1},a_j] \cap (a,b)}$ is $C^2$ for all $j = 1, 2, \ldots, n$.

$$\frac{d}{dx}\frac{\varphi(y)-\varphi(x)}{y-x}\geq 0 \text{ or } \frac{d}{dy}\frac{\varphi(y)-\varphi(x)}{y-x}\geq 0 \text{ for } a<x<y<b.$$

For the first case,

$$\begin{aligned}\frac{d}{dx}\frac{\varphi(y)-\varphi(x)}{y-x} &= \frac{\varphi(y)-\varphi(x)-\varphi'(x)(y-x)}{(y-x)^2}\\ &= \int_0^1 \varphi''(x+t(y-x))(1-t)\,dt \geq 0.\end{aligned}$$

Similarly,

$$\frac{d}{dy}\frac{\varphi(y)-\varphi(x)}{y-x} = \frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2}$$

where we now use,

$$\varphi(x)-\varphi(y) = \varphi'(y)(x-y)+(x-y)^2\int_0^1 \varphi''(y+t(x-y))(1-t)\,dt$$

so that

$$\frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2} = (x-y)^2\int_0^1 \varphi''(y+t(x-y))(1-t)\,dt \geq 0$$

again.

**3b.** Let

$$f(t) := \varphi(u)+t(\varphi(v)-\varphi(u))-\varphi(u+t(v-u)).$$

Then $f(0)=f(1)=0$ with $\ddot{f}(t) = -(v-u)^2\varphi''(u+t(v-u)) \leq 0$ for almost all $t$. By the mean value theorem, there exists, $t_0\in(0,1)$ such that $\dot{f}(t_0)=0$ and then by the fundamental theorem of calculus it follows that

$$\dot{f}(t) = \int_{t_0}^t \ddot{f}(\tau)\,dt.$$

In particular, $\dot{f}(t)\leq 0$ for $t>t_0$ and $\dot{f}(t)\geq 0$ for $t<t_0$ and hence $f(t)\geq f(1)=0$ for $t\geq t_0$ and $f(t)\geq f(0)=0$ for $t\leq t_0$, i.e. $f(t)\geq 0$.

**3c.** Let $h:[0,1]\to\mathbb{R}$ be a piecewise $C^2-$function. Then by the fundamental theorem of calculus and integration by parts,

$$h(t) = h(0)+\int_0^t \dot{h}(\tau)\,d\tau = h(0)+t\dot{h}(t)-\int_0^t \ddot{h}(\tau)\,\tau d\tau$$

and

$$h(1) = h(t)+\int_t^1 \dot{h}(\tau)\,d(\tau-1) = h(t)-(t-1)\dot{h}(t)-\int_t^1 \ddot{h}(\tau)(\tau-1)\,d\tau.$$

Thus we have shown,

$$\begin{aligned}h(t) &= h(0)+t\dot{h}(t)-\int_0^t \ddot{h}(\tau)\,\tau d\tau \text{ and}\\ h(t) &= h(1)+(t-1)\dot{h}(t)+\int_t^1 \ddot{h}(\tau)(\tau-1)\,d\tau.\end{aligned}$$

So if we multiply the first equation by $(1-t)$ and add to it the second equation multiplied by $t$ shows,

$$h(t) = (1-t)h(0)+th(1)-\int_0^1 G(t,\tau)\ddot{h}(\tau)\,d\tau, \tag{14.31}$$

where

$$G(t,\tau) := \begin{cases}\tau(1-t) \text{ if } \tau\leq t\\ t(1-\tau) \text{ if } \tau\geq t\end{cases}.$$

(The function $G(t,\tau)$ is the "Green's function" for the operator $-d^2/dt^2$ on $[0,1]$ with Dirichlet boundary conditions. The formula in Eq. (14.31) is a standard representation formula for $h(t)$ which appears naturally in the study of harmonic functions.)

We now take $h(t) := \varphi(x_0+t(x_1-x_0))$ in Eq. (14.31) to learn

$$\begin{aligned}\varphi(x_0+t(x_1-x_0)) &= (1-t)\varphi(x_0)+t\varphi(x_1)\\ &\quad -(x_1-x_0)^2\int_0^1 G(t,\tau)\ddot{\varphi}(x_0+\tau(x_1-x_0))\,d\tau\\ &\leq (1-t)\varphi(x_0)+t\varphi(x_1),\end{aligned}$$

because $\ddot{\varphi}\geq 0$ and $G(t,\tau)\geq 0$. ■

*Example 14.61.* The functions $\exp(x)$ and $-\log(x)$ are convex and $|x|^p$ is convex iff $p\geq 1$ as follows from Lemma 14.60.

*Example 14.62 (Proof of Lemma 12.42).* Taking $\varphi(x)=e^{-x}$ in Lemma 14.60, Eq. (14.29) with $x_0=0$ implies (see Figure 12.1),

$$1-x\leq\varphi(x)=e^{-x} \text{ for all } x\in\mathbb{R}.$$

Taking $\varphi(x)=e^{-2x}$ in Lemma 14.60, Eq. (14.27) with $x_0=0$ and $x_1=1$ implies, for all $t\in[0,1]$,

$$\begin{aligned} e^{-t} &\leq \varphi\left((1-t)\,0 + t\frac{1}{2}\right) \\ &\leq (1-t)\,\varphi(0) + t\varphi\left(\frac{1}{2}\right) = 1 - t + te^{-1} \leq 1 - \frac{1}{2}t, \end{aligned}$$

wherein the last equality we used $e^{-1} < \frac{1}{2}$. Taking $t = 2x$ in this equation then gives (see Figure 12.2)

$$e^{-2x} \leq 1 - x \text{ for } 0 \leq x \leq \frac{1}{2}. \tag{14.32}$$

**Theorem 14.63.** *Suppose that $\varphi : (a,b) \to \mathbb{R}$ is convex and for $x, y \in (a,b)$ with $x < y$, let*[10]

$$F(x,y) := \frac{\varphi(y) - \varphi(x)}{y - x}.$$

*Then;*

1. *$F(x,y)$ is increasing in each of its arguments.*
2. *The following limits exist,*

$$\varphi'_+(x) := F(x, x+) := \lim_{y \downarrow x} F(x,y) < \infty \text{ and} \tag{14.33}$$

$$\varphi'_-(y) := F(y-, y) := \lim_{x \uparrow y} F(x,y) > -\infty. \tag{14.34}$$

3. *The functions, $\varphi'_\pm$ are both increasing functions and further satisfy,*

$$-\infty < \varphi'_-(x) \leq \varphi'_+(x) \leq \varphi'_-(y) < \infty \;\forall\; a < x < y < b. \tag{14.35}$$

4. *For any $t \in \left[\varphi'_-(x), \varphi'_+(x)\right]$,*

$$\varphi(y) \geq \varphi(x) + t(y - x) \text{ for all } x, y \in (a,b). \tag{14.36}$$

5. *For $a < \alpha < \beta < b$, let $K := \max\left\{\left|\varphi'_+(\alpha)\right|, \left|\varphi'_-(\beta)\right|\right\}$. Then*

$$|\varphi(y) - \varphi(x)| \leq K|y - x| \text{ for all } x, y \in [\alpha, \beta].$$

   *That is $\varphi$ is Lipschitz continuous on $[\alpha, \beta]$.*
6. *The function $\varphi'_+$ is right continuous and $\varphi'_-$ is left continuous.*
7. *The set of discontinuity points for $\varphi'_+$ and for $\varphi'_-$ are the same as the set of points of non-differentiability of $\varphi$. Moreover this set is at most countable.*

[10] The same formula would define $F(x,y)$ for $x \neq y$. However, since $F(x,y) = F(y,x)$, we would gain no new information by this extension.

**Proof.** BRUCE: The first two items are a repetition of Lemma 14.58.

1. and 2. If we let $h_t = t\varphi(x_1) + (1-t)\varphi(x_0)$, then $(x_t, h_t)$ is on the line segment joining $(x_0, \varphi(x_0))$ to $(x_1, \varphi(x_1))$ and the statement that $\varphi$ is convex is then equivalent of $\varphi(x_t) \leq h_t$ for all $0 \leq t \leq 1$. Since

$$\frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t},$$

the convexity of $\varphi$ is equivalent to

$$\frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} \leq \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \text{ for all } x_0 \leq x_t \leq x_1$$

and to

$$\frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \leq \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t} \text{ for all } x_0 \leq x_t \leq x_1.$$

Convexity also implies

$$\frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} = \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \leq \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t}.$$

These inequalities may be written more compactly as,

$$\frac{\varphi(v) - \varphi(u)}{v - u} \leq \frac{\varphi(w) - \varphi(u)}{w - u} \leq \frac{\varphi(w) - \varphi(v)}{w - v}, \tag{14.37}$$

valid for all $a < u < v < w < b$, again see Figure 14.4. The first (second) inequality in Eq. (14.37) shows $F(x,y)$ is increasing $y$ $(x)$. This then implies the limits in item 2. are monotone and hence exist as claimed.

3. Let $a < x < y < b$. Using the increasing nature of $F$,

$$-\infty < \varphi'_-(x) = F(x-, x) \leq F(x, x+) = \varphi'_+(x) < \infty$$

and

$$\varphi'_+(x) = F(x, x+) \leq F(y-, y) = \varphi'_-(y)$$

as desired.

4. Let $t \in \left[\varphi'_-(x), \varphi'_+(x)\right]$. Then

$$t \leq \varphi'_+(x) = F(x, x+) \leq F(x,y) = \frac{\varphi(y) - \varphi(x)}{y - x}$$

or equivalently,

$$\varphi(y) \geq \varphi(x) + t(y - x) \text{ for } y \geq x.$$

Therefore Eq. (14.36) holds for $y \geq x$. Similarly, for $y < x$,

$$t \geq \varphi'_{-}(x) = F(x-, x) \geq F(y, x) = \frac{\varphi(x) - \varphi(y)}{x - y}$$

or equivalently,

$$\varphi(y) \geq \varphi(x) - t(x - y) = \varphi(x) + t(y - x) \text{ for } y \leq x.$$

Hence we have proved Eq. (14.36) for all $x, y \in (a, b)$.

5. For $a < \alpha \leq x < y \leq \beta < b$, we have

$$\varphi'_{+}(\alpha) \leq \varphi'_{+}(x) = F(x, x+) \leq F(x, y) \leq F(y-, y) = \varphi'_{-}(y) \leq \varphi'_{-}(\beta) \tag{14.38}$$

and in particular,

$$-K \leq \varphi'_{+}(\alpha) \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \varphi'_{-}(\beta) \leq K.$$

This last inequality implies, $|\varphi(y) - \varphi(x)| \leq K(y - x)$ which is the desired Lipschitz bound.

6. For $a < c < x < y < b$, we have $\varphi'_{+}(x) = F(x, x+) \leq F(x, y)$ and letting $x \downarrow c$ (using the continuity of $F$) we learn $\varphi'_{+}(c+) \leq F(c, y)$. We may now let $y \downarrow c$ to conclude $\varphi'_{+}(c+) \leq \varphi'_{+}(c)$. Since $\varphi'_{+}(c) \leq \varphi'_{+}(c+)$, it follows that $\varphi'_{+}(c) = \varphi'_{+}(c+)$ and hence that $\varphi'_{+}$ is right continuous.

Similarly, for $a < x < y < c < b$, we have $\varphi'_{-}(y) \geq F(x, y)$ and letting $y \uparrow c$ (using the continuity of $F$) we learn $\varphi'_{-}(c-) \geq F(x, c)$. Now let $x \uparrow c$ to conclude $\varphi'_{-}(c-) \geq \varphi'_{-}(c)$. Since $\varphi'_{-}(c) \geq \varphi'_{-}(c-)$, it follows that $\varphi'_{-}(c) = \varphi'_{-}(c-)$, i.e. $\varphi'_{-}$ is left continuous.

7. Since $\varphi_{\pm}$ are increasing functions, they have at most countably many points of discontinuity. Letting $x \uparrow y$ in Eq. (14.35), using the left continuity of $\varphi'_{-}$, shows $\varphi'_{-}(y) = \varphi'_{+}(y-)$. Hence if $\varphi'_{-}$ is continuous at $y$, $\varphi'_{-}(y) = \varphi'_{-}(y+) = \varphi'_{+}(y)$ and $\varphi$ is differentiable at $y$. Conversely if $\varphi$ is differentiable at $y$, then

$$\varphi'_{+}(y-) = \varphi'_{-}(y) = \varphi'(y) = \varphi'_{+}(y)$$

which shows $\varphi'_{+}$ is continuous at $y$. Thus we have shown that set of discontinuity points of $\varphi'_{+}$ is the same as the set of points of non-differentiability of $\varphi$. That the discontinuity set of $\varphi'_{-}$ is the same as the non-differentiability set of $\varphi$ is proved similarly. ∎

**Corollary 14.64.** *If $\varphi : (a, b) \to \mathbb{R}$ is a convex function and $D \subset (a, b)$ is a dense set, then*

$$\varphi(y) = \sup_{x \in D} \left[\varphi(x) + \varphi'_{\pm}(x)(y - x)\right] \text{ for all } x, y \in (a, b).$$

**Proof.** Let $\psi_{\pm}(y) := \sup_{x \in D}[\varphi(x) + \varphi_{\pm}(x)(y - x)]$. According to Eq. (14.36) above, we know that $\varphi(y) \geq \psi_{\pm}(y)$ for all $y \in (a, b)$. Now suppose that $x \in (a, b)$ and $x_n \in \Lambda$ with $x_n \uparrow x$. Then passing to the limit in the estimate, $\psi_{-}(y) \geq \varphi(x_n) + \varphi'_{-}(x_n)(y - x_n)$, shows $\psi_{-}(y) \geq \varphi(x) + \varphi'_{-}(x)(y - x)$. Since $x \in (a, b)$ is arbitrary we may take $x = y$ to discover $\psi_{-}(y) \geq \varphi(y)$ and hence $\varphi(y) = \psi_{-}(y)$. The proof that $\varphi(y) = \psi_{+}(y)$ is similar. ∎

**Lemma 14.65.** *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is a non-decreasing function such that*

$$\varphi\left(\frac{1}{2}(x + y)\right) \leq \frac{1}{2}[\varphi(x) + \varphi(y)] \text{ for all } x, y \in (a, b), \tag{14.39}$$

*then $\varphi$ is convex. The result remains true if $\varphi$ is assumed to be continuous rather than non-decreasing.*

**Proof.** Let $x_0, x_1 \in (a, b)$ and $x_t := x_0 + t(x_1 - x_0)$ as above. For $n \in \mathbb{N}$ let $\mathbb{D}_n = \left\{\frac{k}{2^n} : 1 \leq k < 2^n\right\}$. We are going to being by showing Eq. (14.39) implies

$$\varphi(x_t) \leq (1 - t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in \mathbb{D} := \cup_n \mathbb{D}_n. \tag{14.40}$$

We will do this by induction on $n$. For $n = 1$, this follows directly from Eq. (14.39). So now suppose that Eq. (14.40) holds for all $t \in \mathbb{D}_n$ and now suppose that $t = \frac{2k+1}{2^n} \in \mathbb{D}_{n+1}$. Observing that

$$x_t = \frac{1}{2}\left(x_{\frac{k}{2^{n-1}}} + x_{\frac{k+1}{2^n}}\right)$$

we may again use Eq. (14.39) to conclude,

$$\varphi(x_t) \leq \frac{1}{2}\left(\varphi\left(x_{\frac{k}{2^{n-1}}}\right) + \varphi\left(x_{\frac{k+1}{2^{n-1}}}\right)\right).$$

Then use the induction hypothesis to conclude,

$$\begin{aligned}\varphi(x_t) &\leq \frac{1}{2}\begin{pmatrix} \left(1 - \frac{k}{2^{n-1}}\right)\varphi(x_0) + \frac{k}{2^{n-1}}\varphi(x_1) \\ + \left(1 - \frac{k+1}{2^{n-1}}\right)\varphi(x_0) + \frac{k+1}{2^{n-1}}\varphi(x_1) \end{pmatrix} \\ &= (1 - t)\varphi(x_0) + t\varphi(x_1)\end{aligned}$$

as desired.

For general $t \in (0, 1)$, let $\tau \in \mathbb{D}$ such that $\tau > t$. Since $\varphi$ is increasing and by Eq. (14.40) we conclude,

$$\varphi(x_t) \leq \varphi(x_\tau) \leq (1 - \tau)\varphi(x_0) + \tau\varphi(x_1).$$

We may now let $\tau \downarrow t$ to complete the proof. This same technique clearly also works if we were to assume that $\varphi$ is continuous rather than monotonic. ∎

# 15

# Hilbert Space Basics

**Definition 15.1 (Inner Product Space).** *Let $H$ be a complex vector space. An inner product on $H$ is a function, $\langle\cdot|\cdot\rangle : H \times H \to \mathbb{C}$, such that*

1. $\langle ax + by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ *i.e.* $x \to \langle x|z\rangle$ *is linear.*
2. $\overline{\langle x|y\rangle} = \langle y|x\rangle$.
3. $\|x\|^2 := \langle x|x\rangle \geq 0$ *with* $\|x\|^2 = 0$ *iff* $x = 0$.

Notice that combining properties (1) and (2) that $x \to \langle z|x\rangle$ is conjugate linear for fixed $z \in H$, i.e.

$$\langle z|ax + by\rangle = \bar{a}\langle z|x\rangle + \bar{b}\langle z|y\rangle.$$

The following identity will be used frequently in the sequel without further mention,

$$\begin{aligned}\|x+y\|^2 &= \langle x+y|x+y\rangle = \|x\|^2 + \|y\|^2 + \langle x|y\rangle + \langle y|x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle. \end{aligned} \tag{15.1}$$

**Theorem 15.2 (Schwarz Inequality).** *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, then for all $x, y \in H$*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

*and equality holds iff $x$ and $y$ are linearly dependent.*

**Proof.** If $y = 0$, the result holds trivially. So assume that $y \neq 0$ and observe; if $x = \alpha y$ for some $\alpha \in \mathbb{C}$, then $\langle x|y\rangle = \bar{\alpha}\,\|y\|^2$ and hence

$$|\langle x|y\rangle| = |\alpha|\,\|y\|^2 = \|x\|\|y\|.$$

Now suppose that $x \in H$ is arbitrary, let $z := x - \|y\|^{-2}\langle x|y\rangle y$. (So $\|y\|^{-2}\langle x|y\rangle y$ is the "orthogonal projection" of $x$ along $y$, see Figure 15.1.)Then

$$\begin{aligned} 0 \leq \|z\|^2 &= \left\|x - \frac{\langle x|y\rangle}{\|y\|^2}y\right\|^2 = \|x\|^2 + \frac{|\langle x|y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x|\frac{\langle x|y\rangle}{\|y\|^2}y\rangle \\ &= \|x\|^2 - \frac{|\langle x|y\rangle|^2}{\|y\|^2} \end{aligned}$$

from which it follows that $0 \leq \|y\|^2\|x\|^2 - |\langle x|y\rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^{-2}\langle x|y\rangle y$. ∎

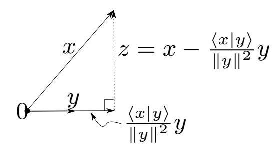



**Fig. 15.1.** The picture behind the proof of the Schwarz inequality.

**Corollary 15.3.** *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x|x\rangle}$. Then the **Hilbertian norm**, $\|\cdot\|$, is a norm on $H$. Moreover $\langle\cdot|\cdot\rangle$ is continuous on $H \times H$, where $H$ is viewed as the normed space $(H, \|\cdot\|)$.*

**Proof.** If $x, y \in H$, then, using Schwarz's inequality,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.\end{aligned}$$

Taking the square root of this inequality shows $\|\cdot\|$ satisfies the triangle inequality.

Checking that $\|\cdot\|$ satisfies the remaining axioms of a norm is now routine and will be left to the reader. If $x, y, \Delta x, \Delta y \in H$, then

$$\begin{aligned}|\langle x + \Delta x|y + \Delta y\rangle - \langle x|y\rangle| &= |\langle x|\Delta y\rangle + \langle \Delta x|y\rangle + \langle \Delta x|\Delta y\rangle| \\ &\leq \|x\|\|\Delta y\| + \|y\|\|\Delta x\| + \|\Delta x\|\|\Delta y\| \\ &\to 0 \text{ as } \Delta x, \Delta y \to 0,\end{aligned}$$

from which it follows that $\langle\cdot|\cdot\rangle$ is continuous. ∎

**Definition 15.4.** *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, we say $x, y \in H$ are **orthogonal** and write $x \perp y$ iff $\langle x|y\rangle = 0$. More generally if $A \subset H$ is a set, $x \in H$ is **orthogonal to** $A$ (write $x \perp A$) iff $\langle x|y\rangle = 0$ for all $y \in A$. Let $A^\perp = \{x \in H : x \perp A\}$ be the set of vectors orthogonal to $A$. A subset $S \subset H$ is an **orthogonal set** if $x \perp y$ for all distinct elements $x, y \in S$. If $S$ further satisfies, $\|x\| = 1$ for all $x \in S$, then $S$ is said to be an **orthonormal set.***

**Proposition 15.5.** *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. ***(Parallelogram Law)***
$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2 \tag{15.2}$$
*for all $a, b \in H$.*
2. ***(Pythagorean Theorem)*** *If $S \subset_f H$ is a finite orthogonal set, then*
$$\left\| \sum_{x \in S} x \right\|^2 = \sum_{x \in S} \|x\|^2. \tag{15.3}$$
3. *If $A \subset H$ is a set, then $A^\perp$ is a* ***closed*** *linear subspace of $H$.*

**Proof.** I will assume that $H$ is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations;

$$\begin{aligned} &\|a+b\|^2 + \|a-b\|^2 \\ &\quad = \|a\|^2 + \|b\|^2 + 2\mathrm{Re}\langle a|b\rangle + \|a\|^2 + \|b\|^2 - 2\mathrm{Re}\langle a|b\rangle \\ &\quad = 2\|a\|^2 + 2\|b\|^2, \end{aligned}$$

and

$$\begin{aligned} \left\| \sum_{x \in S} x \right\|^2 &= \langle \sum_{x \in S} x | \sum_{y \in S} y\rangle = \sum_{x,y \in S} \langle x|y\rangle \\ &= \sum_{x \in S} \langle x|x\rangle = \sum_{x \in S} \|x\|^2. \end{aligned}$$

Item 3. is a consequence of the continuity of $\langle \cdot|\cdot\rangle$ and the fact that

$$A^\perp = \cap_{x \in A} \mathrm{Nul}(\langle \cdot|x\rangle)$$

where $\mathrm{Nul}(\langle \cdot|x\rangle) = \{y \in H : \langle y|x\rangle = 0\}$ – a closed subspace of $H$. Alternatively, if $x_n \in A^\perp$ and $x_n \to x$ in $H$, then

$$0 = \lim_{n\to\infty} 0 = \lim_{n\to\infty} \langle x_n|a\rangle = \left\langle \lim_{n\to\infty} x_n|a \right\rangle = \langle x|a\rangle \ \forall\ a \in A$$

which shows that $x \in A^\perp$. ■

**Definition 15.6.** *A* ***Hilbert space*** *is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

*Example 15.7.* For any measure space, $(\Omega, \mathcal{B}, \mu)$, $H := L^2(\mu)$ with inner product,

$$\langle f|g\rangle = \int_\Omega f(\omega)\, \bar{g}(\omega)\, d\mu(\omega)$$

is a Hilbert space – see Theorem 14.25 for the completeness assertion.

**Definition 15.8.** *A subset $C$ of a vector space $X$ is said to be convex if for all $x, y \in C$ the line segment $[x, y] := \{tx + (1-t)y : 0 \le t \le 1\}$ joining $x$ to $y$ is contained in $C$ as well. (Notice that any vector subspace of $X$ is convex.)*

**Theorem 15.9 (Best Approximation Theorem).** *Suppose that $H$ is a Hilbert space and $M \subset H$ is a closed convex subset of $H$. Then for any $x \in H$ there exists a unique $y \in M$ such that*

$$\|x-y\| = d(x, M) = \inf_{z \in M} \|x-z\|.$$

*Moreover, if $M$ is a vector subspace of $H$, then the point $y$ may also be characterized as the unique point in $M$ such that $(x-y) \perp M$.*

**Proof.** Let $x \in H$, $\delta := d(x, M)$, and $y, z \in M$. Then by the parallelogram law (Eq. (15.2) and the fact that $\frac{y+z}{2} \in M$,

$$\begin{aligned} \|z-y\|^2 &= \|(z-x)+(x-y)\|^2 \\ &= 2\|z-x\|^2 + 2\|x-y\|^2 - \|(z-x)-(x-y)\|^2 \\ &= 2\|z-x\|^2 + 2\|x-y\|^2 - 4\left\| x - \frac{y+z}{2} \right\|^2 \\ &\le 2\|z-x\|^2 + 2\|x-y\|^2 - 4\delta^2, \end{aligned} \tag{15.4}$$

see Figure 15.2.

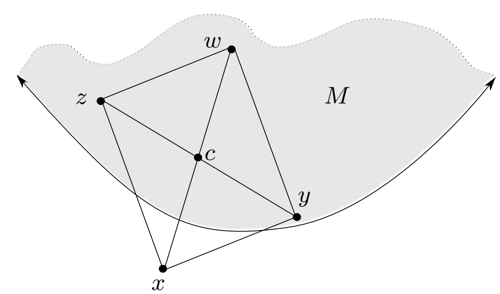



**Fig. 15.2.** In this figure $y, z \in M$ and by convexity, $c = (z+y)/2 \in M$.

**Uniqueness.** If $y, z \in M$ minimize the distance to $x$, then $\|y-x\| = \delta = \|z-x\|$ and it follows from Eq. (15.4) that $y = z$.

**Existence.** Let $y_n \in M$ be chosen such that $\|y_n - x\| = \delta_n \to \delta = d(x, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (15.4) shows

$$\|y_n - y_m\|^2 \le 2\delta_m^2 + 2\delta_n^2 - 4\delta^2 \to 0 \text{ as } m, n \to \infty.$$

Therefore, by completeness of $H$, $\{y_n\}_{n=1}^\infty$ is convergent. Because $M$ is closed, $y := \lim_{n\to\infty} y_n \in M$ and because the norm is continuous,

$$\|y - x\| = \lim_{n\to\infty} \|y_n - x\| = \delta = d(x, M).$$

So $y$ is the desired point in $M$ which is closest to $x$.

**Orthogonality property.** Now suppose $M$ is a closed subspace of $H$ and $x \in H$. Let $y \in M$ be the closest point in $M$ to $x$. Then for $w \in M$, the function

$$g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t\mathrm{Re}\langle x - y|w\rangle + t^2\|w\|^2$$

has a minimum at $t = 0$ and therefore $0 = g'(0) = -2\mathrm{Re}\langle x - y|w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$, see Figure 15.3. Finally

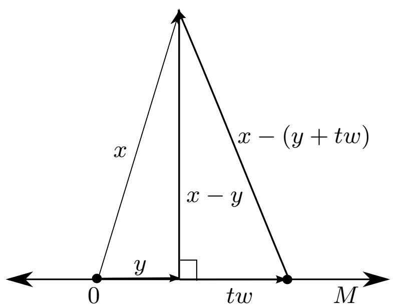



**Fig. 15.3.** The orthogonality relationships of closest points.

suppose $y \in M$ is any point such that $(x - y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2$$

which shows $d(x, M)^2 \geq \|x - y\|^2$. That is to say $y$ is the point in $M$ closest to $x$. ■

*Remark 15.10.* If $M$ is a finite dimensional subspace of $H$ and and $\{e_i\}_{i=1}^n$ is an orthonormal basis for $M$, then $(w := x - \sum_{i=1}^n \langle x|e_i\rangle e_i) \perp M$. Therefore,

$$\|x\|^2 = \|w\|^2 + \left\|\sum_{i=1}^n \langle x|e_i\rangle e_i\right\| = d^2(x, M) + \sum_{i=1}^n |\langle x|e_i\rangle|^2 \tag{15.5}$$

and

$$d^2(x, M) = \left\|x^2\right\| - \sum_{i=1}^n |\langle x|e_i\rangle|^2 . \tag{15.6}$$

Suppose $u \in H \setminus \{0\}$, $M = \mathrm{span}\{u\}$, and $\hat{u} := u/\|u\|$. Then from Eq. (15.6),

$$0 \leq d^2(x, M) = \left\|x^2\right\| - |\langle x|\hat{u}\rangle|^2 = \left\|x^2\right\| - |\langle x|u\rangle|^2 / \|u\|^2$$

from which the Cauchy-Schwarz inequality,

$$|\langle x|u\rangle|^2 \leq \|x\|^2 \left\|u^2\right\|$$

follows. Moreover the proof shows that equality holds iff $x \in M$, i.e. $x = \lambda u$ for some $\lambda \in \mathbb{C}$.

**Notation 15.11** *If $A : X \to Y$ is a linear operator between two normed spaces, we let*

$$\|A\| := \sup_{x \in X\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X} = \sup_{\|x\|_X = 1} \|Ax\|_Y .$$

*We refer to $\|A\|$ as the operator norm of $A$ and call $A$ a bounded operator if $\|A\| < \infty$. We further let $L(X, Y)$ be the set of bounded operators from $X$ to $Y$.*

**Exercise 15.1.** Show that a linear operator, $A : X \to Y$, is a bounded iff it is continuous.

**Definition 15.12.** *Suppose that $A : H \to H$ is a bounded operator. The **adjoint** of $A$, denoted $A^*$, is the unique operator $A^* : H \to H$ such that $\langle Ax|y\rangle = \langle x|A^*y\rangle$. (The proof that $A^*$ exists and is unique will be given in Proposition 15.17 below.) A bounded operator $A : H \to H$ is **self - adjoint** or **Hermitian** if $A = A^*$.*

**Definition 15.13.** *$H$ be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of $H$ onto $M$ is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in $M$ such that $(x - P_M(x)) \perp M$, i.e. $P_M(x)$ is the unique element in $M$ such that*

$$\langle x|m\rangle = \langle P_M(x)|m\rangle \text{ for all } m \in M. \tag{15.7}$$

Given a linear transformation $A$, we will let $\mathrm{Ran}(A)$ and $\mathrm{Nul}(A)$ denote the **range** and the **null-space** of $A$ respectively.

**Theorem 15.14 (Projection Theorem).** *Let $H$ be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection $P_M$ satisfies:*

1. *$P_M$ is linear and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ ($P_M$ is a projection).*
3. *$P_M^* = P_M$ ($P_M$ is self-adjoint).*
4. *$\mathrm{Ran}(P_M) = M$ and $\mathrm{Nul}(P_M) = M^\perp$.*

5. *If $N \subset M \subset H$ is another closed subspace, the $P_N P_M = P_M P_N = P_N$.*

**Proof.**

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{C}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. $P_M$ is linear.
2. Obviously $\mathrm{Ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,
$$\begin{aligned}\langle P_M x|y\rangle &= \langle P_M x|P_M y + y - P_M y\rangle = \langle P_M x|P_M y\rangle \\ &= \langle P_M x + (x - P_M x)|P_M y\rangle = \langle x|P_M y\rangle.\end{aligned}$$
4. We have already seen, $\mathrm{Ran}(P_M) = M$ and $P_M x = 0$ iff $x = x - 0 \in M^\perp$, i.e. $\mathrm{Nul}(P_M) = M^\perp$.
5. If $N \subset M \subset H$ it is clear that $P_M P_N = P_N$ since $P_M = Id$ on $N = \mathrm{Ran}(P_N) \subset M$. Taking adjoints gives the other identity, namely that $P_N P_M = P_N$.
**Alternative proof 1 of** $P_N P_M = P_N$. Let $x \in H$ and $h = P_N P_M x$. Then by definition of $h$, $\langle h|n\rangle = \langle P_M x|n\rangle$ for all $n \in N$. However by definition of $P_M x$ we also know that $\langle P_M x|n\rangle = \langle x|n\rangle$ for all $n \in N \subset M$ and so we may conclude that $\langle h|n\rangle = \langle x|n\rangle$ for all $n \in N$, i.e. $h = P_N x$.
**Alternative proof 2 of** $P_N P_M = P_N$**.** If $x \in H$, then $(x - P_M x) \perp M$ and therefore $(x - P_M x) \perp N$. We also have $(P_M x - P_N P_M x) \perp N$ and therefore,
$$x - P_N P_M x = (x - P_M x) + (P_M x - P_N P_M x) \in N^\perp$$
which shows $P_N P_M x = P_N x$.
**Alternative proof 3 of** $P_N P_M = P_N$. If $x \in H$ and $n \in N$, we have
$$\langle P_N P_M x|n\rangle = \langle P_M x|P_N n\rangle = \langle P_M x|n\rangle = \langle x|P_M n\rangle = \langle x|n\rangle .$$
Since this holds for all $n$ we may conclude that $P_N P_M x = P_N x$.

■

**Corollary 15.15.** *If $M \subset H$ is a proper closed subspace of a Hilbert space $H$, then $H = M \oplus M^\perp$.*

**Proof.** Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x - y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x|x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

**Exercise 15.2.** Suppose $M$ is a subset of $H$, then $M^{\perp\perp} = \overline{\mathrm{span}(M)}$ where (as usual), $\mathrm{span}\,(M)$ denotes all finite linear combinations of elements from $M$.

**Theorem 15.16 (Riesz Theorem).** *Let $H^*$ be the dual space of $H$, i.e. $f \in H^*$ iff $f : H \to \mathbb{F}$ is linear and continuous. The map*
$$z \in H \xrightarrow{j} \langle \cdot|z\rangle \in H^* \tag{15.8}$$
*is a conjugate linear[1] isometric isomorphism, where for $f \in H^*$ we let,*
$$\|f\|_{H^*} := \sup_{x \in H \setminus \{0\}} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|=1} |f(x)| .$$

**Proof.** Let $f \in H^*$ and $M = \mathrm{Nul}(f)$ – a closed proper subspace of $H$ since $f$ is continuous. If $f = 0$, then clearly $f(\cdot) = \langle \cdot|0\rangle$ . If $f \neq 0$ there exists $y \in H \setminus M$. Then for any $\alpha \in \mathbb{C}$ we have $e := \alpha(y - P_M y) \in M^\perp$. We now choose $\alpha$ so that $f(e) = 1$. Hence if $x \in H$,
$$f(x - f(x)e) = f(x) - f(x) f(e) = f(x) - f(x) = 0,$$
which shows $x - f(x)e \in M$. As $e \in M^\perp$ it follows that
$$0 = \langle x - f(x)e|e\rangle = \langle x|e\rangle - f(x)\|e\|^2$$
which shows $f(\cdot) = \langle \cdot|z\rangle = jz$ where $z := e/\|e\|^2$ and thus $j$ is surjective.

The map $j$ is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,
$$|\langle x|z\rangle| \le \|x\| \|z\| \text{ for all } x \in H$$
with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot|z\rangle\|_{H^*} = \|z\|$ . Therefore $j$ is isometric and this implies $j$ is injective.

■

**Proposition 15.17 (Adjoints).** *Let $H$ and $K$ be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*
$$\langle Ax|y\rangle_K = \langle x|A^* y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{15.9}$$
*Moreover, for all $A, B \in L(H, K)$ and $\lambda \in \mathbb{C}$,*

[1] Recall that $j$ is conjugate linear if
$$j(z_1 + \alpha z_2) = jz_1 + \bar{\alpha} jz_2$$
for all $z_1, z_2 \in H$ and $\alpha \in \mathbb{C}$.

1. $(A+\lambda B)^* = A^* + \bar{\lambda} B^*$,
2. $A^{**} := (A^*)^* = A$,
3. $\|A^*\| = \|A\|$ *and*
4. $\|A^*A\| = \|A\|^2$.
5. *If* $K = H$, *then* $(AB)^* = B^*A^*$. *In particular* $A \in L(H)$ *has a bounded inverse iff* $A^*$ *has a bounded inverse and* $(A^*)^{-1} = \left(A^{-1}\right)^*$.

**Proof.** For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in $H^*$ and therefore there exists, by Theorem 15.16, a unique vector $z \in H$ (we will denote this $z$ by $A^*(y)$) such that

$$\langle Ax|y\rangle_K = \langle x|z\rangle_H \text{ for all } x \in H.$$

This shows there is a unique map $A^* : K \to H$ such that $\langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$.

To see $A^*$ is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,

$$\begin{aligned}\langle Ax|y_1 + \lambda y_2\rangle_K &= \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K \\ &= \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_H \\ &= \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_H\end{aligned}$$

and by the uniqueness of $A^*(y_1 + \lambda y_2)$ we find

$$A^*(y_1 + \lambda y_2) = A^*(y_1) + \lambda A^*(y_2).$$

This shows $A^*$ is linear and so we will now write $A^*y$ instead of $A^*(y)$.

Since

$$\langle A^*y|x\rangle_H = \overline{\langle x|A^*y\rangle_H} = \overline{\langle Ax|y\rangle_K} = \langle y|Ax\rangle_K$$

it follows that $A^{**} = A$. The assertion that $(A+\lambda B)^* = A^* + \bar{\lambda}B^*$ is Exercise 15.3.

**Items 3. and 4.** Making use of Schwarz's inequality (Theorem 15.2), we have

$$\begin{aligned}\|A^*\| &= \sup_{k\in K:\|k\|=1} \|A^*k\| \\ &= \sup_{k\in K:\|k\|=1} \sup_{h\in H:\|h\|=1} |\langle A^*k|h\rangle| \\ &= \sup_{h\in H:\|h\|=1} \sup_{k\in K:\|k\|=1} |\langle k|Ah\rangle| = \sup_{h\in H:\|h\|=1} \|Ah\| = \|A\|\end{aligned}$$

so that $\|A^*\| = \|A\|$. Since

$$\|A^*A\| \le \|A^*\| \, \|A\| = \|A\|^2$$

and

$$\begin{aligned}\|A\|^2 &= \sup_{h\in H:\|h\|=1} \|Ah\|^2 = \sup_{h\in H:\|h\|=1} |\langle Ah|Ah\rangle| \\ &= \sup_{h\in H:\|h\|=1} |\langle h|A^*Ah\rangle| \le \sup_{h\in H:\|h\|=1} \|A^*Ah\| = \|A^*A\| \qquad (15.10)\end{aligned}$$

we also have $\|A^*A\| \le \|A\|^2 \le \|A^*A\|$ which shows $\|A\|^2 = \|A^*A\|$.

Alternatively, from Eq. (15.10),

$$\|A\|^2 \le \|A^*A\| \le \|A\| \, \|A^*\| \qquad (15.11)$$

which then implies $\|A\| \le \|A^*\|$. Replacing $A$ by $A^*$ in this last inequality shows $\|A^*\| \le \|A\|$ and hence that $\|A^*\| = \|A\|$. Using this identity back in Eq. (15.11) proves $\|A\|^2 = \|A^*A\|$.

Now suppose that $K = H$. Then

$$\langle ABh|k\rangle = \langle Bh|A^*k\rangle = \langle h|B^*A^*k\rangle$$

which shows $(AB)^* = B^*A^*$. If $A^{-1}$ exists then

$$\begin{aligned}&\left(A^{-1}\right)^* A^* = \left(AA^{-1}\right)^* = I^* = I \text{ and} \\ &A^*\left(A^{-1}\right)^* = \left(A^{-1}A\right)^* = I^* = I.\end{aligned}$$

This shows that $A^*$ is invertible and $(A^*)^{-1} = \left(A^{-1}\right)^*$. Similarly if $A^*$ is invertible then so is $A = A^{**}$. ■

**Exercise 15.3.** Let $H, K, M$ be Hilbert spaces, $A, B \in L(H,K)$, $C \in L(K,M)$ and $\lambda \in \mathbb{C}$. Show $(A+\lambda B)^* = A^* + \bar{\lambda}B^*$ and $(CA)^* = A^*C^* \in L(M,H)$.

**Exercise 15.4.** Let $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z|w\rangle_H = z \cdot \bar{w}$ for $z, w \in H$. Let $A$ be an $m \times n$ matrix thought of as a linear operator from $H$ to $K$. Show the matrix associated to $A^* : K \to H$ is the conjugate transpose of $A$.

**Lemma 15.18.** *Suppose* $A : H \to K$ *is a bounded operator, then:*

1. $\mathrm{Nul}(A^*) = \mathrm{Ran}(A)^\perp$.
2. $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$.
3. *if* $K = H$ *and* $V \subset H$ *is an* $A$ *– invariant subspace (i.e.* $A(V) \subset V$*), then* $V^\perp$ *is* $A^*$ *– invariant.*

**Proof.** An element $y \in K$ is in $\mathrm{Nul}(A^*)$ iff $0 = \langle A^*y|x\rangle = \langle y|Ax\rangle$ for all $x \in H$ which happens iff $y \in \mathrm{Ran}(A)^\perp$. Because, by Exercise 15.2, $\overline{\mathrm{Ran}(A)} = \mathrm{Ran}(A)^{\perp\perp}$, and so by the first item, $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$. Now suppose $A(V) \subset V$ and $y \in V^\perp$, then

$$\langle A^*y|x\rangle = \langle y|Ax\rangle = 0 \text{ for all } x \in V$$

which shows $A^*y \in V^\perp$. ■

The next elementary theorem (referred to as the bounded linear transformation theorem, or B.L.T. theorem for short) is often useful.

**Theorem 15.19 (B. L. T. Theorem).** *Suppose that $Z$ is a normed space, $X$ is a Banach space, and $\mathcal{S} \subset Z$ is a dense linear subspace of $Z$. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \le C\|z\|$ for all $z \in \mathcal{S}$), then $T$ has a unique extension to an element $\bar{T} \in L(Z, X)$ and this extension still satisfies*

$$\|\bar{T}z\| \le C\|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

**Proof.** Let $z \in Z$ and choose $z_n \in \mathcal{S}$ such that $z_n \to z$. Since

$$\|Tz_m - Tz_n\| \le C\|z_m - z_n\| \to 0 \text{ as } m, n \to \infty,$$

it follows by the completeness of $X$ that $\lim_{n\to\infty} Tz_n =: \bar{T}z$ exists. Moreover, if $w_n \in \mathcal{S}$ is another sequence converging to $z$, then

$$\|Tz_n - Tw_n\| \le C\|z_n - w_n\| \to C\|z - z\| = 0$$

and therefore $\bar{T}z$ is well defined. It is now a simple matter to check that $\bar{T} : Z \to X$ is still linear and that

$$\|\bar{T}z\| = \lim_{n\to\infty} \|Tz_n\| \le \lim_{n\to\infty} C\|z_n\| = C\|z\| \text{ for all } x \in Z.$$

Thus $\bar{T}$ is an extension of $T$ to all of the $Z$. The uniqueness of this extension is easy to prove and will be left to the reader. ■

## 15.1 Compactness Results for $L^p$ – Spaces*

In this section we are going to identify the sequentially "weak" compact subsets of $L^p(\Omega, \mathcal{B}, P)$ for $1 \le p < \infty$, where $(\Omega, \mathcal{B}, P)$ is a probability space. The key to our proofs will be the following Hilbert space compactness result.

**Theorem 15.20.** *Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in a Hilbert space $H$ (i.e. $C := \sup_n \|x_n\| < \infty$), then there exists a sub-sequence, $y_k := x_{n_k}$ and an $x \in H$ such that $\lim_{k\to\infty} \langle y_k|h\rangle = \langle x|h\rangle$ for all $h \in H$. We say that $y_k$ converges to $x$ weakly in this case and denote this by $y_k \overset{w}{\to} x$.*

**Proof.** Let $H_0 := \overline{\operatorname{span}(x_k : k \in \mathbb{N})}$. Then $H_0$ is a closed separable Hilbert subspace of $H$ and $\{x_k\}_{k=1}^\infty \subset H_0$. Let $\{h_n\}_{n=1}^\infty$ be a countable dense subset of $H_0$. Since $|\langle x_k|h_n\rangle| \le \|x_k\|\|h_n\| \le C\|h_n\| < \infty$, the sequence, $\{\langle x_k|h_n\rangle\}_{k=1}^\infty \subset \mathbb{C}$, is bounded and hence has a convergent sub-sequence for all $n \in \mathbb{N}$. By the Cantor's diagonalization argument we can find a a sub-sequence, $y_k := x_{n_k}$, of $\{x_n\}$ such that $\lim_{k\to\infty} \langle y_k|h_n\rangle$ exists for all $n \in \mathbb{N}$.

We now show $\varphi(z) := \lim_{k\to\infty} \langle y_k|z\rangle$ exists for all $z \in H_0$. Indeed, for any $k, l, n \in \mathbb{N}$, we have

$$\begin{aligned}|\langle y_k|z\rangle - \langle y_l|z\rangle| &= |\langle y_k - y_l|z\rangle| \le |\langle y_k - y_l|h_n\rangle| + |\langle y_k - y_l|z - h_n\rangle| \\ &\le |\langle y_k - y_l|h_n\rangle| + 2C\|z - h_n\|.\end{aligned}$$

Letting $k, l \to \infty$ in this estimate then shows

$$\limsup_{k,l\to\infty} |\langle y_k|z\rangle - \langle y_l|z\rangle| \le 2C\|z - h_n\|.$$

Since we may choose $n \in \mathbb{N}$ such that $\|z - h_n\|$ is as small as we please, we may conclude that $\limsup_{k,l\to\infty} |\langle y_k|z\rangle - \langle y_l|z\rangle|$, i.e. $\varphi(z) := \lim_{k\to\infty} \langle y_k|z\rangle$ exists.

The function, $\bar{\varphi}(z) = \lim_{k\to\infty} \langle z|y_k\rangle$ is a bounded linear functional on $H$ because

$$|\bar{\varphi}(z)| = \liminf_{k\to\infty} |\langle z|y_k\rangle| \le C\|z\|.$$

Therefore by the Riesz Theorem 15.16, there exists $x \in H_0$ such that $\bar{\varphi}(z) = \langle z|x\rangle$ for all $z \in H_0$. Thus, for this $x \in H_0$ we have shown

$$\lim_{k\to\infty} \langle y_k|z\rangle = \langle x|z\rangle \text{ for all } z \in H_0. \tag{15.12}$$

To finish the proof we need only observe that Eq. (15.12) is valid for all $z \in H$. Indeed if $z \in H$, then $z = z_0 + z_1$ where $z_0 = P_{H_0}z \in H_0$ and $z_1 = z - P_{H_0}z \in H_0^\perp$. Since $y_k, x \in H_0$, we have

$$\lim_{k\to\infty} \langle y_k|z\rangle = \lim_{k\to\infty} \langle y_k|z_0\rangle = \langle x|z_0\rangle = \langle x|z\rangle \text{ for all } z \in H.$$

■

Since unbounded subsets of $H$ are clearly not sequentially weakly compact, Theorem 15.20 states that a set is sequentially precompact in $H$ iff it is bounded. Let us now use Theorem 15.20 to identify the sequentially compact subsets of $L^p(\Omega, \mathcal{B}, P)$ for all $1 \le p < \infty$. We begin with the case $p = 1$.

**Theorem 15.21.** *If $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$, there exists a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ and $X \in L^1(\Omega, \mathcal{B}, P)$ such that*

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in \mathcal{B}_b. \tag{15.13}$$

**Proof.** For each $m \in \mathbb{N}$ let $X_n^m := X_n 1_{|X_n| \le m}$. The truncated sequence $\{X_n^m\}_{n=1}^\infty$ is a bounded subset of the Hilbert space, $L^2(\Omega, \mathcal{B}, P)$, for all $m \in \mathbb{N}$. Therefore by Theorem 15.20, $\{X_n^m\}_{n=1}^\infty$ has a weakly convergent sub-sequence for all $m \in \mathbb{N}$. By Cantor's diagonalization argument, we can find $Y_k^m := X_{n_k}^m$ and $X^m \in L^2(\Omega, \mathcal{B}, P)$ such that $Y_k^m \xrightarrow{w} X^m$ as $m \to \infty$ and in particular

$$\lim_{k\to\infty} \mathbb{E}[Y_k^m h] = \mathbb{E}[X^m h] \text{ for all } h \in \mathcal{B}_b.$$

Our next goal is to show $X^m \to X$ in $L^1(\Omega, \mathcal{B}, P)$. To this end, for $m < M$ and $h \in \mathcal{B}_b$ we have

$$\begin{aligned}\left|\mathbb{E}\left[\left(X^M - X^m\right)h\right]\right| &= \lim_{k\to\infty}\left|\mathbb{E}\left[\left(Y_k^M - Y_k^m\right)h\right]\right| \le \liminf_{k\to\infty}\mathbb{E}\left[\left|Y_k^M - Y_k^m\right||h|\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : M \ge |Y_k| > m\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right].\end{aligned}$$

Taking $h = \overline{\operatorname{sgn}(X^M - X^m)}$ in this inequality shows

$$\mathbb{E}\left[\left|X^M - X^m\right|\right] \le \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right]$$

with the right member of this inequality going to zero as $m, M \to \infty$ with $M \ge m$ by the assumed uniform integrability of the $\{X_n\}$. Therefore there exists $X \in L^1(\Omega, \mathcal{B}, P)$ such that $\lim_{m\to\infty}\mathbb{E}|X - X^m| = 0$.

We are now ready to verify Eq. (15.13) is valid. For $h \in \mathcal{B}_b$,

$$\begin{aligned}|\mathbb{E}[(X - Y_k)h]| &\le |\mathbb{E}[(X^m - Y_k^m)h]| + |\mathbb{E}[(X - X^m)h]| + |\mathbb{E}[(Y_k - Y_k^m)h]| \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot (\mathbb{E}[|X - X^m|] + \mathbb{E}[|Y_k| : |Y_k| > m]) \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot \left(\mathbb{E}[|X - X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right).\end{aligned}$$

Passing to the limit as $k \to \infty$ in the above inequality shows

$$\limsup_{k\to\infty}|\mathbb{E}[(X - Y_k)h]| \le \|h\|_\infty \cdot \left(\mathbb{E}[|X - X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right).$$

Since $X^m \to X$ in $L^1$ and $\sup_l \mathbb{E}[|Y_l| : |Y_l| > m] \to 0$ by uniform integrability, it follows that, $\limsup_{k\to\infty}|\mathbb{E}[(X - Y_k)h]| = 0$. ■

*Example 15.22.* Let $(\Omega, \mathcal{B}, P) = \left((0,1), \mathcal{B}_{(0,1)}, m\right)$ where $m$ is Lebesgue measure and let $X_n(\omega) = 2^n 1_{0<\omega<2^{-n}}$. Then $\mathbb{E}X_n = 1$ for all $n$ and hence $\{X_n\}_{n=1}^\infty$ is bounded in $L^1(\Omega, \mathcal{B}, P)$ (but is not uniformly integrable). Suppose for sake of contradiction that there existed $X \in L^1(\Omega, \mathcal{B}, P)$ and subsequence, $Y_k := X_{n_k}$ such that $Y_k \xrightarrow{w} X$. Then for $h \in \mathcal{B}_b$ and any $\varepsilon > 0$ we would have

$$\mathbb{E}\left[Xh1_{(\varepsilon,1)}\right] = \lim_{k\to\infty}\mathbb{E}\left[Y_k h 1_{(\varepsilon,1)}\right] = 0.$$

Then by DCT it would follow that $\mathbb{E}[Xh] = 0$ for all $h \in \mathcal{B}_b$ and hence that $X \equiv 0$. On the other hand we would also have

$$0 = \mathbb{E}[X \cdot 1] = \lim_{k\to\infty}\mathbb{E}[Y_k \cdot 1] = 1$$

and we have reached the desired contradiction. Hence we must conclude that bounded subset of $L^1(\Omega, \mathcal{B}, P)$ need not be weakly compact and thus we can not drop the uniform integrability assumption made in Theorem 15.21.

When $1 < p < \infty$, the situation is simpler.

**Theorem 15.23.** *Let $p \in (1, \infty)$ and $q = p(p-1)^{-1} \in (1, \infty)$ be its conjugate exponent. If $\{X_n\}_{n=1}^\infty$ is a bounded sequence in $L^p(\Omega, \mathcal{B}, P)$, there exists $X \in L^p(\Omega, \mathcal{B}, P)$ and a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ such that*

$$\lim_{k\to\infty}\mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in L^q(\Omega, \mathcal{B}, P). \tag{15.14}$$

**Proof.** Let $C := \sup_{n\in\mathbb{N}}\|X_n\|_p < \infty$ and recall that Lemma 14.54 guarantees that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$. Therefore by Theorem 15.21, there exists $X \in L^1(\Omega, \mathcal{B}, P)$ and a subsequence, $Y_k := X_{n_k}$, such that Eq. (15.13) holds. We will complete the proof by showing; a) $X \in L^p(\Omega, \mathcal{B}, P)$ and b) and Eq. (15.14) is valid.

a) For $h \in \mathcal{B}_b$ we have

$$|\mathbb{E}[Xh]| \le \liminf_{k\to\infty}\mathbb{E}[|Y_k h|] \le \liminf_{k\to\infty}\|Y_k\|_p \cdot \|h\|_q \le C\|h\|_q.$$

For $M < \infty$, taking $h = \overline{\operatorname{sgn}(X)}\,|X|^{p-1} 1_{|X|\le M}$ in the previous inequality shows

$$\begin{aligned}\mathbb{E}\left[|X|^p 1_{|X|\le M}\right] &\le C\left\|\overline{\operatorname{sgn}(X)}\,|X|^{p-1}1_{|X|\le M}\right\|_q \\ &= C\left(\mathbb{E}\left[|X|^{(p-1)q}1_{|X|\le M}\right]\right)^{1/q} \le C\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/q}\end{aligned}$$

from which it follows that

$$\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/p} \le \left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1-1/q} \le C.$$

Using the monotone convergence theorem, we may let $M \to \infty$ in this equation to find $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p} \le C < \infty$.

b) Now that we know $X \in L^p(\Omega, \mathcal{B}, P)$, in make sense to consider $\mathbb{E}[(X - Y_k)h]$ for all $h \in L^p(\Omega, \mathcal{B}, P)$. For $M < \infty$, let $h^M := h1_{|h|\le M}$, then

$$\begin{aligned}|\mathbb{E}[(X - Y_k)h]| &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + \left|\mathbb{E}\left[(X - Y_k)h1_{|h|>M}\right]\right| \\ &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + \|X - Y_k\|_p\left\|h1_{|h|>M}\right\|_q \\ &\le \left|\mathbb{E}\left[(X - Y_k)h^M\right]\right| + 2C\left\|h1_{|h|>M}\right\|_q.\end{aligned}$$

Since $h^M \in \mathcal{B}_b$, we may pass to the limit $k \to \infty$ in the previous inequality to find,

$$\limsup_{k\to\infty} |\mathbb{E}[(X - Y_k) h]| \le 2C \left\| h 1_{|h|>M} \right\|_q.$$

This completes the proof, since $\left\| h 1_{|h|>M} \right\|_q \to 0$ as $M \to \infty$ by DCT. ■

## 15.2 Exercises

**Exercise 15.5.** Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, $H$. Let $M$ be the closure of $M_0 := \cup_{n=1}^\infty M_n$. Show $\lim_{n\to\infty} P_{M_n} x = P_M x$ for all $x \in H$. **Hint:** first prove this for $x \in M_0$ and then for $x \in M$. Also consider the case where $x \in M^\perp$.

**Exercise 15.6 (A "Martingale" Convergence Theorem).** Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, $H$, $P_n := P_{M_n}$, and $\{x_n\}_{n=1}^\infty$ is a sequence of elements from $H$ such that $x_n = P_n x_{n+1}$ for all $n \in \mathbb{N}$. Show;

1. $P_m x_n = x_m$ for all $1 \le m \le n < \infty$,
2. $(x_n - x_m) \perp M_m$ for all $n \ge m$,
3. $\|x_n\|$ is increasing as $n$ increases,
4. if $\sup_n \|x_n\| = \lim_{n\to\infty} \|x_n\| < \infty$, then $x := \lim_{n\to\infty} x_n$ exists in $M$ and that $x_n = P_n x$ for all $n \in \mathbb{N}$. (**Hint:** show $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence.)

*Remark 15.24.* Let $H = \ell^2 := L^2(\mathbb{N}, \text{counting measure})$,

$$M_n = \{(a(1), \dots, a(n), 0, 0, \dots) : a(i) \in \mathbb{C} \text{ for } 1 \le i \le n\},$$

and $x_n(i) = 1_{i \le n}$, then $x_m = P_m x_n$ for all $n \ge m$ while $\|x_n\|^2 = n \uparrow \infty$ as $n \to \infty$. Thus, we can not drop the assumption that $\sup_n \|x_n\| < \infty$ in Exercise 15.6.

The rest of this section may be safely skipped.

**Exercise 15.7.** *Suppose that $(X, \|\cdot\|)$ is a normed space such that parallelogram law, Eq. (15.2), holds for all $x, y \in X$, then there exists a unique inner product on $\langle\cdot|\cdot\rangle$ such that $\|x\| := \sqrt{\langle x|x\rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.

# Conditional Expectation

In this section let $(\Omega, \mathcal{B}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{B}$ be a sub – sigma algebra of $\mathcal{B}$. We will write $f \in \mathcal{G}_b$ iff $f : \Omega \to \mathbb{C}$ is bounded and $f$ is $(\mathcal{G}, \mathcal{B}_{\mathbb{C}})$ – measurable. If $A \in \mathcal{B}$ and $P(A) > 0$, we will let

$$\mathbb{E}[X|A] := \frac{\mathbb{E}[X : A]}{P(A)} \text{ and } P(B|A) := \mathbb{E}[1_B|A] := \frac{P(A \cap B)}{P(A)}$$

for all integrable random variables, $X$, and $B \in \mathcal{B}$. We will often use the factorization Lemma 8.40 in this section. Because of this let us repeat it here.

**Lemma 16.1.** *Suppose that $(\mathbb{Y}, \mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y), \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $H : \Omega \to \bar{\mathbb{R}}$, there is a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $h : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $H = h \circ Y$.*

**Proof.** First suppose that $H = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $h = 1_B$. More generally if $H = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $H = h \circ Y$ with $h := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $H$, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions $H_n$ converging to $H$. Let $h_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $H_n = h_n \circ Y$. Then it follows that

$$H = \lim_{n\to\infty} H_n = \limsup_{n\to\infty} H_n = \limsup_{n\to\infty} h_n \circ Y = h \circ Y$$

where $h := \limsup_{n\to\infty} h_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$. ■

**Definition 16.2 ($L^2$ – Conditional Expectation).** *Let $\mathbb{E}_{\mathcal{G}} : L^2(\Omega, \mathcal{B}, P) \to L^2(\Omega, \mathcal{G}, P)$ denote orthogonal projection of $L^2(\Omega, \mathcal{B}, P)$ onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. For $f \in L^2(\Omega, \mathcal{B}, P)$, we say that $\mathbb{E}_{\mathcal{G}} f \in L^2(\Omega, \mathcal{G}, P)$ is the* ***conditional expectation of** $f$ **given** $\mathcal{G}$*.

*Remark 16.3 (Basic Properties of $\mathbb{E}_{\mathcal{G}}$).* Some remarks on this definition are in order. Let $f \in \mathrm{L}^2(\Omega, \mathcal{B}, P)$.

1. We are identifying $L^2(\Omega, \mathcal{G}, P)$ with its image,
$$M := \left\{ [g]_{L^2(\Omega,\mathcal{B},P)} : g \in \mathrm{L}^2(\Omega, \mathcal{G}, P) \right\},$$
in $L^2(\Omega, \mathcal{B}, P)$. From Lemma 14.26 we know that $M$ is a closed subspace of $L^2(\Omega, \mathcal{B}, P)$.
2. Thus given $f \in \mathrm{L}^2(\Omega, \mathcal{B}, P)$, we may compute the orthogonal projection (Theorem 15.14) onto $M$, $P_M [f]_{L^2(\Omega,\mathcal{B},P)}$, of $[f]_{L^2(\Omega,\mathcal{B},P)}$. By definition of $M$, $P_M [f]_{L^2(\Omega,\mathcal{B},P)} = [F]_{L^2(\Omega,\mathcal{G},P)}$ for some $F \in \mathrm{L}^2(\Omega, \mathcal{G}, P)$ which is uniquely determined up to sets of measure 0. We will usually abuse notation and write $F = \mathbb{E}_{\mathcal{G}} f$ when this holds. [**Please note:** in general for any fixed $\omega \in \Omega$, $(\mathbb{E}_{\mathcal{G}} f)(\omega)$ is not well defined unless $P(\{\omega\}) > 0$. In is only then that changing $F = \tilde{F}$ a.s. would imply that $F(\omega) = \tilde{F}(\omega)$.]
3. By the orthogonal projection Theorem 15.14) we know that $F \in L^2(\Omega, \mathcal{G}, P)$ is $\mathbb{E}_{\mathcal{G}} f$ a.s. iff either of the following two conditions hold;
   a) $\|f - F\|_2 \le \|f - g\|_2$ for all $g \in \mathrm{L}^2(\Omega, \mathcal{G}, P)$ or
   b) $\mathbb{E}[fh] = \mathbb{E}[Fh]$ for all $h \in \mathrm{L}^2(\Omega, \mathcal{G}, P)$.
4. $L^1(P)$ – contractivity: $\mathbb{E}|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}|f|$ for all $f \in L^2(\Omega, \mathcal{B}, P)$. To prove this, let $F := \mathbb{E}_{\mathcal{G}} f$ (i.e. $F$ is a version of $\mathbb{E}_{\mathcal{G}} f$) and take $h := 1_{F\neq 0} \frac{\bar{F}}{F}$ in item 3b. above to find;
$$\mathbb{E}[|F|] = \mathbb{E}\left[F 1_{F\neq 0} \frac{\bar{F}}{F}\right] = \mathbb{E}[fh] \le \mathbb{E}[|fh|] \le \mathbb{E}|f|.$$
5. Moreover if $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$ then $L^2(\Omega, \mathcal{G}_0, P) \subset L^2(\Omega, \mathcal{G}_1, P) \subset L^2(\Omega, \mathcal{B}, P)$ and therefore by Theorem 15.14,
$$\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^2(\Omega, \mathcal{B}, P). \tag{16.1}$$

**Lemma 16.4.** *If $f \in \mathrm{L}^2(\Omega, \mathcal{B}, P)$ and $F \in \mathrm{L}^2(\Omega, \mathcal{G}, P)$ then the following are equivalent;*

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}, \tag{16.2}$$
$$\mathbb{E}[fh] = \mathbb{E}[Fh] \text{ for all } h \in \mathcal{G}_b, \text{ and} \tag{16.3}$$
$$F = \mathbb{E}_{\mathcal{G}} f \text{ a.s.}$$

**Proof.** If Eq. (16.2) holds, then by linearity we have $\mathbb{E}[fh] = \mathbb{E}[Fh]$ for all $\mathcal{G}$ – measurable simple functions, $h$ and hence by the approximation Theorem 8.39 and the DCT for all $h \in \mathcal{G}_b$. Therefore Eq. (16.2) implies Eq. (16.3). If Eq. (16.3) holds and $h \in L^2(\Omega, \mathcal{G}, P)$, we may use DCT to show

$$\mathbb{E}[fh] \overset{\text{DCT}}{=} \lim_{n\to\infty} \mathbb{E}\left[fh1_{|h|\le n}\right] \overset{(16.3)}{=} \lim_{n\to\infty} \mathbb{E}\left[Fh1_{|h|\le n}\right] \overset{\text{DCT}}{=} \mathbb{E}[Fh],$$

by Condition 3b of Remark 16.3 shows $F = \mathbb{E}_{\mathcal{G}} f$ a.s.. Taking $h = 1_A$ with $A \in \mathcal{G}$ Condition 3b. or Remark 16.3, we learn that Eq. (16.2) is satisfied as well. ■

**Theorem 16.5.** *Let $(\Omega, \mathcal{B}, P)$ and $\mathcal{G} \subset \mathcal{B}$ be as above and let $f, g \in L^1(\Omega, \mathcal{B}, P)$. The operator $\mathbb{E}_{\mathcal{G}} : L^2(\Omega, \mathcal{B}, P) \to L^2(\Omega, \mathcal{G}, P)$ extends uniquely to a linear contraction from $L^1(\Omega, \mathcal{B}, P)$ to $L^1(\Omega, \mathcal{G}, P)$. This extension enjoys the following properties;*

1. *If $f \ge 0$, $P$ – a.s. then $\mathbb{E}_{\mathcal{G}} f \ge 0$, $P$ – a.s.*
2. ***Monotonicity.** If $f \ge g$, $P$ – a.s. there $\mathbb{E}_{\mathcal{G}} f \ge \mathbb{E}_{\mathcal{G}} g$, $P$ – a.s.*
3. *$L^\infty$ – **contraction property.** $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, $P$ – a.s.*
4. ***Averaging Property.** If $f \in L^1(\Omega, \mathcal{B}, P)$ then $F = \mathbb{E}_{\mathcal{G}} f$ iff $F \in L^1(\Omega, \mathcal{G}, P)$ and*
$$\mathbb{E}(Fh) = \mathbb{E}(fh) \text{ for all } h \in \mathcal{G}_b. \tag{16.4}$$
5. ***Pull out property or product rule.** If $g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}}(gf) = g \cdot \mathbb{E}_{\mathcal{G}} f$, $P$ – a.s.*
6. ***Tower or smoothing property.** If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*

$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^1(\Omega, \mathcal{B}, P). \tag{16.5}$$

**Proof.** By the definition of orthogonal projection, $f \in L^2(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$,

$$\mathbb{E}(fh) = \mathbb{E}(f \cdot \mathbb{E}_{\mathcal{G}} h) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h). \tag{16.6}$$

Taking

$$h = \overline{\text{sgn}(\mathbb{E}_{\mathcal{G}} f)} := \frac{\overline{\mathbb{E}_{\mathcal{G}} f}}{\mathbb{E}_{\mathcal{G}} f} 1_{|\mathbb{E}_{\mathcal{G}} f| > 0} \tag{16.7}$$

in Eq. (16.6) shows

$$\mathbb{E}(|\mathbb{E}_{\mathcal{G}} f|) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \le \mathbb{E}(|fh|) \le \mathbb{E}(|f|). \tag{16.8}$$

It follows from this equation and the BLT (Theorem 15.19) that $\mathbb{E}_{\mathcal{G}}$ extends uniquely to a contraction form $L^1(\Omega, \mathcal{B}, P)$ to $L^1(\Omega, \mathcal{G}, P)$. Moreover, by a simple limiting argument, Eq. (16.6) remains valid for all $f \in L^1(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$. Indeed, (without reference to Theorem 15.19) if $f_n := f 1_{|f| \le n} \in L^2(\Omega, \mathcal{B}, P)$, then $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f_n - \mathbb{E}_{\mathcal{G}} f_m|] = \mathbb{E}[|\mathbb{E}_{\mathcal{G}}(f_n - f_m)|] \le \mathbb{E}[|f_n - f_m|] \to 0 \text{ as } m, n \to \infty.$$

By the completeness of $L^1(\Omega, \mathcal{G}, P)$, $F := L^1(\Omega, \mathcal{G}, P)\text{-}\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n$ exists. Moreover the function $F$ satisfies,

$$\mathbb{E}(F \cdot h) = \mathbb{E}(\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n \cdot h) = \lim_{n\to\infty} \mathbb{E}(f_n \cdot h) = \mathbb{E}(f \cdot h) \tag{16.9}$$

for all $h \in \mathcal{G}_b$ and by Proposition 9.22 there is at most one, $F \in L^1(\Omega, \mathcal{G}, P)$, which satisfies Eq. (16.9). We will again denote $F$ by $\mathbb{E}_{\mathcal{G}} f$. This proves the existence and uniqueness of $F$ satisfying the defining relation in Eq. (16.4) of item 4. The same argument used in Eq. (16.8) again shows $\mathbb{E}|F| \le \mathbb{E}|f|$ and therefore that $\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P)$ is a contraction.

Items 1 and 2. If $f \in L^1(\Omega, \mathcal{B}, P)$ with $f \ge 0$, then

$$\mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \ge 0 \; \forall \; h \in \mathcal{G}_b \text{ with } h \ge 0. \tag{16.10}$$

An application of Lemma 9.23 then shows that $\mathbb{E}_{\mathcal{G}} f \ge 0$ a.s.[1] The proof of item 2. follows by applying item 1. with $f$ replaced by $f - g \ge 0$.

Item 3. If $f$ is real, $\pm f \le |f|$ and so by Item 2., $\pm\mathbb{E}_{\mathcal{G}} f \le \mathbb{E}_{\mathcal{G}} |f|$, i.e. $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, $P$ – a.s. For complex $f$, let $h \ge 0$ be a bounded and $\mathcal{G}$ – measurable function. Then

$$\begin{aligned}\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f| h] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{G}} f \cdot \overline{\text{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right] = \mathbb{E}\left[f \cdot \overline{\text{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right] \\ &\le \mathbb{E}[|f| h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} |f| \cdot h].\end{aligned}$$

Since $h \ge 0$ is an arbitrary $\mathcal{G}$ – measurable function, it follows, by Lemma 9.23, that $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, $P$ – a.s. Recall the item 4. has already been proved.

Item 5. If $h, g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{B}, P)$, then

$$\mathbb{E}[(g\mathbb{E}_{\mathcal{G}} f) h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} f \cdot hg] = \mathbb{E}[f \cdot hg] = \mathbb{E}[gf \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}}(gf) \cdot h].$$

Thus $\mathbb{E}_{\mathcal{G}}(gf) = g \cdot \mathbb{E}_{\mathcal{G}} f$, $P$ – a.s.

Item 6., by the item 5. of the projection Theorem 15.14, Eq. (16.5) holds on $L^2(\Omega, \mathcal{B}, P)$. By continuity of conditional expectation on $L^1(\Omega, \mathcal{B}, P)$ and the density of $L^1$ probability spaces in $L^2$ – probability spaces shows that Eq. (16.5) continues to hold on $L^1(\Omega, \mathcal{B}, P)$.

**Second Proof.** For $h \in (\mathcal{G}_0)_b$, we have

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_0} f \cdot h]$$

which shows $\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_0} f$ a.s. By the product rule in item 5., it also follows that

$$\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f] = \mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f \cdot 1] = \mathbb{E}_{\mathcal{G}_0} f \cdot \mathbb{E}_{\mathcal{G}_1}[1] = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s.}$$

Notice that $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ need only be $\mathcal{G}_1$ – measurable. What the statement says there are representatives of $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ which is $\mathcal{G}_0$ – measurable and any such representative is also a representative of $\mathbb{E}_{\mathcal{G}_0} f$. ■

[1] This can also easily be proved directly here by taking $h = 1_{\mathbb{E}_{\mathcal{G}} f < 0}$ in Eq. (16.10).

*Remark 16.6.* There is another standard construction of $\mathbb{E}_{\mathcal{G}} f$ based on the characterization in Eq. (16.4) and the Radon Nikodym Theorem 17.8 below. It goes as follows, for $0 \le f \in L^1(P)$, let $Q$ be the measure defined by $dQ := f dP$. Then $Q|_{\mathcal{G}} \ll P|_{\mathcal{G}}$ and hence there exists $0 \le g \in L^1(\Omega, \mathcal{G}, P)$ such that $dQ|_{\mathcal{G}} = g dP|_{\mathcal{G}}$. This then implies that

$$\int_A f dP = Q(A) = \int_A g dP \text{ for all } A \in \mathcal{G},$$

i.e. $g = \mathbb{E}_{\mathcal{G}} f$. For general real valued, $f \in L^1(P)$, define $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} f_+ - \mathbb{E}_{\mathcal{G}} f_-$ and then for complex $f \in L^1(P)$ let $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} \operatorname{Re} f + i \mathbb{E}_{\mathcal{G}} \operatorname{Im} f$.

**Notation 16.7** *In the future, we will often write $\mathbb{E}_{\mathcal{G}} f$ as $\mathbb{E}[f|\mathcal{G}]$. Moreover, if $(\mathbb{X}, \mathcal{M})$ is a measurable space and $X : \Omega \to \mathbb{X}$ is a measurable map. We will often simply denote $\mathbb{E}[f|\sigma(X)]$ simply by $\mathbb{E}[f|X]$. We will further let $P(A|\mathcal{G}) := \mathbb{E}[1_A|\mathcal{G}]$ be the **conditional probability of** $A$ **given** $\mathcal{G}$, and $P(A|X) := P(A|\sigma(X))$ be **conditional probability of** $A$ **given** $X$.*

**Exercise 16.1.** Suppose $f \in L^1(\Omega, \mathcal{B}, P)$ and $f > 0$ a.s. Show $\mathbb{E}[f|\mathcal{G}] > 0$ a.s. (i.e. show $g > 0$ a.s. for any version, $g$, of $\mathbb{E}[f|\mathcal{G}]$.) Use this result to conclude if $f \in (a, b)$ a.s. for some $a, b$ such that $-\infty \le a < b \le \infty$, then $\mathbb{E}[f|\mathcal{G}] \in (a, b)$ a.s. More precisely you are to show that any version, $g$, of $\mathbb{E}[f|\mathcal{G}]$ satisfies, $g \in (a, b)$ a.s.

## 16.1 Examples

*Example 16.8.* Suppose $\mathcal{G}$ is the trivial $\sigma$ – algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$. In this case $\mathbb{E}_{\mathcal{G}} f = \mathbb{E} f$ a.s.

*Example 16.9.* On the opposite extreme, if $\mathcal{G} = \mathcal{B}$, then $\mathbb{E}_{\mathcal{G}} f = f$ a.s.

**Exercise 16.2 (Exercise 5.21 revisited.).** Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\mathcal{P} := \{A_i\}_{i=1}^\infty \subset \mathcal{B}$ is a partition of $\Omega$. (Recall this means $\Omega = \sum_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the $\sigma$ – algebra generated by $\mathcal{P}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i \in \Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.
2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}$ – measurable iff $g = \sum_{i=1}^\infty \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{B}, P)$, let $\mathbb{E}[f|A_i] := \mathbb{E}[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $\mathbb{E}[f|A_i] = 0$ otherwise. Show

$$\mathbb{E}_{\mathcal{G}} f = \sum_{i=1}^\infty \mathbb{E}[f|A_i] 1_{A_i} \text{ a.s.} \tag{16.11}$$

*Example 16.10.* If $S$ is a countable or finite set equipped with the $\sigma$ – algebra, $2^S$, and $X : \Omega \to S$ is a measurable map. Then

$$\mathbb{E}[Z|X] = \sum_{s \in S} \mathbb{E}[Z|X = s] 1_{X=s} \text{ a.s.}$$

where by convention we set $\mathbb{E}[Z|X = s] = 0$ if $P(X = s) = 1$. This is an immediate consequence of Exercise 16.2 with $\mathcal{G} = \sigma(X)$ which is generated by the partition, $\{X = s\}$ for $s \in S$. Thus if we define $F(s) := \mathbb{E}[Z|X = s]$, we will have $\mathbb{E}[Z|X] = F(X)$ a.s.

**Lemma 16.11.** *Suppose $(\mathbb{X}, \mathcal{M})$ is a measurable space, $X : \Omega \to \mathbb{X}$ is a measurable function, and $\mathcal{G}$ is a sub-$\sigma$-algebra of $\mathcal{B}$. If $X$ is independent of $\mathcal{G}$ and $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)]$ a.s.. Conversely if $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)]$ a.s. for all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, then $X$ is independent of $\mathcal{G}$.*

**Proof.** Suppose that $X$ is independent of $\mathcal{G}$, $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(\Omega, \mathcal{B}, P)$, $\mu := \mathbb{E}[f(X)]$, and $A \in \mathcal{G}$. Then, by independence,

$$\mathbb{E}[f(X) : A] = \mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A] = \mathbb{E}[\mu 1_A] = \mathbb{E}[\mu : A].$$

Therefore $\mathbb{E}_{\mathcal{G}}[f(X)] = \mu = \mathbb{E}[f(X)]$ a.s.

Conversely if $\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}[f(X)] = \mu$ and $A \in \mathcal{G}$, then

$$\mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X) : A] = \mathbb{E}[\mu : A] = \mu \mathbb{E}[1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A].$$

Since this last equation is assumed to hold true for all $A \in \mathcal{G}$ and all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, $X$ is independent of $\mathcal{G}$. ■

The following remark is often useful in computing conditional expectations. The following Exercise should help you gain some more intuition about conditional expectations.

*Remark 16.12 (**Note well.**).* According to Lemma 16.1, $\mathbb{E}(f|X) = \tilde{f}(X)$ a.s. for some measurable function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$. So computing $\mathbb{E}(f|X) = \tilde{f}(X)$ is equivalent to finding a function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f \cdot h(X)] = \mathbb{E}\left[\tilde{f}(X) h(X)\right] \tag{16.12}$$

for all bounded and measurable functions, $h : \mathbb{X} \to \mathbb{R}$. "The" function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, is often denoted by writing $\tilde{f}(x) = \mathbb{E}(f|X = x)$. If $P(X = x) > 0$, then $\mathbb{E}(f|X = x) = \mathbb{E}(f : X = x) / P(X = x)$ consistent with our previous definitions – compare with Example 16.10. If $P(X = x) = 0$, $\mathbb{E}(f|X = x)$ is not given a value but is **just** a convenient notational way to denote a function $\tilde{f} : \mathbb{X} \to \mathbb{R}$ such that Eq. (16.12) holds. (Roughly speaking, you should think that $\mathbb{E}(f|X = x) = \mathbb{E}[f \cdot \delta_x(X)] / \mathbb{E}[\delta_x(X)]$ where $\delta_x$ is the "Dirac delta function" at $x$. If this last comment is confusing to you, please ignore it!)

*Example 16.13.* Suppose that $X$ is a random variable, $t \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(P)$. We wish to compute $\mathbb{E}[f(X)|X \wedge t] = h(X \wedge t)$. So we are looking for a function, $h : (-\infty, t] \to \mathbb{R}$ such that

$$\mathbb{E}[f(X)u(X \wedge t)] = \mathbb{E}[h(X \wedge t)u(X \wedge t)] \tag{16.13}$$

for all bounded measurable functions, $u : (-\infty, t] \to \mathbb{R}$. Taking $u = 1_{\{t\}}$ in Eq. (16.13) implies,

$$\mathbb{E}[f(X) : X \geq t] = h(t) P(X \geq t)$$

and therefore we should take,

$$h(t) = \mathbb{E}[f(X)|X \geq t]$$

which by convention we set to be (say) zero if $P(X \geq t) = 0$. Now suppose that $u(t) = 0$, then Eq. (10.6) becomes,

$$\mathbb{E}[f(X)u(X) : X < t] = \mathbb{E}[h(X)u(X) : X < t]$$

from which it follows that $f(X)1_{X<t} = h(X)1_{X<t}$ a.s. Thus we can take

$$h(x) := \begin{cases} f(x) & \text{if } x < t \\ \mathbb{E}[f(X)|X \geq t] & \text{if } x = t \end{cases}$$

and we have shown,

$$\begin{aligned} \mathbb{E}[f(X)|X \wedge t] &= 1_{X<t} f(X) + 1_{X \geq t} \mathbb{E}[f(X)|X \geq t] \\ &= 1_{X \wedge t < t} f(X) + 1_{X \wedge t = t} \mathbb{E}[f(X)|X \geq t]. \end{aligned}$$

**Exercise 16.3.** Let $(\Omega, \mathcal{B}, P)$ be a probability space, $(\mathbb{X}, \mathcal{M})$ and $(\mathbb{Y}, \mathcal{N})$ be measurable spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ be measurable functions. Let $(\mathbb{X}, \mathcal{M}), (\mathbb{Y}, \mathcal{N})$ be measurable spaces, $(\Omega, \mathcal{F}, P)$ a probability space, and $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ be measurable functions. Further assume that $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$–algebra such that $X$ is $\mathcal{G}/\mathcal{M}$ – measurable and $Y$ is independent of $\mathcal{G}$. Then for any bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable function $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ we have

$$\mathbb{E}[f(X,Y)|\mathcal{G}] = h_f(X) = \mathbb{E}[f(x,Y)]|_{x=X} \text{ a.s.} \tag{16.14}$$

where if $\mu := \mathrm{Law}_P(Y)$,

$$h_f(x) := \mathbb{E}[f(x,Y)] = \int_{\mathbb{Y}} f(x,y)\, d\mu(y). \tag{16.15}$$

[This exercise is essentially a special case of Exercise 16.5 below.]

**Proposition 16.14.** *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $(\mathbb{X}, \mathcal{M}, \mu)$ and $(\mathbb{Y}, \mathcal{N}, \nu)$ are two $\sigma$ – finite measure spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions, and there exists $0 \leq \rho \in L^1(\Omega, \mathcal{B}, \mu \otimes \nu)$ such that $P((X,Y) \in U) = \int_U \rho(x,y)\, d\mu(x)\, d\nu(y)$ for all $U \in \mathcal{M} \otimes \mathcal{N}$. Let*

$$\bar{\rho}(x) := \int_{\mathbb{Y}} \rho(x,y)\, d\nu(y) \tag{16.16}$$

*and $x \in \mathbb{X}$ and $B \in \mathcal{N}$, let*

$$Q(x,B) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_B \rho(x,y)\, d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ \delta_{y_0}(B) & \text{if } \bar{\rho}(x) \in \{0,\infty\} \end{cases} \tag{16.17}$$

*where $y_0$ is some arbitrary but fixed point in $Y$. Then for any bounded (or non-negative) measurable function, $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, we have*

$$\mathbb{E}[f(X,Y)|X] = Q(X, f(X,\cdot)) =: \int_{\mathbb{Y}} f(X,y)\, Q(X,dy) = g(X) \text{ a.s.} \tag{16.18}$$

*where,*

$$g(x) := \int_{\mathbb{Y}} f(x,y)\, Q(x,dy) = Q(x, f(x,\cdot)).$$

*As usual we use the notation,*

$$Q(x,v) := \int_{\mathbb{Y}} v(y)\, Q(x,dy) = \begin{cases} \frac{1}{\bar{\rho}(x)} \int_{\mathbb{Y}} v(y) \rho(x,y)\, d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ \delta_{y_0}(v) = v(y_0) & \text{if } \bar{\rho}(x) \in \{0,\infty\}. \end{cases}$$

*for all bounded measurable functions, $v : \mathbb{Y} \to \mathbb{R}$.*

**Proof.** Our goal is to compute $\mathbb{E}[f(X,Y)|X]$. According to Remark 16.12, we are searching for a bounded measurable function, $g : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f(X,Y)h(X)] = \mathbb{E}[g(X)h(X)] \text{ for all } h \in \mathcal{M}_b. \tag{16.19}$$

(Throughout this argument we are going to repeatedly use the Tonelli - Fubini theorems.) We now explicitly write out both sides of Eq. (16.19);

$$\begin{aligned} \mathbb{E}[f(X,Y)h(X)] &= \int_{\mathbb{X} \times \mathbb{Y}} h(x) f(x,y) \rho(x,y)\, d\mu(x)\, d\nu(y) \\ &= \int_{\mathbb{X}} h(x) \left[ \int_{\mathbb{Y}} f(x,y) \rho(x,y)\, d\nu(y) \right] d\mu(x) \end{aligned} \tag{16.20}$$

$$\begin{aligned} \mathbb{E}[g(X)h(X)] &= \int_{\mathbb{X} \times \mathbb{Y}} h(x) g(x) \rho(x,y)\, d\mu(x)\, d\nu(y) \\ &= \int_{\mathbb{X}} h(x) g(x) \bar{\rho}(x)\, d\mu(x). \end{aligned} \tag{16.21}$$

Since the right sides of Eqs. (16.20) and (16.21) must be equal for all $h \in \mathcal{M}_b$, we must demand (see Lemma 9.23 and 9.24) that

$$\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y) = g(x)\,\bar{\rho}(x) \text{ for } \mu-\text{a.e. } x. \tag{16.22}$$

There are two possible problems in solving this equation for $g(x)$ at a particular point $x$; the first is when $\bar{\rho}(x) = 0$ and the second is when $\bar{\rho}(x) = \infty$. Since

$$\int_{\mathbb{X}} \bar{\rho}(x)\,d\mu(x) = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} \rho(x,y)\,d\nu(y)\right] d\mu(x) = 1,$$

we know that $\bar{\rho}(x) < \infty$ for $\mu-$ a.e. $x$ and therefore it does not matter how $g$ is defined on $\{\bar{\rho} = \infty\}$ as long as it is measurable. If

$$0 = \bar{\rho}(x) = \int_{\mathbb{Y}} \rho(x,y)\,d\nu(y),$$

then $\rho(x,y) = 0$ for $\nu-$ a.e. $y$ and therefore,

$$\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y) = 0. \tag{16.23}$$

Hence Eq. (16.22) will be valid no matter how we choose $g(x)$ for $x \in \{\bar{\rho} = 0\}$. So a valid solution of Eq. (16.22) is

$$g(x) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ f(x,y_0) = \delta_{y_0}(f(x,\cdot)) & \text{if } \bar{\rho}(x) \in \{0,\infty\} \end{cases}$$

and with this choice we will have $\mathbb{E}[f(X,Y)|X] = g(X) = Q(X,f)$ a.s. as desired. (Observe here that when $\bar{\rho}(x) < \infty$, $\rho(x,\cdot) \in L^1(\nu)$ and hence $\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y)$ is a well defined integral.) ■

It is comforting to observe that

$$P(X \in \{\bar{\rho} = 0\}) = P(\bar{\rho}(X) = 0) = \int_{\mathbb{X}} 1_{\bar{\rho}=0}\bar{\rho}\,d\mu = 0$$

and similarly

$$P(X \in \{\bar{\rho} = \infty\}) = \int_{\mathbb{X}} 1_{\bar{\rho}=\infty}\bar{\rho}\,d\mu = 0.$$

Thus it follows that $P(X \in \{x \in \mathbb{X} : \bar{\rho}(x) = 0 \text{ of } \infty\}) = 0$ while the set $\{x \in \mathbb{X} : \bar{\rho}(x) = 0 \text{ of } \infty\}$ is precisely where there is ambiguity in defining $g(x)$. Just for added security, let us check directly that $g(X) = \mathbb{E}[f(X,Y)|X]$ a.s. According to Eq. (16.21) we have

$$\begin{aligned}
\mathbb{E}[g(X)\,h(X)] &= \int_{\mathbb{X}} h(x)\,g(x)\,\bar{\rho}(x)\,d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\,g(x)\,\bar{\rho}(x)\,d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\,\bar{\rho}(x)\left(\frac{1}{\bar{\rho}(x)}\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y)\right) d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\left(\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y)\right) d\mu(x) \\
&= \int_{\mathbb{X}} h(x)\left(\int_{\mathbb{Y}} f(x,y)\,\rho(x,y)\,d\nu(y)\right) d\mu(x) \\
&= \mathbb{E}[f(X,Y)\,h(X)] \qquad \text{(by Eq. (16.20))},
\end{aligned}$$

wherein we have repeatedly used $\mu(\bar{\rho} = \infty) = 0$ and Eq. (16.23) holds when $\bar{\rho}(x) = 0$. This completes the verification that $g(X) = \mathbb{E}[f(X,Y)|X]$ a.s..

Proposition 16.14 shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. Proposition 16.14 also gives an example of regular conditional probabilities which we now define.

### 16.1.1 Conditioning Gaussian Random Vectors

**Theorem 16.15.** *Suppose that $Z = (X,Y)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$. Let $C = C_X := \mathbb{E}[XX^{\mathrm{tr}}]$ and then let*

$$W := Y - \mathbb{E}\left[YX^{\mathrm{tr}}\right] C^{-1}X \tag{16.24}$$

*where $C^{-1} = C|_{\mathrm{Ran}(C)}^{-1}P$ is as in Example 11.50 below. Then $(X,W)^{\mathrm{tr}}$ is again a Gaussian random vector and moreover $W$ is independent of $X$. The covariance matrix for $W$ is*

$$C_W := \mathbb{E}\left[WW^{\mathrm{tr}}\right] = \mathbb{E}\left[YY^{\mathrm{tr}}\right] - \mathbb{E}\left[YX^{\mathrm{tr}}\right] C^{-1}\mathbb{E}\left[XY^{\mathrm{tr}}\right]. \tag{16.25}$$

**Proof.** Let $A$ be any $k \times l$ matrix and let $W := Y - AX$. Since

$$\begin{pmatrix} X \\ W \end{pmatrix} = \begin{pmatrix} I & 0 \\ -A & I \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix},$$

according to Lemma 11.42 $(X,W)^{\mathrm{tr}}$ is still Gaussian. So according to Lemma 12.25, in order to make $W$ independent of $X$ it suffices to choose $A$ so that $W$ and $X$ are uncorrelated, i.e.

$$0 = \mathrm{Cov}\left(W_j, X_i\right) = \mathrm{Cov}\left(Y_j - \sum_k A_{jk} X_k, X_i\right)$$
$$= \mathbb{E}\left[Y_j X_i\right] - \sum_k A_{jk} \mathbb{E}\left(X_k X_i\right).$$

In matrix notation, we want to choose $A$ so that

$$\mathbb{E}\left[Y X^{\mathrm{tr}}\right] = A \mathbb{E}\left[X X^{\mathrm{tr}}\right]. \tag{16.26}$$

In the case $C := \mathbb{E}\left[X X^{\mathrm{tr}}\right]$ is non-degenerate, we see that $A := \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1}$ is the desired solution. In fact this works for general $C$ where $C^{-1}$ is defined in Example 11.50. To see this is correct, recall

$$v \cdot C v = v \cdot \mathbb{E}\left[X X^{\mathrm{tr}} v\right] = \mathbb{E}\left[(v \cdot X)^2\right]$$

from which it follows that

$$\mathrm{Nul}\left(C\right) = \left\{v \in \mathbb{R}^k : v \cdot X = 0\right\}.$$

Hence it follows that

$$\mathbb{E}\left[Y X^{\mathrm{tr}}\right] v = A \mathbb{E}\left[X X^{\mathrm{tr}}\right] v \text{ for all } v \in \mathrm{Nul}\left(C\right)$$

no matter how $A$ is chosen. On the other hand if $v \in \mathrm{Ran}\left(C\right) = \mathrm{Nul}\left(C\right)^{\perp}$,

$$A \mathbb{E}\left[X X^{\mathrm{tr}}\right] v = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1} C v = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] v$$

as desired.

To prove Eq. (16.25) let $B := \mathbb{E}\left[Y X^{\mathrm{tr}}\right]$ so that

$$W := Y - B C^{-1} X.$$

We then have

$$\begin{aligned}
\mathbb{E}\left[W W^{\mathrm{tr}}\right] &= \mathbb{E}\left[\left(Y - B C^{-1} X\right)\left(Y - B C^{-1} X\right)^{\mathrm{tr}}\right] \\
&= \mathbb{E}\left[\left(Y - B C^{-1} X\right)\left(Y^{\mathrm{tr}} - X^{\mathrm{tr}} C^{-1} B^{\mathrm{tr}}\right)\right] \\
&= \mathbb{E}\left[Y Y^{\mathrm{tr}} - Y X^{\mathrm{tr}} C^{-1} B^{\mathrm{tr}} - B C^{-1} X Y^{\mathrm{tr}} + B C^{-1} X X^{\mathrm{tr}} C^{-1} B^{\mathrm{tr}}\right] \\
&= \mathbb{E}\left[Y Y^{\mathrm{tr}}\right] - B C^{-1} B^{\mathrm{tr}} - B C^{-1} B^{\mathrm{tr}} + B C^{-1} C C^{-1} B^{\mathrm{tr}} \\
&= \mathbb{E}\left[Y Y^{\mathrm{tr}}\right] - B C^{-1} B^{\mathrm{tr}} \\
&= \mathbb{E}\left[Y Y^{\mathrm{tr}}\right] - \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1} \mathbb{E}\left[X Y^{\mathrm{tr}}\right].
\end{aligned}$$

■

**Corollary 16.16.** *Suppose that $Z = (X, Y)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$,*

$$A := \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1},$$
$$C_W := \mathbb{E}\left[Y Y^{\mathrm{tr}}\right] - \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1} \mathbb{E}\left[X Y^{\mathrm{tr}}\right],$$

*and suppose $W \stackrel{d}{=} N\left(C_W, 0\right)$. If $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}$ is a bounded measurable function, then*

$$\mathbb{E}\left[f\left(X, Y\right) | X\right] = \mathbb{E}\left[f\left(x, A x + W\right)\right]|_{x = X}.$$

*As an important special case, if $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then*

$$\mathbb{E}\left[e^{i(x \cdot X + y \cdot Y)} | X\right] = e^{i(x \cdot X + y \cdot A X)} e^{-\frac{1}{2} \mathrm{Var}(y \cdot W)} = e^{i(x \cdot X + y \cdot A X)} e^{-\frac{1}{2} C_W y \cdot y}. \tag{16.27}$$

**Proof.** Using the notation in Theorem 16.15,

$$\mathbb{E}\left[f\left(X, Y\right) | X\right] = \mathbb{E}\left[f\left(X, A X + W\right) | X\right]$$

where $W \stackrel{d}{=} N\left(C_W, 0\right)$ and $W$ is independent of $X$. The result now follows by an application of Exercise 16.6. Let us now specialize to the case where $f\left(X, Y\right) = e^{i(x \cdot X + y \cdot Y)}$ in which case

$$\begin{aligned}
\mathbb{E}\left[e^{i(x \cdot X + y \cdot Y)} | X\right] &= \mathbb{E}\left[e^{i\left(x \cdot x' + y \cdot \left(A x' + W\right)\right)} | X\right]|_{x' = X} = e^{i(x \cdot X + y \cdot A X)} \mathbb{E}\left[e^{i y \cdot W}\right] \\
&= e^{i(x \cdot X + y \cdot A X)} e^{-\frac{1}{2} \mathrm{Var}(y \cdot W)} = e^{i(x \cdot X + y \cdot A X)} e^{-\frac{1}{2} C_W y \cdot y}.
\end{aligned}$$

■

*Remark 16.17.* Let us go through the above calculations in the special case where $X$ is non-degenerate. Notice that $\left(Y X^{\mathrm{tr}}\right)_{ij} = Y_i X_j$ so that the the condition that $W := Y - A X$ is independent of $X$ may be written as

$$0 = \mathbb{E}\left[W X^{\mathrm{tr}}\right] = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] - A \mathbb{E}\left[X X^{\mathrm{tr}}\right]$$

and hence

$$A = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] \left(\mathbb{E}\left[X X^{\mathrm{tr}}\right]\right)^{-1} = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C_X^{-1}$$

and so $W$ defined in Eq. (16.24) is independent of $X$. Since $Z = (X, Y)^{\mathrm{tr}} = (X, A X + W)^{\mathrm{tr}}$ it then follows that

$$\mathbb{E}\left[f\left(Z\right) | X\right] = \mathbb{E}\left[f\left(\left(X, A X + W\right)^{\mathrm{tr}}\right) | X\right] = \mathbb{E}\left[f\left(\left(x, A x + W\right)^{\mathrm{tr}}\right)\right]|_{x = X}.$$

In particular we have,

$$\mathbb{E}\left[e^{i(x\cdot X+y\cdot Y)}|X\right]=\mathbb{E}\left[e^{i(x\cdot X+y\cdot AX+y\cdot W)}|X\right]=e^{i(x\cdot X+y\cdot AX)}\mathbb{E}\left[e^{iy\cdot W}|X\right]$$
$$=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}\operatorname{Var}(y\cdot W)}=e^{i(x\cdot X+y\cdot AX)}e^{-\frac{1}{2}C_W y\cdot y}.$$

Here we have

$$\begin{aligned}C_W =&\mathbb{E}\left[W\,W^{\rm tr}\right]=\mathbb{E}\left[(Y-AX)\,(Y-AX)^{\rm tr}\right]=\mathbb{E}\left[(Y-AX)\left[Y^{\rm tr}-X^{\rm tr}A^{\rm tr}\right]\right]\\ =&\mathbb{E}\left[YY^{\rm tr}-YX^{\rm tr}A^{\rm tr}\right]-A\mathbb{E}\left[XY^{\rm tr}-XX^{\rm tr}A^{\rm tr}\right]\\ =&C_Y-\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]-\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]\\ &+\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}C_XC_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]\\ =&C_Y-\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]-\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]\\ &+\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right]\\ =&C_Y-\mathbb{E}\left[YX^{\rm tr}\right]C_X^{-1}\mathbb{E}\left[XY^{\rm tr}\right].\end{aligned}$$

*Remark 16.18.* Suppose that $Z:\Omega\to\mathbb{R}^N$ is a mean zero Gaussian random vector, $C_Z=\mathbb{E}\left[ZZ^{\rm tr}\right]$ so that

$$\mathbb{E}\left[(Z\cdot a)\,(Z\cdot b)\right]=\mathbb{E}\left[a\cdot ZZ^{\rm tr}b\right]=a\cdot C_Zb\;\forall\;a,b\in\mathbb{R}^N.$$

Let us assume that $C_Z>0$, i.e. $Z$ is non-degenerate. Further suppose that $S:\mathbb{R}^N\to\mathbb{R}^k$ and $T:\mathbb{R}^N\to\mathbb{R}^l$ are linear maps, $X:=SZ$, and $Y=TZ$. We then have

$$\mathbb{E}\left[XY^{\rm tr}\right]=\mathbb{E}\left[SZZ^{\rm tr}T^{\rm tr}\right]=SC_ZT^{\rm tr}$$

from which it follows that $X$ is independent of $Y$ iff $SC_ZT^{\rm tr}=0$. The above result specialize to

$$C_X=\mathbb{E}\left[XX^{\rm tr}\right]=\mathbb{E}\left[SZZ^{\rm tr}S^{\rm tr}\right]=SC_ZS^{\rm tr}\text{ and}$$
$$C_Y=\mathbb{E}\left[YY^{\rm tr}\right]=\mathbb{E}\left[TZZ^{\rm tr}T^{\rm tr}\right]=TC_ZT^{\rm tr}.$$

Our next goal is to find $A:\mathbb{R}^k\to\mathbb{R}^N$ linear so that

$$W:=Z-AX=(I-AS)\,Z$$

is independent of $X=SZ$ which happens iff

$$0=(I-AS)\,C_ZS^{\rm tr}=C_ZS^{\rm tr}-AC_X\iff AC_X=C_ZS^{\rm tr}.$$

Let us observe that $C_Xx=0$ implies

$$0=C_Xx\cdot x=SC_ZS^{\rm tr}x\cdot x=C_ZS^{\rm tr}x\cdot S^{\rm tr}x$$

and since $C_Z>0$ we may conclude that $\operatorname{Nul}(C_X)=\operatorname{Nul}(S^{\rm tr})=\operatorname{Ran}(S)^\perp$. Hence if we assume that $S$ is surjective it will follows that $C_X$ is invertible and we must take

$$A=C_ZS^{\rm tr}C_X^{-1}.$$

In summary,

$$Z=AX+W$$

where $W\perp\!\!\!\perp X$ and

$$\begin{aligned}C_W&=(I-AS)\,C_Z\,(I-AS)^{\rm tr}=(I-AS)\,C_Z\left(I-S^{\rm tr}A^{\rm tr}\right)\\&=C_Z-ASC_Z-C_ZS^{\rm tr}A^{\rm tr}+ASC_ZS^{\rm tr}A^{\rm tr}\\&=C_Z-ASC_Z-C_ZS^{\rm tr}A^{\rm tr}+AC_XA^{\rm tr}\\&=C_Z-ASC_Z-C_ZS^{\rm tr}A^{\rm tr}+C_ZS^{\rm tr}A^{\rm tr}\\&=C_Z-ASC_Z=(I-AS)\,C_Z\\&=\left(I-C_ZS^{\rm tr}C_X^{-1}S\right)C_Z=C_Z-C_ZS^{\rm tr}C_X^{-1}SC_Z.\end{aligned}$$

From all of this it now follows that

$$\begin{aligned}\mathbb{E}\left[f\left(Z\right)|X\right]&=\mathbb{E}\left[f\left(AX+W\right)|X\right]=\mathbb{E}\left[f\left(Ax+W\right)\right]|_{x=X}\\&=\mathbb{E}\left[f\left(Ax+W\right)\right]|_{x=X}\end{aligned}$$

where $W\overset{d}{=}N\left(0,C_W\right)$. In particular,

$$\mathbb{E}\left[e^{i\lambda\cdot Z}|X\right]=\mathbb{E}\left[e^{i\lambda\cdot(AX+W)}|X\right]=e^{i\lambda\cdot AX}\cdot\mathbb{E}\left[e^{i\lambda\cdot W}\right]=\exp\left(-\frac{1}{2}C_W\lambda\cdot\lambda\right)e^{i\lambda\cdot AX}.$$

**Exercise 16.4.** Suppose now that $(X,Y,Z)^{\rm tr}$ is a mean zero Gaussian random vector with $X\in\mathbb{R}^k$, $Y\in\mathbb{R}^l$, and $Z\in\mathbb{R}^m$. Show for all $y\in\mathbb{R}^l$ and $z\in\mathbb{R}^m$ that

$$\begin{aligned}&\mathbb{E}\left[\exp\left(i\left(y\cdot Y+z\cdot Z\right)\right)|X\right]\\&\quad=\exp\left(-\operatorname{Cov}\left(y\cdot W_1,z\cdot W_2\right)\right)\cdot\mathbb{E}\left[\exp\left(iy\cdot Y\right)|X\right]\cdot\mathbb{E}\left[\exp\left(iz\cdot Z\right)|X\right].\end{aligned}\tag{16.28}$$

In performing these computations please use the following definitions,

$$C:=C_X:=\mathbb{E}\left[XX^{\rm tr}\right],\tag{16.29}$$
$$A:=\mathbb{E}\left[\begin{bmatrix}Y\\Z\end{bmatrix}X^{\rm tr}\right]C^{-1}=\begin{bmatrix}\mathbb{E}\left[YX^{\rm tr}\right]C^{-1}\\\mathbb{E}\left[ZX^{\rm tr}\right]C^{-1}\end{bmatrix}=:\begin{bmatrix}A_1\\A_2\end{bmatrix},\tag{16.30}$$

and

$$W:=\begin{bmatrix}W_1\\W_2\end{bmatrix}=\begin{bmatrix}Y\\Z\end{bmatrix}-AX=\begin{bmatrix}Y-A_1X\\Z-A_2X\end{bmatrix}.\tag{16.31}$$

### 16.1.2 Probability Kernels and Regular Conditional Distributions

**Definition 16.19.** *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces. A function, $Q:\mathbb{X}\times\mathcal{N}\to[0,1]$ is a* ***probability kernel on*** *$\mathbb{X}\times\mathbb{Y}$ if*

1. *$Q(x,\cdot):\mathcal{N}\to[0,1]$ is a probability measure on $(\mathbb{Y},\mathcal{N})$ for each $x\in\mathbb{X}$ and*
2. *$Q(\cdot,B):\mathbb{X}\to[0,1]$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $B\in\mathcal{N}$.*

If $Q$ is a probability kernel on $\mathbb{X}\times\mathbb{Y}$ and $f:\mathbb{Y}\to\mathbb{R}$ is a bounded measurable function or a positive measurable function, then $x\to Q(x,f):=\int_{\mathbb{Y}}f(y)\,Q(x,dy)$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable. This is clear for simple functions and then for general functions via simple limiting arguments.

**Definition 16.20.** *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces and $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ be measurable functions. A probability kernel, $Q$, on $\mathbb{X}\times\mathbb{Y}$ is said to be a* ***regular conditional distribution of*** *$Y$* ***given*** *$X$ iff $Q(X,B)$ is a version of $P(Y\in B|X)$ for each $B\in\mathcal{N}$. Equivalently, we should have $Q(X,f)=\mathbb{E}[f(Y)|X]$ a.s. for all $f\in\mathcal{N}_b$.*

The probability kernel, $Q$, defined in Eq. (16.17) is an example of a regular conditional distribution of $Y$ given $X$.

*Remark 16.21.* Unfortunately, regular conditional distributions do not always exists, see Doob [8, p. 624]. However, if we require $\mathbb{Y}$ to be a "standard Borel space," (i.e. $\mathbb{Y}$ is isomorphic to a Borel subset of $\mathbb{R}$), then a conditional distribution of $Y$ given $X$ will always exists. See Theorem 16.41 in the appendix to this chapter. Moreover, it is known that "reasonable" measure spaces are standard Borel spaces, see Section 11.11 above for more details. So in most instances of interest a regular conditional distribution of $Y$ given $X$ **will** exist.

**Exercise 16.5 (Jazzed up pull out property).** Let $(\Omega,\mathcal{B},P)$ be a probability space, $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces, $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ be measurable functions, and assume there exists a regular conditional distribution, $Q$, of $Y$ given $X$. Show:

1. For all bounded measurable functions, $f:(\mathbb{X}\times\mathbb{Y},\mathcal{M}\otimes\mathcal{N})\to\mathbb{R}$, the function $\mathbb{X}\ni x\to Q(x,f(x,\cdot))$ is measurable and
$$\mathbb{E}[f(X,Y)|X](\omega)=Q(X(\omega),f(X(\omega),\cdot)) \tag{16.32}$$
$$=\int_{\mathbb{Y}}Q(X(\omega),dy)\,f(X(\omega),y)\ \text{ for } P-\text{a.e. }\omega.. \tag{16.33}$$
**Hint:** let $\mathbb{H}$ denote the set of bounded measurable functions, $f$, on $\mathbb{X}\times\mathbb{Y}$ such that the two assertions are valid.
2. If $A\in\mathcal{M}\otimes\mathcal{N}$ and $\mu:=P\circ X^{-1}$ be the law of $X$, then
$$P((X,Y)\in A)=\int_{\mathbb{X}}Q(x,1_A(x,\cdot))\,d\mu(x)=\int_{\mathbb{X}}d\mu(x)\int_{\mathbb{Y}}1_A(x,y)\,Q(x,dy). \tag{16.34}$$

**Note:** If we take $\mathbb{X}=\Omega$, $X:\Omega\to\mathbb{X}=\Omega$ to be the identity map, and $\mathcal{M}$ to be a sub-sigma algebra of $\mathcal{B}$, then the results above reduce to;
$$Q(\omega,f(\omega,\cdot))=\mathbb{E}[f(\cdot,Y(\cdot))|\mathcal{M}](\omega)\ \text{ for } P-\text{a.e. }\omega$$
and
$$P(\{\omega:(\omega,Y(\omega))\in A\})=\int_{\Omega}dP(\omega)\int_{\mathbb{Y}}1_A(\omega,y)\,Q(\omega,dy).$$

**Exercise 16.6 (Compare with Exercise 16.3).** Let $(\Omega,\mathcal{B},P)$ be a probability space, $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces, $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ be measurable functions. Further assume that $X$ and $Y$ are independent. Find a regular conditional distribution of $Y$ given $X$ and prove
$$\mathbb{E}[f(X,Y)|X]=h_f(X)\ \text{ a.s. }\forall\text{ bounded measurable } f:\mathbb{X}\times\mathbb{Y}\to\mathbb{R},$$
where
$$h_f(x):=\mathbb{E}[f(x,Y)]\ \text{ for all } x\in X,$$
i.e.
$$\mathbb{E}[f(X,Y)|X]=\mathbb{E}[f(x,Y)]|_{x=X}\ \text{a.s.}$$

**Exercise 16.7.** Suppose $(\Omega,\mathcal{B},P)$ and $(\Omega',\mathcal{B}',P')$ are two probability spaces, $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ are measurable spaces, $X:\Omega\to\mathbb{X}$, $X':\Omega'\to\mathbb{X}$, $Y:\Omega\to\mathbb{Y}$,and $Y':\Omega\to\mathbb{Y}$ are measurable functions such that $P\circ(X,Y)^{-1}=P'\circ(X',Y')$, i.e. $(X,Y)\overset{d}{=}(X',Y')$. If $f:(\mathbb{X}\times\mathbb{Y},\mathcal{M}\otimes\mathcal{N})\to\mathbb{R}$ is a bounded measurable function and $\tilde{f}:(\mathbb{X},\mathcal{M})\to\mathbb{R}$ is a measurable function such that $\tilde{f}(X)=\mathbb{E}[f(X,Y)|X]$ $P$ - a.s. then
$$\mathbb{E}'[f(X',Y')|X']=\tilde{f}(X')\ \ P'\text{ a.s.}$$

**Exercise 16.8.** Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. integrable random variables such that $\mathbb{E}X_n=0$. Further set $S_0=0$ and for $n\in\mathbb{N}$ let $S_n:=X_1+\dots+X_n$ and $\mathcal{B}_{-n}:=\sigma(S_n,S_{n+1},S_{n+2},\dots)$. Show
$$\mathbb{E}[X_1|\mathcal{B}_{-n}]=\mathbb{E}[X_1|S_n,S_{n+1},S_{n+2},\dots]=\frac{S_n}{n}\ \text{a.s.}$$
[This problem will be used in Example 20.82 to give a proof of the strong law of large numbers.]

**Hint:** Use Exercise 16.7 to show
$$\mathbb{E}[X_j|\mathcal{B}_{-n}]=\mathbb{E}[X_1|\mathcal{B}_{-n}]\ \text{ a.s. for all } j\le n. \tag{16.35}$$

Let now suppose that $\mathcal{G}$ is a sub-$\sigma$-algebra of $\mathcal{B}$ and let $P_{\mathcal{G}} : \mathcal{B} \to L^1(\Omega, \mathcal{G}, P)$ be defined by, $P_{\mathcal{G}}(B) = P(B|\mathcal{G}) := \mathbb{E}_{\mathcal{G}} 1_B \in L^1(\Omega, \mathcal{B}, P)$ for all $B \in \mathcal{B}$. If $B = \sum_{n=1}^{\infty} B_n$ with $B_n \in \mathcal{B}$, then $1_B = \sum_{n=1}^{\infty} 1_{B_n}$ and this sum converges in $L^1(P)$ (in fact in all $L^p(P)$) by the DCT. Since $\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P)$ is a contraction and therefore continuous it follows that

$$P_{\mathcal{G}}(B) = \mathbb{E}_{\mathcal{G}} 1_B = \mathbb{E}_{\mathcal{G}} \sum_{n=1}^{\infty} 1_{B_n} = \sum_{n=1}^{\infty} \mathbb{E}_{\mathcal{G}} 1_{B_n} = \sum_{n=1}^{\infty} P_{\mathcal{G}}(B_n) \tag{16.36}$$

where all equalities are in $L^1(\Omega, \mathcal{G}, P)$. Now suppose that we have chosen a representative, $\bar{P}_{\mathcal{G}}(B) : \Omega \to [0,1]$, of $P_{\mathcal{G}}(B)$ for each $B \in \mathcal{B}$. From Eq. (16.36) it follows that

$$\bar{P}_{\mathcal{G}}(B)(\omega) = \sum_{n=1}^{\infty} \bar{P}_{\mathcal{G}}(B_n)(\omega) \text{ for } P\text{-a.e. } \omega. \tag{16.37}$$

However, **note well**, the exceptional set of $\omega$'s depends on the sets $B, B_n \in \mathcal{B}$. The goal of regular conditioning is to carefully choose the representative, $\bar{P}_{\mathcal{G}}(B) : \Omega \to [0,1]$, such that Eq. (16.37) holds for all $\omega \in \Omega$ and all $B, B_n \in \mathcal{B}$ with $B = \sum_{n=1}^{\infty} B_n$.

**Definition 16.22.** *If $\mathcal{G}$ is a sub-$\sigma$ – algebra of $\mathcal{B}$, a **regular conditional distribution given** $\mathcal{G}$ is a probability kernel on $Q : (\Omega, \mathcal{G}) \times (\Omega, \mathcal{B}) \to [0,1]$ such that*

$$Q(\cdot, B) = P(B|\mathcal{G})(\cdot) \text{ a.s. for every } B \in \mathcal{B}. \tag{16.38}$$

*This corresponds to the $Q$ in Definition 16.20 provided, $(\mathbb{X}, \mathcal{M}) = (\Omega, \mathcal{G})$, $(\mathbb{Y}, \mathcal{N}) = (\Omega, \mathcal{B})$, and $X(\omega) = Y(\omega) = \omega$ for all $\omega \in \Omega$.*

## 16.2 Additional Properties of Conditional Expectations

The next theorem is devoted to extending the notion of conditional expectations to all non-negative functions and to proving conditional versions of the MCT, DCT, and Fatou's lemma.

*Remark 16.23.* For the reader's convenience let us recall Lemma 9.24 asserts that if $F$ and $G$ are non-negative extended $\mathcal{G}$ – measurable functions on $\Omega$ such that $\mathbb{E}[F : A] \geq \mathbb{E}[G : A]$ for all $A \in \mathcal{G}$ then $F \geq G$ a.s. Here is a repeat of the proof in this case.

For each $n \in \mathbb{N}$ let $A_n := \{F < G : F \leq n\}$ in which case

$$\infty > n \geq \mathbb{E}[F 1_{A_n}] \geq \mathbb{E}[G 1_{A_n}]$$



from which it follows $G 1_{A_n}$ is integrable. Therefore, we may now conclude that $\mathbb{E}[F 1_{A_n} - G 1_{A_n}] \geq 0 \implies \mathbb{E}[G 1_{A_n} - F 1_{A_n}] \leq 0$. As $G 1_{A_n} - F 1_{A_n} \geq 0$ and $G - F > 0$ on $A_n$, it follows that $P(A_n) = 0$. Since $A_n \uparrow \{F < G\}$, we find $P(F < G) = 0$, i.e. $F \geq G$ a.s.

**Theorem 16.24 (Extending $\mathbb{E}_{\mathcal{G}}$).** *If $f : \Omega \to [0, \infty]$ is $\mathcal{B}$ – measurable, there is a $\mathcal{G}$ – measurable function, $F : \Omega \to [0, \infty]$, satisfying*

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}. \tag{16.39}$$

*By Lemma 9.24, the function $F$ is uniquely determined up to sets of measure zero and hence we denote any such version of $F$ by $\mathbb{E}_{\mathcal{G}} f$.*

1. *Properties 2., 5. (with $0 \leq g \in \mathcal{G}_b$), and 6. of Theorem 16.5 still hold for any $\mathcal{B}$ – measurable functions such that $0 \leq f \leq g$. Namely;*
   a) ***Order Preserving.*** *$\mathbb{E}_{\mathcal{G}} f \leq \mathbb{E}_{\mathcal{G}} g$ a.s. when $0 \leq f \leq g$,*
   b) ***Pull out Property.*** *$\mathbb{E}_{\mathcal{G}}[hf] = h \mathbb{E}_{\mathcal{G}}[f]$ a.s. for all $h \geq 0$ and $\mathcal{G}$ – measurable.*
   c) ***Tower or smoothing property.*** *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*
   $$\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s.}$$
2. ***Conditional Monotone Convergence (cMCT).*** *Suppose that, almost surely, $0 \leq f_n \leq f_{n+1}$ for all $n$, then $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} f_n]$ a.s.*
3. ***Conditional Fatou's Lemma (cFatou).*** *Suppose again that $0 \leq f_n \in L^1(\Omega, \mathcal{B}, P)$ a.s., then*
$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] \leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f_n] \text{ a.s.} \tag{16.40}$$

**Proof.** Since $f \wedge n \in L^1(\Omega, \mathcal{B}, P)$ and $f \wedge n$ is increasing, it follows that $F := \uparrow \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n]$ exists a.s. Moreover, by two applications of the standard MCT, we have for any $A \in \mathcal{G}$, that

$$\mathbb{E}[F : A] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}_{\mathcal{G}}[f \wedge n] : A] = \lim_{n\to\infty} \mathbb{E}[f \wedge n : A] = \lim_{n\to\infty} \mathbb{E}[f : A].$$

Thus Eq. (16.39) holds and this uniquely determines $F$ follows from Lemma 9.24.

Item 1. a) If $0 \leq f \leq g$, then

$$\mathbb{E}_{\mathcal{G}} f = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n] \leq \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[g \wedge n] = \mathbb{E}_{\mathcal{G}} g \text{ a.s.}$$

and so $\mathbb{E}_{\mathcal{G}}$ still preserves order. We will prove items 1b and 1c at the end of this proof.

Item 2. Suppose that, almost surely, $0 \le f_n \le f_{n+1}$ for all $n$, then $\mathbb{E}_{\mathcal{G}} f_n$ is a.s. increasing in $n$. Hence, again by two applications of the MCT, for any $A \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}\left[\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n : A\right] &= \lim_{n\to\infty} \mathbb{E}\left[\mathbb{E}_{\mathcal{G}} f_n : A\right] = \lim_{n\to\infty} \mathbb{E}\left[f_n : A\right] \\ &= \mathbb{E}\left[\lim_{n\to\infty} f_n : A\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{G}}\left[\lim_{n\to\infty} f_n\right] : A\right]\end{aligned}$$

which combined with Lemma 9.24 implies that $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}\left[\lim_{n\to\infty} f_n\right]$ a.s.

Item 3. For $0 \le f_n$, let $g_k := \inf_{n \ge k} f_n$. Then $g_k \le f_k$ for all $k$ and $g_k \uparrow \liminf_{n\to\infty} f_n$ and hence by cMCT and item 1.,

$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] = \lim_{k\to\infty} \mathbb{E}_{\mathcal{G}} g_k \le \liminf_{k\to\infty} \mathbb{E}_{\mathcal{G}} f_k \text{ a.s.}$$

Item 1. b) If $h \ge 0$ is a $\mathcal{G}$ – measurable function and $f \ge 0$, then by cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}\left[hf\right] &\overset{\text{cMCT}}{=} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}\left[(h \wedge n)(f \wedge n)\right] \\ &= \lim_{n\to\infty} (h \wedge n)\,\mathbb{E}_{\mathcal{G}}\left[(f \wedge n)\right] \overset{\text{cMCT}}{=} h\mathbb{E}_{\mathcal{G}} f \text{ a.s.}\end{aligned}$$

Item 1. c) Similarly by multiple uses of cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f &= \mathbb{E}_{\mathcal{G}_0} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}(f \wedge n) \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f &= \mathbb{E}_{\mathcal{G}_1} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0}\left[f \wedge n\right] \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f.\end{aligned}$$

■

**Theorem 16.25 (Conditional Dominated Convergence (cDCT)).** *If $f_n \overset{a.s}{\to} f$, and $|f_n| \le g \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ a.s.*

**Proof.** From Corollary 14.9 we know that $f_n \to f$ in $L^1(P)$ and therefore $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ in $L^1(P)$ as conditional expectation is a contraction on $L^1(P)$. So we need only prove the almost sure convergence. As usual it suffices to consider the real case.

Following the proof of the Dominated convergence theorem, we start with the fact that $0 \le g \pm f_n$ a.s. for all $n$. Hence by cFatou,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(g \pm f) &= \mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty}(g \pm f_n)\right] \\ &\le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(g \pm f_n) = \mathbb{E}_{\mathcal{G}} g + \begin{cases} \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } + \text{ case} \\ -\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } - \text{ case,}\end{cases}\end{aligned}$$

where the above equations hold a.s. Cancelling $\mathbb{E}_{\mathcal{G}} g$ from both sides of the equation then implies

$$\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \le \mathbb{E}_{\mathcal{G}} f \le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \text{ a.s.}$$

■

*Remark 16.26.* Suppose that $f_n \overset{P}{\to} f$, $|f_n| \le g_n \in L^1(\Omega, \mathcal{B}, P)$, $g_n \overset{P}{\to} g \in L^1(\Omega, \mathcal{B}, P)$ and $\mathbb{E}g_n \to \mathbb{E}g$. Then by the DCT in Corollary 14.9, we know that $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$. Since $\mathbb{E}_{\mathcal{G}}$ is a contraction, it follows that $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence $\mathbb{E}_{\mathcal{G}} f_n \overset{P}{\to} \mathbb{E}_{\mathcal{G}} f$.

The next result in Lemma 16.29 shows how to localize conditional expectations. In order to state and prove the lemma we need a little ground work first.

**Definition 16.27.** *Suppose that $\mathcal{F}$ and $\mathcal{G}$ are sub-$\sigma$-fields of $\mathcal{B}$ and $A \in \mathcal{B}$. We say that $\mathcal{F} = \mathcal{G}$ on $A$ iff $\mathcal{F}_A = \mathcal{G}_A$. Recall that $\mathcal{F}_A = \{B \cap A : B \in \mathcal{F}\}$.*

**Lemma 16.28.** *If $A \in \mathcal{F} \cap \mathcal{G}$, then $\mathcal{F}_A \cap \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A$ and $\mathcal{F}_A = \mathcal{G}_A$ implies*

$$\mathcal{F}_A = \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A. \tag{16.41}$$

**Proof.** If $A \in \mathcal{F}$ we have $B \in \mathcal{F}_A$ iff there exists $B' \in \mathcal{F}$ such that $B = A \cap B'$. As $A \in \mathcal{F}$ it follows that $B \in \mathcal{F}$ and therefore we have

$$\mathcal{F}_A = \{B \subset A : B \in \mathcal{F}\}.$$

Thus if $A \in \mathcal{F} \cap \mathcal{G}$ it follows that $\mathcal{F}_A = \{B \subset A : B \in \mathcal{F}\}$ and $\mathcal{G}_A = \{B \subset A : B \in \mathcal{G}\}$ and therefore

$$\mathcal{F}_A \cap \mathcal{G}_A = \{B \subset A : B \in \mathcal{F} \cap \mathcal{G}\} = [\mathcal{F} \cap \mathcal{G}]_A.$$

Equation (16.41) now clearly follows from this identity when $\mathcal{F}_A = \mathcal{G}_A$. ■

**Lemma 16.29 (Localizing Conditional Expectations).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\mathcal{F}$ and $\mathcal{G}$ be sub-sigma-fileds of $\mathcal{B}$, $X, Y \in L^1(\Omega, \mathcal{B}, P)$ or $X, Y : (\Omega, \mathcal{B}) \to [0, \infty]$ are measurable, and $A \in \mathcal{F} \cap \mathcal{G}$. If $\mathcal{F} = \mathcal{G}$ on $A$ and $X = Y$ a.s. on $A$, then*

$$\mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}} Y = \mathbb{E}_{\mathcal{G}} Y \text{ a.s. on } A. \tag{16.42}$$

**Proof.** It suffices to prove, $\mathbb{E}_{\mathcal{F}}X = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y$ a.s. on $A$ and this is equivalent to $1_A\mathbb{E}_{\mathcal{F}}X = 1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y$ a.s. Since

$$1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[1_A Y\right] = \mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[1_A X\right] = 1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[X\right] \text{ a.s.}$$

it suffices to show

$$\mathbb{E}_{\mathcal{F}}\left[1_A X\right] = 1_A\mathbb{E}_{\mathcal{F}}X = 1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[X\right] \text{ a.s.} \tag{16.43}$$

For $B \in \mathcal{F}$, we have by Lemma 16.28 that

$$A \cap B \in \mathcal{F}_A = \mathcal{G}_A = \left[\mathcal{F}\cap\mathcal{G}\right]_A \subset \mathcal{F}\cap\mathcal{G}$$

and therefore,

$$\begin{aligned}\mathbb{E}\left[1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[X\right]:B\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[X\right]:A\cap B\right] = \mathbb{E}\left[X:A\cap B\right] \\ &= \mathbb{E}\left[1_A X:B\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{F}}\left[1_A X\right]:B\right]\end{aligned}$$

from which Eq. (16.43) follows. ∎

*Example 16.30.* Let us use Lemma 16.29 to show $\mathbb{E}\left[f(X)|X\wedge t\right] = f(X) = f(X\wedge t)$ on $\{X<t\}$ – a fact we have already seen to be true in Example 16.13. Let us begin by observing that $\{X<t\} = \{X\wedge t<t\} \in \sigma(X)\cap\sigma(X\wedge t)$. Moreover, using $\sigma(X)_A = \sigma(X|_A)$ for all $A\in\mathcal{B}$,[2] we see that

$$\sigma(X)_{\{X<t\}} = \sigma\left(X|_{\{X<t\}}\right) = \sigma\left((X\wedge t)|_{\{X<t\}}\right) = \sigma(X\wedge t)_{\{X<t\}}.$$

Therefore it follows that

$$\mathbb{E}\left[f(X)|X\wedge t\right] = \mathbb{E}\left[f(X)|\sigma(X\wedge t)\right] = \mathbb{E}\left[f(X)|\sigma(X)\right] = f(X) \text{ a.s. on } \{X<t\}.$$

What goes wrong with the above argument if you replace $\{X<t\}$ by $\{X\le t\}$ everywhere? (Notice that the same argument shows; if $X=Y$ on $A\in\sigma(X)\cap\sigma(Y)$ then $\mathbb{E}\left[f(X)|Y\right] = f(Y) = f(X)$ a.s. on $A$.)

**Theorem 16.31 (Conditional Jensen's inequality).** *Let $(\Omega,\mathcal{B},P)$ be a probability space, $-\infty\le a<b\le\infty$, and $\varphi:(a,b)\to\mathbb{R}$ be a convex function. Assume $f\in L^1(\Omega,\mathcal{B},P;\mathbb{R})$ is a random variable satisfying, $f\in(a,b)$ a.s. and $\varphi(f)\in L^1(\Omega,\mathcal{B},P;\mathbb{R})$. Then $\varphi(\mathbb{E}_{\mathcal{G}}f)\in L^1(\Omega,\mathcal{G},P)$,*

$$\varphi(\mathbb{E}_{\mathcal{G}}f) \le \mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right] \text{ a.s.} \tag{16.44}$$

*and*

$$\mathbb{E}\left[\varphi(\mathbb{E}_{\mathcal{G}}f)\right] \le \mathbb{E}\left[\varphi(f)\right] \tag{16.45}$$

**Proof.** Let $\Lambda := \mathbb{Q}\cap(a,b)$ – a countable dense subset of $(a,b)$. By Theorem 14.63 (also see Lemma 14.60) and Figure 14.5 when $\varphi$ is $C^1$)

$$\varphi(y) \ge \varphi(x) + \varphi'_-(x)(y-x) \text{ for all for all } x,y\in(a,b), \tag{16.46}$$

where $\varphi'_-(x)$ is the left hand derivative of $\varphi$ at $x$. Taking $y=f$ and then taking conditional expectations imply,

$$\mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right] \ge \mathbb{E}_{\mathcal{G}}\left[\varphi(x)+\varphi'_-(x)(f-x)\right] = \varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x) \text{ a.s.} \tag{16.47}$$

Since this is true for all $x\in(a,b)$ (and hence all $x$ in the countable set, $\Lambda$) we may conclude that

$$\mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right] \ge \sup_{x\in\Lambda}\left[\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right] \text{ a.s.}$$

By Exercise 16.1, $\mathbb{E}_{\mathcal{G}}f\in(a,b)$, and hence it follows from Corollary 14.64 that

$$\sup_{x\in\Lambda}\left[\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right] = \varphi(\mathbb{E}_{\mathcal{G}}f) \text{ a.s.}$$

Combining the last two estimates proves Eq. (16.44).

From Eq. (16.44) and Eq. (16.46) with $y=\mathbb{E}_{\mathcal{G}}f$ and $x\in(a,b)$ fixed we find,

$$\varphi(x)+\varphi'_-(x)\left(\mathbb{E}_{\mathcal{G}}f-x\right) \le \varphi(\mathbb{E}_{\mathcal{G}}f) \le \mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right]. \tag{16.48}$$

Therefore

$$\left|\varphi(\mathbb{E}_{\mathcal{G}}f)\right| \le \left|\mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right]\right| \vee \left|\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right| \in L^1(\Omega,\mathcal{G},P) \tag{16.49}$$

which implies that $\varphi(\mathbb{E}_{\mathcal{G}}f)\in L^1(\Omega,\mathcal{G},P)$. Taking expectations of Eq. (16.44) is now allowed and immediately gives Eq. (16.45). ∎

*Remark 16.32 (On Theorem 16.31 and its proof.).* *

1. From Eq. (16.46),

$$\varphi(f) \ge \varphi(\mathbb{E}_{\mathcal{G}}f) + \varphi'_-(\mathbb{E}_{\mathcal{G}}f)(f-\mathbb{E}_{\mathcal{G}}f). \tag{16.50}$$

Therefore taking $\mathbb{E}_{\mathcal{G}}$ of this equation "implies" that

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}\left[\varphi(f)\right] &\ge \varphi(\mathbb{E}_{\mathcal{G}}f) + \mathbb{E}_{\mathcal{G}}\left[\varphi'_-(\mathbb{E}_{\mathcal{G}}f)(f-\mathbb{E}_{\mathcal{G}}f)\right] \\ &= \varphi(\mathbb{E}_{\mathcal{G}}f) + \varphi'_-(\mathbb{E}_{\mathcal{G}}f)\mathbb{E}_{\mathcal{G}}\left[(f-\mathbb{E}_{\mathcal{G}}f)\right] = \varphi(\mathbb{E}_{\mathcal{G}}f).\end{aligned} \tag{16.51}$$

The technical problem with this argument is the justification that $\mathbb{E}_{\mathcal{G}}\left[\varphi'_-(\mathbb{E}_{\mathcal{G}}f)(f-\mathbb{E}_{\mathcal{G}}f)\right] = \varphi'_-(\mathbb{E}_{\mathcal{G}}f)\mathbb{E}_{\mathcal{G}}\left[(f-\mathbb{E}_{\mathcal{G}}f)\right]$ since there is no reason for $\varphi'_-$ to be a bounded function. The proof we give in Theorem 16.31 circumvents this technical detail.

[2] Here is the verification that $\sigma(X)_A = \sigma(X|_A)$. Let $i_A : A\to\Omega$ be the inclusion map. Since $\sigma(X) = X^{-1}(\mathcal{B}_{\mathbb{R}})$ and $\sigma(X)_A = i_A^{-1}\sigma(X)$ it follows that

$$\begin{aligned}\sigma(X)_A &= i_A^{-1}\left(X^{-1}(\mathcal{B}_{\mathbb{R}})\right) = (X\circ i_A)^{-1}(\mathcal{B}_{\mathbb{R}}) \\ &= \sigma(X\circ i_A) = \sigma(X|_A).\end{aligned}$$

2. On the other hand let us now suppose that $\varphi$ is $C^1(\mathbb{R})$ is convex and for the moment that $|f| \le M < \infty$ a.s. Then $\mathbb{E}_\mathcal{G} f \in [-M, M]$ a.s. and hence $\varphi'_-(\mathbb{E}_\mathcal{G} f) = \varphi'(\mathbb{E}_\mathcal{G} f)$ is bounded and Eq. (16.51) is now valid. Moreover, taking $x = 0$ in Eq. (16.48) shows

$$\varphi(0) + \varphi'(0)\mathbb{E}_\mathcal{G} f \le \varphi(\mathbb{E}_\mathcal{G} f) \le \mathbb{E}_\mathcal{G}[\varphi(f)].$$

If $f$ is unbounded we may apply the above inequality with $f$ replaced by $f_M := f \cdot 1_{|f| \le M}$ in order to conclude,

$$\varphi(0) + \varphi'(0)\mathbb{E}_\mathcal{G} f_M \le \varphi(\mathbb{E}_\mathcal{G} f_M) \le \mathbb{E}_\mathcal{G}[\varphi(f_M)].$$

If we further assume that $\varphi(f_M) \ge 0$ is increasing as $M$ increase (for example this is the case if $\varphi(x) = |x|^p$ for some $p > 1$), then by passing to the limit $(M \uparrow \infty)$ along a nicely chosen subsequence it follows that

$$\varphi(0) + \varphi'(0)\mathbb{E}_\mathcal{G} f \le \varphi(\mathbb{E}_\mathcal{G} f) \le \mathbb{E}_\mathcal{G}[\varphi(f)]$$

where we used $\mathbb{E}_\mathcal{G}[\varphi(f_M)] \to \mathbb{E}_\mathcal{G}[\varphi(f)]$ by cMCT.

3. Since $\varphi(x) = x^p$ is the most important case we need later let us write out the argument for this particular case or more generally for $\varphi : [0, \infty) \to [0, \infty)$, $C^1$ – and convex such that $\varphi(0) = 0$ and $\varphi$ is increasing. Let $0 \le f < \infty$ be a measurable function. The convexity of $\varphi$ may be stated as,

$$\varphi(c) + \varphi'(c)(x - c) \le \varphi(x) \text{ for all } x, c \ge 0.$$

So letting $f_n := f 1_{f \le n}$, if $x = f_n$ and $c = \mathbb{E}_\mathcal{G} f_n$ in the previous equation, then

$$\varphi(\mathbb{E}_\mathcal{G} f_n) + \varphi'(\mathbb{E}_\mathcal{G} f_n)(f_n - \mathbb{E}_\mathcal{G} f_n) \le \varphi(f_n).$$

Under the given hypothesis, $\varphi'(\mathbb{E}_\mathcal{G} f_n)$ is bounded and $\mathcal{G}$ – measurable and therefore conditioning the previous equation on $\mathcal{G}$ implies,

$$\varphi(\mathbb{E}_\mathcal{G} f_n) = \mathbb{E}_\mathcal{G}[\varphi(\mathbb{E}_\mathcal{G} f_n) + \varphi'(\mathbb{E}_\mathcal{G} f_n)(f_n - \mathbb{E}_\mathcal{G} f_n)] \le \mathbb{E}_\mathcal{G}[\varphi(f_n)].$$

Using *cMCT* we may let $n \uparrow \infty$ in order to conclude

$$\varphi(\mathbb{E}_\mathcal{G} f) = \varphi\left(\lim_{n\to\infty} \mathbb{E}_\mathcal{G} f_n\right) = \lim_{n\to\infty} \varphi(\mathbb{E}_\mathcal{G} f_n) \le \lim_{n\to\infty} \mathbb{E}_\mathcal{G}[\varphi(f_n)] = \mathbb{E}_\mathcal{G}\left[\lim_{n\to\infty} \varphi(f_n)\right] = \mathbb{E}_\mathcal{G}[\varphi(f)].$$

**Corollary 16.33.** *The conditional expectation operator, $\mathbb{E}_\mathcal{G}$ maps $L^p(\Omega, \mathcal{B}, P)$ into $L^p(\Omega, \mathcal{B}, P)$ and the map remains a contraction for all $1 \le p \le \infty$.*

**Proof.** The case $p = \infty$ and $p = 1$ have already been covered in Theorem 16.5. So now suppose, $1 < p < \infty$, and apply Jensen's inequality with $\varphi(x) = |x|^p$ to find $|\mathbb{E}_\mathcal{G} f|^p \le \mathbb{E}_\mathcal{G} |f|^p$ a.s. Taking expectations of this inequality gives the desired result. ■

**Exercise 16.9 (Martingale Convergence Theorem for $p = 1$ and 2.").** Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\{\mathcal{B}_n\}_{n=1}^\infty$ be an increasing sequence of sub-$\sigma$-algebras of $\mathcal{B}$. Show;

1. The closure, $M$, of $\cup_{n=1}^\infty L^2(\Omega, \mathcal{B}_n, P)$ is $L^2(\Omega, \mathcal{B}_\infty, P)$ where $\mathcal{B}_\infty = \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. **Hint:** make use of Theorem 14.29.
2. For every $X \in L^2(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^2(P)$. **Hint:** see Exercise 15.5.
3. For every $X \in L^1(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^1(P)$. **Hint:** make use of item 2. by a truncation argument using the contractive properties of conditional expectations.

(Eventually we will show that $X_n = \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ a.s. as well.)

**Exercise 16.10 (Martingale Convergence Theorem for general $p$).** Let $1 \le p < \infty$, $(\Omega, \mathcal{B}, P)$ be a probability space, and $\{\mathcal{B}_n\}_{n=1}^\infty$ be an increasing sequence of sub-$\sigma$-algebras of $\mathcal{B}$. Show for all $X \in L^p(\Omega, \mathcal{B}, P)$, $X_n := \mathbb{E}[X|\mathcal{B}_n] \to \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^p(P)$. (**Hint:** show that $\{|\mathbb{E}[X|\mathcal{B}_n]|^p\}_{n=1}^\infty$ is uniformly integrable and $\mathbb{E}[X|\mathcal{B}_n] \xrightarrow{P} \mathbb{E}[X|\mathcal{B}_\infty]$ with the aid of item 3. of Exercise 16.9.)

**Theorem 16.34.** *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and assume that $\mathcal{B}$ is countably generated.*[3] *Then for all $1 \le p < \infty$, $L^p(\mu)^* \cong L^q(\mu)$ where $q = p/(p-1)$.*

**Proof.** The case where $1 \le p \le 2$ may be dealt with using Hilbert space theory and the fact that $L^2(\mu) \subset L^p(\mu)$ for $p \in [1, 2]$. So the difficult case is where $2 < p < \infty$. Moreover, the hard part is to show for all $\varphi \in L^p(\mu)^*$ there exists $g \in L^q(\mu)$ so that

$$\varphi(f) = \int_\Omega f g d\mu \text{ for all } f \in L^p(\mu).$$

Using the countably generated assumption on $\mathcal{B}$ we may find finite subalgebras, $\mathcal{B}_n$, of $\mathcal{B}$ so that $\mathcal{B}_n \subset \mathcal{B}_{n+1}$ and $\mathcal{B} = \sigma(\cup_n \mathcal{B}_n)$. Given $\varphi \in L^p(\mu)^*$ we let $\varphi_n \in \varphi|_{L^p(\Omega, \mathcal{B}_n, \mu)} \in L^p(\Omega, \mathcal{B}_n, \mu)^*$. As $L^p(\Omega, \mathcal{B}_n, \mu)$ is a finite dimensional subspace we may use **finite dimensional** Hilbert space theory to conclude that there exists

$$g_n \in L^2(\Omega, \mathcal{B}_n, \mu) = L^p(\Omega, \mathcal{B}_n, \mu) = L^q(\Omega, \mathcal{B}_n, \mu)$$

such that

[3] This theorem is true without this assumption or the finiteness assumption on $\mu$, but we do not prove this here.

$$\varphi(f) = \varphi_n(f) = \int_\Omega f g_n d\mu \text{ for all } f \in L^2(\Omega, \mathcal{B}_n, \mu) = L^p(\Omega, \mathcal{B}_n, \mu).$$

Now taking $f = \operatorname{sgn}(g_n)|g_n|^{q-1}$ in the above identity shows

$$\|g_n\|_q^q = \int_\Omega |g_n|^q d\mu = \varphi(f) \le \|\varphi\|_{L^p(\Omega,\mathcal{B},\mu)^*} \cdot \|f\|_p$$

where

$$\|f\|_p^p = \int_\Omega |g_n|^{(\frac{p}{p-1}-1)p} d\mu = \int_\Omega |g_n|^{\frac{p}{p-1}} d\mu = \|g_n\|_q^q.$$

Therefore it follows that

$$\|g_n\|_q^q \le \|\varphi\|_{L^p(\Omega,\mathcal{B},\mu)^*} \cdot \left(\|g_n\|_q^q\right)^{1/p} \implies \infty > \|\varphi\|_{L^p(\Omega,\mathcal{B},\mu)^*} \ge \left(\|g_n\|_q^q\right)^{1-\frac{1}{p}} = \|g_n\|_q.$$

Moreover it is easy to check using $\varphi(f) = \int_\Omega f g_n d\mu$ that $\{g_n\}_{n=1}^\infty$ is a $\{\mathcal{B}_n\}_{n=1}^\infty$-martingale. Therefore by the martingale convergence theorem, $g := L^q(\mu)\text{-}\lim_{n\to\infty} g_n$ exists and we have $g_n = \mathbb{E}[g|\mathcal{B}_n]$ for all $n$. That is $\varphi(f) = \int_\Omega f g d\mu$ for all $f \in L^p(\Omega, \mathcal{B}_n, \mu)$ for all $n$. Using the martingale convergence theorem again; if $f \in L^p(\mu)$, then

$$\begin{aligned}\varphi(f) &= \lim_{n\to\infty} \varphi(\mathbb{E}[f|\mathcal{B}_n]) = \lim_{n\to\infty} \int_\Omega \mathbb{E}[f|\mathcal{B}_n] g d\mu \\ &= \lim_{n\to\infty} \int_\Omega f \mathbb{E}[g|\mathcal{B}_n] d\mu = \lim_{n\to\infty} \int_\Omega f g d\mu.\end{aligned}$$

This completes the existence proof.

We could also use the multiplicative system theorem to finish off the proof since $\mathbb{M} := \cup_n L^p(\Omega, \mathcal{B}_n, \mu)$ is a multiplicative system of bounded function containing 1 and hence $\varphi(f) = \int_\Omega f g d\mu$ for all bounded measurable $f$. We then do a simple cutoff argument to show the relation holds for all $f \in L^p(\mu)$. ■

## 16.3 Conditional Independence

**Definition 16.35 (Conditional Independence).** *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\mathcal{B}_i \subset \mathcal{B}$ be a sub-sigma algebra of $\mathcal{B}$ for $i = 1, 2, 3$. We say that $\mathcal{B}_1$ is **independent of** $\mathcal{B}_3$ **conditioned** on $\mathcal{B}_2$ (written $\mathcal{B}_1 \overset{\mathcal{B}_2}{\perp\!\!\!\perp} \mathcal{B}_3$) provided,*

$$P(A \cap B|\mathcal{B}_2) = P(A|\mathcal{B}_2) \cdot P(B|\mathcal{B}_2) \text{ a.s.}$$

*for all $A \in \mathcal{B}_1$ and $B \in \mathcal{B}_3$. This can be equivalently stated as*

$$\mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \text{ a.s.}$$

*for all $f \in (\mathcal{B}_1)_b$ and $g \in (\mathcal{B}_3)_b$, where $\mathcal{B}_b$ denotes the **bounded** $B$ – **measurable functions.** If $X, Y, Z$ are measurable functions on $(\Omega, \mathcal{B})$, we say that $X$ is independent of $Z$ conditioned on $Y$ (written as $X \overset{Y}{\perp\!\!\!\perp} Z$) provided $\sigma(X) \overset{\sigma(Y)}{\perp\!\!\!\perp} \sigma(Z)$.*

*Example 16.36.* Let $X$ and $Y$ be two i.i.d. random variables such that $P(X = 1) = 1/2 = P(Y = 1)$ and $P(X = 2) = 1/2 = P(Y = 2)$. Then

$$\mathbb{E}[Y|X = Y] = \mathbb{E}[X|X = Y] = \frac{\frac{1}{4}(1+2)}{\frac{1}{4} + \frac{1}{4}} = \frac{3}{2}$$

and

$$\mathbb{E}[XY|X = Y] = \frac{\frac{1}{4}(1+4)}{\frac{1}{4} + \frac{1}{4}} = \frac{5}{4}.$$

Notice that

$$\mathbb{E}[XY|X = Y] = \frac{5}{4} \neq \frac{9}{4} = \mathbb{E}[Y|X = Y] \cdot \mathbb{E}[X|X = Y].$$

So independence does not necessarily imply conditional independence!

See Exercise 16.12 and Theorem 19.4 for a couple more examples involving conditional independence.

**Exercise 16.11.** Suppose $\mathbb{M}_i \subset (\mathcal{B}_i)_b$ for $i = 1$ and $i = 3$ are multiplicative systems such that $\mathcal{B}_i = \sigma(\mathbb{M}_i)$. Show $\mathcal{B}_1 \overset{\mathcal{B}_2}{\perp\!\!\!\perp} \mathcal{B}_3$ iff

$$\mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \text{ a.s. } \forall f \in \mathbb{M}_1 \text{ and } g \in \mathbb{M}_3. \tag{16.52}$$

**Hint:** Do this by two applications of the functional form of the multiplicative systems theorem, see Theorems 10.20 and 10.5 of Chapter 10. For the first application, fix an $f \in \mathbb{M}_1$ and let

$$\mathbb{H} := \{g \in (\mathcal{B}_3)_b : \mathbb{E}(f \cdot g|\mathcal{B}_2) = \mathbb{E}(f|\mathcal{B}_2) \cdot \mathbb{E}(g|\mathcal{B}_2) \text{ a.s.}\}.$$

(See the proof of Theorem 19.4 if you get stuck.)

**Exercise 16.12.** Suppose now that $(X, Y, Z)^{\mathrm{tr}}$ is a mean zero Gaussian random vector with $X \in \mathbb{R}^k$, $Y \in \mathbb{R}^l$, and $Z \in \mathbb{R}^m$ as in Exercise 16.4. Show $Y \overset{X}{\perp\!\!\!\perp} Z$ (see Definition 16.35) iff

$$\mathbb{E}\left[Y Z^{\mathrm{tr}}\right] = \mathbb{E}\left[Y X^{\mathrm{tr}}\right] C^{-1} \mathbb{E}\left[X Z^{\mathrm{tr}}\right].$$

where

$$C = C_X := \mathbb{E}\left[X X^{\mathrm{tr}}\right].$$

In solving this problem, please continue to use the notation setup in Exercise 16.4.

## 16.4 Construction of Regular Conditional Distributions*

**Lemma 16.37.** *Suppose that $h : \mathbb{Q} \to [0,1]$ is an increasing (i.e. non-decreasing) function and $H(t) := \inf\{h(s) : t < s \in \mathbb{Q}\}$ for all $t \in \mathbb{R}$. Then $H : \mathbb{R} \to [0,1]$ is an increasing right continuous function.*

**Proof.** If $t_1 < t_2$, then

$$\{h(s) : t_1 < s \in \mathbb{Q}\} \subset \{h(s) : t_2 < s \in \mathbb{Q}\}$$

and therefore $H(t_1) \le H(t_2)$. Let $H(t+) := \lim_{\tau \downarrow t} H(\tau)$. Then for any $s \in \mathbb{Q}$ with $s > t$ we have $H(t) \le H(t+) \le h(s)$ and then taking the infimum over such $s$ we learn that $H(t) \le H(t+) \le H(t)$, i.e. $H(t+) = H(t)$. ∎

**Lemma 16.38.** *Suppose that $(\mathbb{X}, \mathcal{M})$ is a measurable space and $F : \mathbb{X} \times \mathbb{R} \to \mathbb{R}$ is a function such that; 1) $F(\cdot, t) : \mathbb{X} \to \mathbb{R}$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $t \in \mathbb{R}$, and 2) $F(x, \cdot) : \mathbb{R} \to \mathbb{R}$ is right continuous for all $x \in \mathbb{X}$. Then $F$ is $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable.*

**Proof.** For $n \in \mathbb{N}$, the function,

$$F_n(x,t) := \sum_{k=-\infty}^{\infty} F\left(x, (k+1)2^{-n}\right) 1_{(k2^{-n},(k+1)2^{-n}]}(t),$$

is $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable. Using the right continuity assumption, it follows that $F(x,t) = \lim_{n\to\infty} F_n(x,t)$ for all $(x,t) \in \mathbb{X} \times \mathbb{R}$ and therefore $F$ is also $\mathcal{M} \otimes \mathcal{B}_{\mathbb{R}}/\mathcal{B}_{\mathbb{R}}$ – measurable. ∎

**Proposition 16.39.** *Let $\mathcal{B}_{\mathbb{R}}$ be the Borel $\sigma$ – algebra on $\mathbb{R}$. Then $\mathcal{B}_{\mathbb{R}}$ contains a countable sub-algebra, $\mathcal{A}_{\mathbb{R}} \subset \mathcal{B}_{\mathbb{R}}$, which generates $\mathcal{B}_{\mathbb{R}}$ and has the amazing property that every finitely additive probability measure on $\mathcal{A}_{\mathbb{R}}$ extends uniquely to a countably additive probability measure on $\mathcal{B}_{\mathbb{R}}$.*

**Proof.** By the results in Appendix 11.11, we know that $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measure theoretically isomorphic to $\left(\{0,1\}^{\mathbb{N}}, \mathcal{F}\right)$ where $\mathcal{F}$ is the product $\sigma$ – algebra. As we saw in Section 6.10, $\mathcal{F}$ is generated by the countable algebra, $\mathcal{A} := \cup_{n=1}^{\infty} \mathcal{A}_n$ where

$$\mathcal{A}_n := \{B \times \Omega : B \subset \{0,1\}^n\} \text{ for all } n \in \mathbb{N}.$$

According to the baby Kolmogorov Theorem 6.63, any finitely additive probability measure on $\mathcal{A}$ has a unique extension to a probability measure on $\mathcal{F}$. The algebra $\mathcal{A}$ may now be transferred by the measure theoretic isomorphism to the desired sub-algebra, $\mathcal{A}_{\mathbb{R}}$, of $\mathcal{B}_{\mathbb{R}}$. ∎

**Theorem 16.40.** *Suppose that $(\mathbb{X}, \mathcal{M})$ is a measurable space, $X : \Omega \to \mathbb{X}$ is a measurable function and $Y : \Omega \to \mathbb{R}$ is a random variable. Then there exists a probability kernel, $Q$, on $\mathbb{X} \times \mathbb{R}$ such that $\mathbb{E}[f(Y)|X] = Q(X, f)$, $P$ – a.s., for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$.*

**Proof. First proof.** For each $r \in \mathbb{Q}$, let $q_r : \mathbb{X} \to [0,1]$ be a measurable function such that

$$\mathbb{E}[1_{Y \le r}|X] = q_r(X) \text{ a.s.}$$

Let $\nu := P \circ X^{-1}$ be the law of $X$. Then using the basic properties of conditional expectation, $q_r \le q_s$ $\nu$ – a.s. for all $r \le s$, $\lim_{r\uparrow\infty} q_r = 1$ and $\lim_{r\downarrow\infty} q_r = 0$, $\nu$ – a.s. Hence the set, $\mathbb{X}_0 \subset \mathbb{X}$ where $q_r(x) \le q_s(x)$ for all $r \le s$, $\lim_{r\uparrow\infty} q_r(x) = 1$, and $\lim_{r\downarrow\infty} q_r(x) = 0$ satisfies, $\nu(\mathbb{X}_0) = P(X \in \mathbb{X}_0) = 1$. For $t \in \mathbb{R}$, let

$$F(x,t) := 1_{\mathbb{X}_0}(x) \cdot \inf\{q_r(x) : r > t\} + 1_{\mathbb{X} \setminus \mathbb{X}_0}(x) \cdot 1_{t \ge 0}.$$

Then $F(\cdot, t) : \mathbb{X} \to \mathbb{R}$ is measurable for each $t \in \mathbb{R}$ and by Lemma 16.37, $F(x, \cdot)$ is a distribution function on $\mathbb{R}$ for each $x \in \mathbb{X}$. Hence an application of Lemma 16.38 shows $F : \mathbb{X} \times \mathbb{R} \to [0,1]$ is measurable.

For each $x \in \mathbb{X}$ and $B \in \mathcal{B}_{\mathbb{R}}$, let $Q(x,B) = \mu_{F(x,\cdot)}(B)$ where $\mu_F$ denotes the probability measure on $\mathbb{R}$ determined by a distribution function, $F : \mathbb{R} \to [0,1]$.

We will now show that $Q$ is the desired probability kernel. To prove this, let $\mathbb{H}$ be the collection of bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{X} \ni x \to Q(x,f) \in \mathbb{R}$ is measurable and $\mathbb{E}[f(Y)|X] = Q(X,f)$, $P$ – a.s. It is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. We will finish the proof by showing that $\mathbb{H}$ contains the multiplicative class, $\mathbb{M} = \left\{1_{(-\infty,t]} : t \in \mathbb{R}\right\}$ so that multiplicative systems Theorem 10.2 may be applied.

Notice that $Q\left(x, 1_{(-\infty,t]}\right) = F(x,t)$ is measurable. Now let $r \in \mathbb{Q}$ and $g : \mathbb{X} \to \mathbb{R}$ be a bounded measurable function, then

$$\begin{aligned}\mathbb{E}[1_{Y \le r} \cdot g(X)] &= \mathbb{E}[\mathbb{E}[1_{Y \le r}|X]\, g(X)] = \mathbb{E}[q_r(X)\, g(X)] \\ &= \mathbb{E}[q_r(X)\, 1_{\mathbb{X}_0}(X)\, g(X)].\end{aligned}$$

For $t \in \mathbb{R}$, we may let $r \downarrow t$ in the above equality (use DCT) to learn,

$$\mathbb{E}[1_{Y \le t} \cdot g(X)] = \mathbb{E}[F(X,t)\, 1_{\mathbb{X}_0}(X)\, g(X)] = \mathbb{E}[F(X,t)\, g(X)].$$

Since $g$ was arbitrary, we may conclude that

$$Q\left(X, 1_{(-\infty,t]}\right) = F(X,t) = \mathbb{E}[1_{Y \le t}|X] \text{ a.s.}$$

This completes the proof.

**Second proof.** Let $\mathcal{A} := \mathcal{A}_{\mathbb{R}}$ be the algebra described in Proposition 16.39. For each $A \in \mathcal{A}$, let $\mu_A : \mathbb{X} \to \mathbb{R}$ be a measurable function such that $\mu_A(X) = P(Y \in A|X)$ a.s. If $A = A_1 \cup A_2$ with $A_i \in \mathcal{A}$ and $A_1 \cap A_2 = \emptyset$, then

$$\begin{aligned}\mu_{A_1}(X) + \mu_{A_2}(X) &= P(Y \in A_1|X) + P(Y \in A_2|X) \\ &= P(Y \in A_1 \cup A_2|X) = \mu_{A_1+A_2}(X) \text{ a.s.}\end{aligned}$$

Thus if $\nu := \mathrm{Law}_P(X)$, we have $\mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1+A_2}(x)$ for $\nu$ – a.e. $x$. Since

$$\mu_{\mathbb{R}}(X) = P(Y \in \mathbb{R}|X) = 1 \text{ a.s.}$$

we know that $\mu_{\mathbb{R}}(x) = 1$ for $\nu$ – a.e. $x$.

Thus if we let $X_0$ denote those $x \in X$ such that $\mu_{\mathbb{R}}(x) = 1$ and $\mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1+A_2}(x)$ for all disjoint pairs, $(A_1, A_2) \in \mathcal{A}^2$, we have $\nu(X_0) = 1$ and $\mathcal{A} \ni A \to Q_0(x, A) := \mu_A(x)$ is a finitely additive probability measure on $\mathcal{A}$. According to Proposition 16.39, $Q_0(x, \cdot)$ extends to a probability measure, $Q(x, \cdot)$ on $\mathcal{B}_{\mathbb{R}}$ for all $x \in X_0$. For $x \notin X_0$ we let $Q_0(x, \cdot) = \delta_0$ where $\delta_0(B) = 1_B(0)$ for all $B \in \mathcal{B}_{\mathbb{R}}$.

We will now show that $Q$ is the desired probability kernel. To prove this, let $\mathbb{H}$ be the collection of bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{X} \ni x \to Q(x, f) \in \mathbb{R}$ is measurable and $\mathbb{E}[f(Y)|X] = Q(X, f)$, $P$ – a.s. By construction, $\mathbb{H}$ contains the multiplicative system, $\{1_A : A \in \mathcal{A}\}$. Moreover it is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. Therefore by the multiplicative systems Theorem 10.2, $\mathbb{H}$ consists of all bounded measurable functions on $\mathbb{R}$. ■

This result leads fairly immediately to the following far reaching generalization.

**Theorem 16.41.** *Suppose that* $(\mathbb{X}, \mathcal{M})$ *is a measurable space and* $(\mathbb{Y}, \mathcal{N})$ *is a standard Borel space*[4]*, see Appendix 11.11. Suppose that* $X : \Omega \to \mathbb{X}$ *and* $Y : \Omega \to \mathbb{Y}$ *are measurable functions. Then there exists a probability kernel,* $Q$*, on* $\mathbb{X} \times \mathbb{Y}$ *such that* $\mathbb{E}[f(Y)|X] = Q(X, f)$, $P$ *– a.s., for all bounded measurable functions,* $f : \mathbb{Y} \to \mathbb{R}$.

**Proof.** By definition of a standard Borel space, we may assume that $\mathbb{Y} \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{N} = \mathcal{B}_{\mathbb{Y}}$. In this case $Y$ may also be viewed to be a measurable map form $\Omega \to \mathbb{R}$ such that $Y(\Omega) \subset \mathbb{Y}$. By Theorem 16.40, we may find a probability kernel, $Q_0$, on $\mathbb{X} \times \mathbb{R}$ such that

$$\mathbb{E}[f(Y)|X] = Q_0(X, f), \ P - a.s., \tag{16.53}$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$.

Taking $f = 1_{\mathbb{Y}}$ in Eq. (16.53) shows

$$1 = \mathbb{E}[1_{\mathbb{Y}}(Y)|X] = Q_0(X, \mathbb{Y}) \text{ a.s..}$$

[4] According to the counter example in Doob [8, p. 624], it is not sufficient to assume that $\mathcal{N}$ is countably generated!

Thus if we let $\mathbb{X}_0 := \{x \in \mathbb{X} : Q_0(x, \mathbb{Y}) = 1\}$, we know that $P(X \in \mathbb{X}_0) = 1$. Let us now define

$$Q(x, B) := 1_{\mathbb{X}_0}(x) Q_0(x, B) + 1_{\mathbb{X} \setminus \mathbb{X}_0}(x) \delta_y(B) \text{ for } (x, B) \in \mathbb{X} \times \mathcal{B}_{\mathbb{Y}},$$

where $y$ is an arbitrary but fixed point in $Y$. Then and hence $Q$ is a probability kernel on $\mathbb{X} \times \mathbb{Y}$. Moreover if $B \in \mathcal{B}_{\mathbb{Y}} \subset \mathcal{B}_{\mathbb{R}}$, then

$$Q(X, B) = 1_{\mathbb{X}_0}(X) Q_0(X, B) = 1_{\mathbb{X}_0}(X) \mathbb{E}[1_B(Y)|X] = \mathbb{E}[1_B(Y)|X] \text{ a.s.}$$

This shows that $Q$ is the desired regular conditional probability. ■

**Corollary 16.42.** *Suppose* $\mathcal{G}$ *is a sub-*$\sigma$ *– algebra of* $\mathcal{B}$*,* $(\mathbb{Y}, \mathcal{N})$ *is a standard Borel space, and* $Y : \Omega \to \mathbb{Y}$ *is a measurable function. Then there exists a probability kernel,* $Q$*, on* $(\Omega, \mathcal{G}) \times (\mathbb{Y}, \mathcal{N})$ *such that* $\mathbb{E}[f(Y)|\mathcal{G}] = Q(\cdot, f)$, $P$ *- a.s. for all bounded measurable functions,* $f : \mathbb{Y} \to \mathbb{R}$.

**Proof.** This is a special case of Theorem 16.41 applied with $(\mathbb{X}, \mathcal{M}) = (\Omega, \mathcal{G})$ and $X : \Omega \to \Omega$ being the identity map which is $\mathcal{B}/\mathcal{G}$ – measurable. ■

**Corollary 16.43.** *Suppose that* $(\Omega, \mathcal{B}, P)$ *is a probability space such that* $(\Omega, \mathcal{B})$ *is a standard Borel space and* $\mathcal{G}$ *is a sub-*$\sigma$ *– algebra* $\mathcal{B}$*. Then there exists a probability kernel,* $Q$ *on* $(\Omega, \mathcal{G}) \times (\Omega, \mathcal{B})$ *such that* $\mathbb{E}[Z|\mathcal{G}] = Q(\cdot, Z)$, $P$ *- a.s. for all bounded* $\mathcal{B}$ *– measurable random variables,* $Z : \Omega \to \mathbb{R}$.

**Proof.** This is a special case of Corollary 16.42 with $(\mathbb{Y}, \mathcal{N}) = (\Omega, \mathcal{B})$ and $Y : \Omega \to \Omega$ being the identity map which is $\mathcal{B}/\mathcal{B}$ – measurable. ■

*Remark 16.44.* It turns out that every standard Borel space $(\mathbb{X}, \mathcal{M})$ possess a countable sub-algebra $\mathcal{A}$ generating $\mathcal{M}$ with the property that every finitely additive probability measure on $\mathcal{A}$ extends to a probability measure on $\mathcal{M}$, see [4]. With this in hand, the second proof of Theorem 16.40 extends easily to give another proof of Theorem 16.41 all in one go. As the next example shows it is a bit tricky to produce the algebra $\mathcal{A}$.

*Example 16.45.* Let $\Omega := \{0,1\}^{\mathbb{N}}$, $\pi_i : \Omega \to \{0,1\}$ be projection onto the $i^{\text{th}}$ component and $\mathcal{B} := \sigma(\pi_1, \pi_2, \dots)$ be the product $\sigma$ – algebra on $\Omega$. Further let $\mathcal{A} := \cup_{n=1}^{\infty} \mathcal{A}_n$ where

$$\mathcal{A}_n := \{B \times \Omega : B \subset \{0,1\}^n\} \text{ for all } n \in \mathbb{N}.$$

Suppose that $X = \{e_n\}_{n=1}^{\infty} \subset \Omega$ where $e_n(i) = \delta_{in}$ for $i, n \in \mathbb{N}$. I now claim that

$$\mathcal{A}_X = \{A \subset X : \#(A) < \infty \text{ or } \#(A^c) < \infty\} =: \mathcal{C}$$

is the so called cofinite $\sigma$ – algebra. To see this observe that $\mathcal{A}$ is generated by sets of the form $\{\pi_i = 1\}$ for $i \in \mathbb{N}$. Therefore $\mathcal{A}_X$ is generated by sets of the form $\{\pi_i = 1\}_X = \{e_i\}$. But these one point sets are easily seen to generate $\mathcal{C}$.

Now suppose that $\lambda : X \to [0,1]$ is a function such that $Z := \sum_{n\in\mathbb{N}} \lambda(e_n) \in (0,1)$ and let $\mu(B) := \sum_{a\in B} \lambda(a)$ for all $B \subset X$. Then $\mu$ is a measure on $2^X$ with $\mu(X) = Z < 1$.

Using this measure $\mu$, we may define $P_0 : \mathcal{A}_X = \mathcal{C} \to [0,1]$ by,

$$P_0(A) := \begin{cases} \mu(A) & \text{if } \#(A) < \infty \\ 1 - \mu(A^c) & \text{if } \#(A^c) < \infty \end{cases}.$$

I claim that $P_0$ is a finitely additive probability measure on $\mathcal{A}_X = \mathcal{C}$ which has no -extension to a probability measure on $2^X$. To see that $P_0$ is finitely additive, let $A, B \in \mathcal{C}$ be disjoint sets. If both $A$ and $B$ are finite sets, then

$$P_0(A \cup B) = \mu(A \cup B) = \mu(A) + \mu(B) = P_0(A) + P_0(B).$$

If one of the sets is an infinite set, say $B$, then $\#(B^c) < \infty$ and $\#(A) < \infty$ for otherwise $A \cap B \neq \emptyset$. As $A \cap B = \emptyset$ we know that $A \subset B^c$ and therefore,

$$\begin{aligned} P_0(A \cup B) &= 1 - \mu([A \cup B]^c) = 1 - \mu(A^c \cap B^c) \\ &= 1 - \mu(B^c \setminus A) = 1 - (\mu(B^c) - \mu(A)) \\ &= 1 - \mu(B^c) + \mu(A) = P_0(B) + P_0(A). \end{aligned}$$

Thus we have shown that $P_0 : \mathcal{A}_X \to [0,1]$ is a finitely additive probability measure. If $P$ were a countably additive extension of $P_0$, we would have to have,

$$\begin{aligned} 1 = P_0(X) = P(X) &= \sum_{n=1}^{\infty} P(\{e_n\}) \\ &= \sum_{n=1}^{\infty} P_0(\{e_n\}) = \sum_{n=1}^{\infty} \mu(\{e_n\}) = Z < 1 \end{aligned}$$

which is clearly a contradiction.

There is however a way to fix this example as shown in [4]. It is to replace $\mathcal{A}_X$ in this example by the algebra, $\mathcal{A}$, generated by $\mathcal{E} := \{\{n\} : n \geq 2\}$. This algebra may be described as those $A \subset \mathbb{N}$ such that either $A \subset_f \{2, 3, \dots\}$ for $1 \in A$ and $\#(A^c) < \infty$. Thus if $A_k \in \mathcal{A}$ with $A_k \downarrow \emptyset$ we must have that $1 \notin A_k$ for $k$ large and therefore $\#(A_k) < \infty$ for $k$ large. Moreover $\#(A_k)$ is decreasing in $k$. If $\lim_{k\to\infty} \#(A_k) = m > 0$, we must have that $A_k = A_l$ for all $k, l$ large and therefore $\cap A_k \neq \emptyset$. Thus we must conclude that $A_k = \emptyset$ for large $k$. We therefore may conclude that any finitely additive probability measure, $P_0$, on $\mathcal{A}$ has a unique extension to a probability measure on $\sigma(\mathcal{A}) = 2^{\mathbb{N}}$.

# 17

# The Radon-Nikodym Theorem

**Theorem 17.1 (A Baby Radon-Nikodym Theorem).** *Suppose $(X, \mathcal{M})$ is a measurable space, $\lambda$ and $\nu$ are two finite positive measures on $\mathcal{M}$ such that $\nu(A) \leq \lambda(A)$ for all $A \in \mathcal{M}$. Then there exists a measurable function, $\rho : X \to [0, 1]$ such that $d\nu = \rho d\lambda$.*

**Proof.** If $f$ is a non-negative simple function, then

$$\nu(f) = \sum_{a \geq 0} a\nu(f = a) \leq \sum_{a \geq 0} a\lambda(f = a) = \lambda(f).$$

In light of Theorem 8.39 and the MCT, this inequality continues to hold for all non-negative measurable functions. Furthermore if $f \in L^1(\lambda)$, then $\nu(|f|) \leq \lambda(|f|) < \infty$ and hence $f \in L^1(\nu)$ and

$$|\nu(f)| \leq \nu(|f|) \leq \lambda(|f|) \leq \lambda(X)^{1/2} \cdot \|f\|_{L^2(\lambda)}.$$

Therefore, $L^2(\lambda) \ni f \to \nu(f) \in \mathbb{C}$ is a continuous linear functional on $L^2(\lambda)$. By the Riesz representation Theorem 15.16, there exists a unique $\rho \in L^2(\lambda)$ such that

$$\nu(f) = \int_X f\rho d\lambda \text{ for all } f \in L^2(\lambda).$$

In particular this equation holds for all bounded measurable functions, $f : X \to \mathbb{R}$ and for such a function we have

$$\nu(f) = \operatorname{Re}\nu(f) = \operatorname{Re}\int_X f\rho d\lambda = \int_X f \operatorname{Re}\rho \, d\lambda. \tag{17.1}$$

Thus by replacing $\rho$ by $\operatorname{Re}\rho$ if necessary we may assume $\rho$ is real.

Taking $f = 1_{\rho<0}$ in Eq. (17.1) shows

$$0 \leq \nu(\rho < 0) = \int_X 1_{\rho<0}\rho \, d\lambda \leq 0,$$

from which we conclude that $1_{\rho<0}\rho = 0$, $\lambda$ – a.e., i.e. $\lambda(\rho < 0) = 0$. Therefore $\rho \geq 0$, $\lambda$ – a.e. Similarly for $\alpha > 1$,

$$\lambda(\rho > \alpha) \geq \nu(\rho > \alpha) = \int_X 1_{\rho>\alpha}\rho \, d\lambda \geq \alpha\lambda(\rho > \alpha)$$

which is possible iff $\lambda(\rho > \alpha) = 0$. Letting $\alpha \downarrow 1$, it follows that $\lambda(\rho > 1) = 0$ and hence $0 \leq \rho \leq 1$, $\lambda$ - a.e. ∎

**Definition 17.2.** *Let $\mu$ and $\nu$ be two positive measure on a measurable space, $(X, \mathcal{M})$. Then:*

1. *$\mu$ and $\nu$ are **mutually singular** (written as $\mu \perp \nu$) if there exists $A \in \mathcal{M}$ such that $\nu(A) = 0$ and $\mu(A^c) = 0$. We say that $\nu$ lives on $A$ and $\mu$ lives on $A^c$.*
2. *The measure $\nu$ **is absolutely continuous relative to** $\mu$ (written as $\nu \ll \mu$) provided $\nu(A) = 0$ whenever $\mu(A) = 0$.*

As an example, suppose that $\mu$ is a positive measure and $\rho \geq 0$ is a measurable function. Then the measure, $\nu := \rho\mu$ is absolutely continuous relative to $\mu$. Indeed, if $\mu(A) = 0$ then

$$\nu(A) = \int_A \rho d\mu = 0.$$

We will eventually show that if $\mu$ and $\nu$ are $\sigma$ – finite and $\nu \ll \mu$, then $d\nu = \rho d\mu$ for some measurable function, $\rho \geq 0$.

**Definition 17.3 (Lebesgue Decomposition).** *Let $\mu$ and $\nu$ be two positive measure on a measurable space, $(X, \mathcal{M})$. Two positive measures $\nu_a$ and $\nu_s$ form a **Lebesgue decomposition** of $\nu$ relative to $\mu$ if $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$, and $\nu_s \perp \mu$.*

**Lemma 17.4.** *If $\mu_1$, $\mu_2$ and $\nu$ are positive measures on $(X, \mathcal{M})$ such that $\mu_1 \perp \nu$ and $\mu_2 \perp \nu$, then $(\mu_1 + \mu_2) \perp \nu$. More generally if $\{\mu_i\}_{i=1}^\infty$ is a sequence of positive measures such that $\mu_i \perp \nu$ for all $i$ then $\mu = \sum_{i=1}^\infty \mu_i$ is singular relative to $\nu$.*

**Proof.** It suffices to prove the second assertion since we can then take $\mu_j \equiv 0$ for all $j \geq 3$. Choose $A_i \in \mathcal{M}$ such that $\nu(A_i) = 0$ and $\mu_i(A_i^c) = 0$ for all $i$. Letting $A := \cup_i A_i$ we have $\nu(A) = 0$. Moreover, since $A^c = \cap_i A_i^c \subset A_m^c$ for all $m$, we have $\mu_i(A^c) = 0$ for all $i$ and therefore, $\mu(A^c) = 0$. This shows that $\mu \perp \nu$. ∎

**Lemma 17.5.** *Let $\nu$ and $\mu$ be positive measures on $(X, \mathcal{M})$. If there exists a Lebesgue decomposition, $\nu = \nu_s + \nu_a$, of the measure $\nu$ relative to $\mu$ then this decomposition is unique. Moreover: if $\nu$ is a $\sigma$ – finite measure then so are $\nu_s$ and $\nu_a$.*

**Proof.** Since $\nu_s \perp \mu$, there exists $A \in \mathcal{M}$ such that $\mu(A) = 0$ and $\nu_s(A^c) = 0$ and because $\nu_a \ll \mu$, we also know that $\nu_a(A) = 0$. So for $C \in \mathcal{M}$,

$$\nu(C \cap A) = \nu_s(C \cap A) + \nu_a(C \cap A) = \nu_s(C \cap A) = \nu_s(C) \tag{17.2}$$

and

$$\nu(C \cap A^c) = \nu_s(C \cap A^c) + \nu_a(C \cap A^c) = \nu_a(C \cap A^c) = \nu_a(C). \tag{17.3}$$

Now suppose we have another Lebesgue decomposition, $\nu = \tilde{\nu}_a + \tilde{\nu}_s$ with $\tilde{\nu}_s \perp \mu$ and $\tilde{\nu}_a \ll \mu$. Working as above, we may choose $\tilde{A} \in \mathcal{M}$ such that $\mu(\tilde{A}) = 0$ and $\tilde{A}^c$ is $\tilde{\nu}_s$ – null. Then $B = A \cup \tilde{A}$ is still a $\mu$ – null set and and $B^c = A^c \cap \tilde{A}^c$ is a null set for both $\nu_s$ and $\tilde{\nu}_s$. Therefore we may use Eqs. (17.2) and (17.3) with $A$ being replaced by $B$ to conclude,

$$\begin{aligned} \nu_s(C) &= \nu(C \cap B) = \tilde{\nu}_s(C) \text{ and} \\ \nu_a(C) &= \nu(C \cap B^c) = \tilde{\nu}_a(C) \text{ for all } C \in \mathcal{M}. \end{aligned}$$

Lastly if $\nu$ is a $\sigma$ – finite measure then there exists $X_n \in \mathcal{M}$ such that $X = \sum_{n=1}^{\infty} X_n$ and $\nu(X_n) < \infty$ for all $n$. Since $\infty > \nu(X_n) = \nu_a(X_n) + \nu_s(X_n)$, we must have $\nu_a(X_n) < \infty$ and $\nu_s(X_n) < \infty$, showing $\nu_a$ and $\nu_s$ are $\sigma$ – finite as well. ■

**Lemma 17.6.** *Suppose $\mu$ is a positive measure on $(X, \mathcal{M})$ and $f, g : X \to [0, \infty]$ are functions such that the measures, $fd\mu$ and $gd\mu$ are $\sigma$ – finite and further satisfy,*

$$\int_A f d\mu = \int_A g d\mu \text{ for all } A \in \mathcal{M}. \tag{17.4}$$

*Then $f(x) = g(x)$ for $\mu$ – a.e. $x$. (BRUCE: this lemma is very closely related to Lemma 9.24 above.)*

**Proof.** By assumption there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\int_{X_n} f d\mu < \infty$ and $\int_{X_n} g d\mu < \infty$ for all $n$. Replacing $A$ by $A \cap X_n$ in Eq. (17.4) implies

$$\int_A 1_{X_n} f d\mu = \int_{A \cap X_n} f d\mu = \int_{A \cap X_n} g d\mu = \int_A 1_{X_n} g d\mu$$

for all $A \in \mathcal{M}$. Since $1_{X_n} f$ and $1_{X_n} g$ are in $L^1(\mu)$ for all $n$, this equation implies $1_{X_n} f = 1_{X_n} g$, $\mu$ – a.e. Letting $n \to \infty$ then shows that $f = g$, $\mu$ – a.e. ■

*Remark 17.7.* Lemma 17.6 is in general false without the $\sigma$ – finiteness assumption. A trivial counterexample is to take $\mathcal{M} = 2^X$, $\mu(A) = \infty$ for all non-empty $A \in \mathcal{M}$, $f = 1_X$ and $g = 2 \cdot 1_X$. Then Eq. (17.4) holds yet $f \neq g$.

**Theorem 17.8 (Radon Nikodym Theorem for Positive Measures).** *Suppose that $\mu$ and $\nu$ are $\sigma$ – finite positive measures on $(X, \mathcal{M})$. Then $\nu$ has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to $\mu$ and there exists a unique (modulo sets of $\mu$ – measure 0) function $\rho : X \to [0, \infty)$ such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$.*

**Proof.** The uniqueness assertions follow directly from Lemmas 17.5 and 17.6.

**Existence when $\mu$ and $\nu$ are both finite measures.** (Von-Neumann's Proof. See Remark 17.9 for the motivation for this proof.) First suppose that $\mu$ and $\nu$ are **finite** measures and let $\lambda = \mu + \nu$. By Theorem 17.1, $d\nu = hd\lambda$ with $0 \le h \le 1$ and this implies, for all non-negative measurable functions $f$, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh) \tag{17.5}$$

or equivalently

$$\nu(f(1-h)) = \mu(fh). \tag{17.6}$$

Taking $f = 1_{\{h=1\}}$ in Eq. (17.6) shows that

$$\mu(\{h = 1\}) = \nu(1_{\{h=1\}}(1-h)) = 0,$$

i.e. $0 \le h(x) < 1$ for $\mu$ - a.e. $x$. Let

$$\rho := 1_{\{h<1\}} \frac{h}{1-h}$$

and then take $f = g 1_{\{h<1\}}(1-h)^{-1}$ with $g \ge 0$ in Eq. (17.6) to learn

$$\nu(g 1_{\{h<1\}}) = \mu(g 1_{\{h<1\}}(1-h)^{-1} h) = \mu(\rho g).$$

Hence if we define

$$\nu_a := 1_{\{h<1\}} \nu \text{ and } \nu_s := 1_{\{h=1\}} \nu,$$

we then have $\nu_s \perp \mu$ (since $\nu_s$ "lives" on $\{h = 1\}$ while $\mu(h = 1) = 0$) and $\nu_a = \rho\mu$ and in particular $\nu_a \ll \mu$. Hence $\nu = \nu_a + \nu_s$ is the desired Lebesgue decomposition of $\nu$. If we further assume that $\nu \ll \mu$, then $\mu(h = 1) = 0$ implies $\nu(h = 1) = 0$ and hence that $\nu_s = 0$ and we conclude that $\nu = \nu_a = \rho\mu$.

**Existence when $\mu$ and $\nu$ are $\sigma$-finite measures.** Write $X = \sum_{n=1}^{\infty} X_n$ where $X_n \in \mathcal{M}$ are chosen so that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$ for all $n$. Let $d\mu_n = 1_{X_n} d\mu$ and $d\nu_n = 1_{X_n} d\nu$. Then by what we have just proved there exists $\rho_n \in L^1(X, \mu_n) \subset L^1(X, \mu)$ and measure $\nu_n^s$ such that $d\nu_n = \rho_n d\mu_n + d\nu_n^s$ with $\nu_n^s \perp \mu_n$. Since $\mu_n$ and $\nu_n^s$ "live" on $X_n$ there exists $A_n \in \mathcal{M}_{X_n}$ such that $\mu(A_n) = \mu_n(A_n) = 0$ and

$$\nu_n^s(X \setminus A_n) = \nu_n^s(X_n \setminus A_n) = 0.$$

This shows that $\nu_n^s \perp \mu$ for all $n$ and so by Lemma 17.4, $\nu_s := \sum_{n=1}^\infty \nu_n^s$ is singular relative to $\mu$. Since

$$\nu = \sum_{n=1}^\infty \nu_n = \sum_{n=1}^\infty (\rho_n \mu_n + \nu_n^s) = \sum_{n=1}^\infty (\rho_n 1_{X_n}\mu + \nu_n^s) = \rho\mu + \nu_s, \tag{17.7}$$

where $\rho := \sum_{n=1}^\infty 1_{X_n}\rho_n$, it follows that $\nu = \nu_a + \nu_s$ with $\nu_a = \rho\mu$. Hence this is the desired Lebesgue decomposition of $\nu$ relative to $\mu$. ∎

*Remark 17.9.* Here is the motivation for the above construction. Suppose that $d\nu = d\nu_s + \rho d\mu$ is the Radon-Nikodym decomposition and $X = A \sum B$ such that $\nu_s(B) = 0$ and $\mu(A) = 0$. Then we find

$$\nu_s(f) + \mu(\rho f) = \nu(f) = \lambda(hf) = \nu(hf) + \mu(hf).$$

Letting $f \to 1_A f$ then implies that

$$\nu(1_A f) = \nu_s(1_A f) = \nu(1_A hf)$$

which show that $h = 1$, $\nu$ –a.e. on $A$. Also letting $f \to 1_B f$ implies that

$$\mu(\rho 1_B f) = \nu(h 1_B f) + \mu(h 1_B f) = \mu(\rho h 1_B f) + \mu(h 1_B f)$$

which implies, $\rho = \rho h + h$, $\mu$ – a.e. on $B$, i.e.

$$\rho(1-h) = h,\ \mu\text{– a.e. on } B.$$

In particular it follows that $h < 1$, $\mu = \nu$ – a.e. on $B$ and that $\rho = \frac{h}{1-h}1_{h<1}$, $\mu$ – a.e. So up to sets of $\nu$ – measure zero, $A = \{h = 1\}$ and $B = \{h < 1\}$ and therefore,

$$d\nu = 1_{\{h=1\}}d\nu + 1_{\{h<1\}}d\nu = 1_{\{h=1\}}d\nu + \frac{h}{1-h}1_{h<1}d\mu.$$

## 17.1 Proof of the Change of Variables Theorem* 11.21

*This section is still very rough.

**A better proof of Theorem 11.21.** Here is perhaps the proof I should use in this book for the change of variables theorem. As usual let $T : \Omega \to T(\Omega)$ be a $C^1$ – diffeomorpshism and assume both $T$ and $T^{-1}$ have globally bounded Lipschitz constants (this can be achieved by shrinking $\Omega$ if necessary). We will work in the $\ell^\infty$ – norm on $\mathbb{R}^d$.



1. If $f : T(\Omega) \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ we have
$$\int_\Omega f \circ T dm = \int_{T(\Omega)} f d\left(m \circ T^{-1}\right) \tag{17.8}$$
and
$$\int_{T(\Omega)} g \circ T^{-1} dm = \int_\Omega g d(m \circ T). \tag{17.9}$$
2. Referring to the math 140 notes, show $|m \circ T(A)| \le Km(A)$ and similarly $\left|m \circ T^{-1}(A)\right| \le Km(A)$. Therefore by the easiest version of the the Radon – Nykodym there are bounded non-negative functions, $\alpha$ and $\beta$ such that
$$\frac{d(m \circ T)}{dm} = \alpha : \Omega \to \mathbb{R} \text{ and}$$
$$\frac{d\left(m \circ T^{-1}\right)}{dm} = \beta : T(\Omega) \to \mathbb{R}.$$
In other words we now have,
$$\int_\Omega f \circ T dm = \int_{T(\Omega)} f\beta dm \text{ and} \tag{17.10}$$
$$\int_{T(\Omega)} g \circ T^{-1} dm = \int_\Omega g\alpha dm. \tag{17.11}$$
3. There is a relationship between $\alpha$ and $\beta$. Indeed taking $g = (f\beta) \circ T$ shows
$$\int_\Omega f \circ T dm = \int_{T(\Omega)} f\beta dm = \int_{T(\Omega)} ((f\beta) \circ T) \circ T^{-1} dm = \int_\Omega (f \circ T) \cdot (\beta \circ T)\alpha dm$$
from which we conclude
$$(\beta \circ T)\alpha = 1 \text{ a.e.} \iff \alpha = \frac{1}{\beta \circ T} \tag{17.12}$$
We now wish to compute the functions $\alpha$ and $\beta$ by taking limits and for this we will use the Lebesgue differentiation theorem or the easier fact that $\delta_r * \alpha \to \alpha$ in $L^1_{loc}$ as $r \downarrow 0$ where $\delta_r(x) := \frac{1}{m(B_r(0))}1_{B_r(0)}$.
4. So fix and $x \in \Omega$ and consider
$$\begin{aligned} T'(x)^{-1}[T(y) - T(x)] &= T'(x)^{-1}\left[\int_0^1 T'(x + t(y-x))(y-x)dt\right] \\ &= \left[\int_0^1 T'(x)^{-1}T'(x + t(y-x))(y-x)dt\right] \\ &= y - x + \varepsilon(x,y)(y-x) \end{aligned}$$

where

$$\varepsilon(x,y) := \int_0^1 \left[T'(x)^{-1} T'(x + t(y-x)) - I\right] dt$$

satisfies,

$$\|\varepsilon(x,y)\| \le \int_0^1 \left\|T'(x)^{-1} T'(x + t(y-x)) - I\right\| dt = \varepsilon(\|y-x\|).$$

From this it follows if $y \in B_r(x) = x + B_r(0)$, then

$$\left\|T'(x)^{-1}[T(y) - T(x)]\right\| \le r(1+\varepsilon(r))$$

and hence

$$T'(x)^{-1}[T(B_r(x)) - T(x)] \subset B_{r(1+\varepsilon(r))}(0)$$

or equivalently that

$$T(B_r(x)) \subset T(x) + T'(x) B_{r(1+\varepsilon(r))}(0). \tag{17.13}$$

5. Therefore it follows that

$$\begin{aligned}\frac{m(T(B_r(x)))}{m(B_r(x))} &\le \frac{m\left(T(x) + T'(x) B_{r(1+\varepsilon(r))}(0)\right)}{m(B_r(x))} \\ &= |\det T'(x)| \cdot (1+\varepsilon(r))^d.\end{aligned}$$

Letting $r \downarrow 0$ and using the Lebesgue differentiation theorem (see Theorem **??**) which is a rather deep result!)

$$\alpha(x) = \frac{d(m \circ T)}{dm}(x) \le |\det T'(x)| \text{ for a.e. } x \in \Omega. \tag{17.14}$$

Applying this result with $T$ replaced by $T^{-1}$ then shows,

$$\beta(y) = \frac{d\left(m \circ T^{-1}\right)}{dm}(y) \le \left|\det\left(T^{-1}\right)'(y)\right| \text{ for a.e. } y \in T(\Omega). \tag{17.15}$$

Alternatively let us observe that

$$\frac{m(T(B_r(x)))}{m(B_r(x))} = \frac{1}{m(B_r(x))}\int_{B_r(0)} \alpha(x+y)\,dy = \delta_r * \alpha(x).$$

By the easier approximate identity Theorem **??** we know $\delta_r * \alpha \to \alpha$ in $L^1_{loc}$ and so there exists $r_n \downarrow 0$ such that $\delta_{r_n} * \alpha \to \alpha$ a.e. as $n \to \infty$. Thus we again learn that for a.e. $x$,

$$\begin{aligned}\alpha(x) &= \lim_{n\to\infty} \delta_{r_n} * \alpha(x) = \lim_{n\to\infty} \frac{m(T(B_{r_n}(x)))}{m(B_{r_n}(x))} \\ &\le \lim_{n\to\infty} |\det T'(x)| \cdot (1+\varepsilon(r_n))^d = |\det T'(x)|.\end{aligned}$$

6. We now use $\alpha = \frac{1}{\beta \circ T}$ from Eq. (17.12) along with the inequality in Eq. (17.15) to learn

$$\alpha(x) = \frac{1}{\beta \circ T(x)} \ge \frac{1}{\left|\det\left(T^{-1}\right)'(T(x))\right|}.$$

On the other hand, since $T^{-1} \circ T = I$, it follows by the chain rule that $\left(T^{-1}\right)'(T(x))\,T'(x) = I$ and therefore

$$\frac{1}{\left|\det\left(T^{-1}\right)'(T(x))\right|} = |\det T'(x)|$$

and we may conclude $\alpha(x) \ge |\det T'(x)|$. This result along with the inequality in Eq. (17.14) shows

$$\alpha(x) = |\det T'(x)|.$$

7. Using this result back in Eq. (17.11) with $g = f \circ T$ for some function on $f : T(\Omega) \to \mathbb{R}$ gives,

$$\int_{T(\Omega)} f dm = \int_\Omega f \circ T \cdot |\det T'|\, dm.$$

**Note:** By working harder as in the inverse function Theorem **??**, we could have proved the stronger version of Eq. (17.13);

$$T(x) + T'(x) B_{r(1-\varepsilon(r))}(0) \subset T(B_r(x)) \subset T(x) + T'(x) B_{r(1+\varepsilon(r))}(0).$$

If we had done this we could have avoided discussing $\beta$ altogether. Indeed, the method of step 5. would then give

$$|\det T'(x)| \cdot (1-\varepsilon(r))^d \le \frac{m(T(B_r(x)))}{m(B_r(x))} \le |\det T'(x)| \cdot (1+\varepsilon(r))^d$$

which upon letting $r \downarrow 0$ would have shown $\frac{d[m \circ T]}{dm}(x) = |\det T'(x)|$. ■

# Some Ergodic Theory

The goal of this chapter is to show (in certain circumstances) that "time averages" are the same as "spatial averages." We start with a "simple" Hilbert space version of the type of theorem that we are after. For more on the following mean Ergodic theorem, see [19] and [11].

**Theorem 18.1 (Von-Neumann's Mean Ergodic Theorem).** *Let $U : H \to H$ be an isometry on a Hilbert space $H$, $M = \mathrm{Nul}(U-I)$, $P = P_M$ be orthogonal projection onto $M$, and $S_n = \sum_{k=0}^{n-1} U^k$. Show $\frac{S_n}{n} \to P_M$* ***strongly*** *by which we mean $\lim_{n\to\infty} \frac{S_n}{n} x = P_M x$ for all $x \in H$.*

**Proof.** Since $U$ is an isometry we have $(Ux, Uy) = (x, y)$ for all $x, y \in H$ and therefore that $U^*U = I$. In general it is not true that $UU^* = I$ but instead, $UU^* = P_{\mathrm{Ran}(U)}$. Thus $UU^* = I$ iff $U$ is surjective, i.e. $U$ is unitary.

Before starting the proof in earnest we need to prove

$$\mathrm{Nul}(U^* - I) = \mathrm{Nul}(U - I).$$

If $x \in \mathrm{Nul}\,(U - I)$ then $x = Ux$ and therefore $U^*x = U^*Ux = x$, i.e. $x \in \mathrm{Nul}(U^* - I)$. Conversely if $x \in \mathrm{Nul}(U^* - I)$ then $U^*x = x$ and we have

$$\begin{aligned}\|Ux - x\|^2 &= 2\|x\|^2 - 2\,\mathrm{Re}\,(Ux, x)\\ &= 2\|x\|^2 - 2\,\mathrm{Re}\,(x, U^*x) = 2\|x\|^2 - 2\,\mathrm{Re}\,(x, x) = 0\end{aligned}$$

which shows that $Ux = x$, i.e. $x \in \mathrm{Nul}\,(U - I)$. With this remark in hand we can easily complete the proof.

Let us first observe that

$$\frac{S_n}{n}(U - I) = \frac{1}{n}[U^n - I] \to 0 \text{ as } n \to \infty.$$

Thus if $x = (U - I)y \in \mathrm{Ran}(U - I)$, we have

$$\frac{S_n}{n} x = \frac{1}{n}(U^n y - y) \to 0 \text{ as } n \to \infty.$$

More generally if $x \in \overline{\mathrm{Ran}(U - I)}$ and $x' \in \mathrm{Ran}(U-I)$, we have, since $\left\|\frac{S_n}{n}\right\| \le 1$, that

$$\left\|\frac{S_n}{n}x - \frac{S_n}{n}x'\right\| \le \|x - x'\|$$

and hence

$$\limsup_{n\to\infty}\left\|\frac{S_n}{n}x\right\| = \limsup_{n\to\infty}\left\|\frac{S_n}{n}x - \frac{S_n}{n}x'\right\| \le \|x - x'\|.$$

Letting $x' \in \mathrm{Ran}\,(U - I)$ tend to $x \in \overline{\mathrm{Ran}(U - I)}$ allows us to conclude that $\limsup_{n\to\infty}\left\|\frac{S_n}{n}x\right\| = 0$.

For

$$x \in \overline{\mathrm{Ran}(U - I)}^{\perp} = \mathrm{Ran}\,(U - I)^{\perp} = \mathrm{Nul}\,(U^* - I) = \mathrm{Nul}\,(U - I) = M$$

we have $\frac{S_n}{n}x = x$. So for general $x \in H$, we have $x = P_M x + y$ with $y \in M^{\perp} = \overline{\mathrm{Ran}(U - I)}$ and therefore,

$$\frac{S_n}{n}x = \frac{S_n}{n}P_M x + \frac{S_n}{n}y = P_M x + \frac{S_n}{n}y \to P_M x \text{ as } n \to \infty.$$

■

For the rest of this section, suppose that $(\Omega, \mathcal{B}, \mu)$ is a $\sigma$ – finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_*\mu = \mu$. After Theorem 18.6 we will further assume that $\mu = P$ is a probability measure. For more results along the lines of this chapter, the reader is referred to Kallenberg [26, Chapter 10]. The reader may also benefit from Norris's notes in [33].

**Definition 18.2.** *Let*

$$\begin{aligned}\mathcal{B}_\theta &:= \left\{A \in \mathcal{B} : \theta^{-1}(A) = A\right\} \text{ and}\\ \mathcal{B}'_\theta &:= \left\{A \in \mathcal{B} : \mu\left(\theta^{-1}(A)\,\Delta A\right) = 0\right\}\end{aligned}$$

*be the* ***invariant*** *$\sigma$* ***– field*** *and* ***almost invariant*** *$\sigma$ –fields respectively.*

In what follows we will make use of the following easily proved set identities. Let $\{A_n\}_{n=1}^{\infty}$, $\{B_n\}_{n=1}^{\infty}$, and $A, B, C$ be a collection of subsets of $\Omega$, then;

1. $A \triangle C \subset [A \triangle B] \cup [B \triangle C]$,
2. $[\cup_{n=1}^{\infty} A_n]\,\Delta\,[\cup_{n=1}^{\infty} B_n] \subset \cup_{n=1}^{\infty} A_n \Delta B_n$,
3. $[\cap_{n=1}^{\infty} A_n]\,\Delta\,[\cap_{n=1}^{\infty} B_n] \subset \cup_{n=1}^{\infty} A_n \Delta B_n$,
4. $B \Delta \{A_n \text{ i.o.}\} \subset \cup_{n=1}^{\infty}[B \Delta A_n]$.

**Lemma 18.3.** *The elements of $\mathcal{B}'_\theta$ are the same as the elements in $\mathcal{B}_\theta$ modulo null sets, i.e.*

$$\mathcal{B}'_\theta = \{B \in \mathcal{B} : \exists\ A \in \mathcal{B}_\theta\ \ni\ \mu(A \triangle B) = 0\}.$$

*Moreover if $B \in \mathcal{B}'_\theta$, then*

$$A := \left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ i.o. } k\right\} \in \mathcal{B}_\theta \tag{18.1}$$

*and $\mu(A \triangle B) = 0$. (We could have just as well taken $A$ to be equal to $\left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ a.a.}\right\}$.)*

**Proof.** If $A \in \mathcal{B}_\theta$ and $B \in \mathcal{B}$ such that $\mu(A \triangle B) = 0$, then

$$\mu\left(A\Delta\theta^{-1}(B)\right) = \mu\left(\theta^{-1}(A)\,\Delta\theta^{-1}(B)\right) = \mu\theta^{-1}(A \triangle B) = \mu(A \triangle B) = 0$$

and therefore it follows that

$$\mu\left(B\Delta\theta^{-1}(B)\right) \le \mu(B\Delta A) + \mu\left(A\Delta\theta^{-1}(B)\right) = 0.$$

This shows that $B \in \mathcal{B}'_\theta$.

Conversely if $B \in \mathcal{B}'_\theta$ then by the invariance of $\mu$ under $\theta$ it follows that $\mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0$ for all $k = 0, 1, 2, 3 \dots$. In particular we learn that

$$\begin{aligned}\mu\left(\theta^{-k}(B)\,\Delta B\right) &= \mu\left(\left|1_{\theta^{-k}(B)} - 1_B\right|\right)\\ &\le \sum_{l=0}^{k-1}\mu\left(\left|1_{\theta^{-l}(B)} - 1_{\theta^{-(l+1)}(B)}\right|\right)\\ &= \sum_{l=0}^{k-1}\mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0.\end{aligned}$$

Thus if $A = \left\{\theta^{-k}(B) \text{ i.o. } k\right\}$ as in Eq. (18.1) we have,

$$\mu(B\Delta A) \le \sum_{k=1}^{\infty}\mu\left(B\Delta\theta^{-k}(B)\right) = 0.$$

This completes the proof since

$$\theta^{-1}(A) = \left\{\omega \in \Omega : \theta^{k+1}(\omega) \in B \text{ i.o. } k\right\} = A$$

and thus $A \in \mathcal{B}_\theta$. ■

**Definition 18.4.** *A $\mathcal{B}$ – measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff $f \circ \theta = f$ ($f \circ \theta = f$ a.s.).*

**Lemma 18.5.** *A $\mathcal{B}$ – measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff $f$ is $\mathcal{B}_\theta$ ($\mathcal{B}'_\theta$) measurable. Moreover, if $f$ is almost invariant, then there exists and invariant function, $g : \Omega \to \mathbb{R}$, such that $f = g$, $\mu$ – a.e. (This latter assertion has already been explained in Exercises 14.3 and 14.4.)*

**Proof.** If $f$ is invariant, $f \circ \theta = f$, then $\theta^{-1}(\{f \le x\}) = \{f \circ \theta \le x\} = \{f \le x\}$ which shows that $\{f \le x\} \in \mathcal{B}_\theta$ for all $x \in \mathbb{R}$ and therefore $f$ is $\mathcal{B}_\theta$ – measurable. Similarly if $f$ is almost invariant so that $f \circ \theta = f$ ($\mu$ – a.e.), then

$$\begin{aligned}\mu\left(\left|1_{\theta^{-1}(\{f\le x\})} - 1_{\{f\le x\}}\right|\right) &= \mu\left(\left|1_{\{f\circ\theta\le x\}} - 1_{\{f\le x\}}\right|\right)\\ &= \mu\left(\left|1_{(-\infty,x]}\circ f\circ\theta - 1_{(-\infty,x]}\circ f\right|\right) = 0\end{aligned}$$

from which it follows that $\{f \le x\} \in \mathcal{B}'_\theta$ for all $x \in \mathbb{R}$, that is $f$ is $\mathcal{B}'_\theta$ – measurable.

Conversely if $f : \Omega \to \mathbb{R}$ is ($\mathcal{B}'_\theta$) $\mathcal{B}_\theta$ -measurable, then for all $-\infty < a < b < \infty$, ($\{a < f \le b\} \in \mathcal{B}'_\theta$) $\{a < f \le b\} \in \mathcal{B}_\theta$ from which it follows that $1_{\{a<f\le b\}}$ is (almost) invariant. Thus for every $N \in \mathbb{N}$ the function defined by;

$$f_N := \sum_{n=-N^2}^{N^2}\frac{n}{N}1_{\left\{\frac{n-1}{N}<f\le\frac{n}{N}\right\}},$$

is (almost) invariant. As $f = \lim_{N\to\infty} f_N$, it follows that $f$ is (almost) invariant as well.

In the case where $f$ is almost invariant, we can choose $D_N(n) \in \mathcal{B}_\theta$ such that $\mu\left(D_N(n)\,\Delta\left\{\frac{n-1}{N} < f \le \frac{n}{N}\right\}\right) = 0$ for all $n$ and $N$ and then set

$$g_N := \sum_{n=-N^2}^{N^2}\frac{n}{N}1_{D_N(n)}.$$

We then have $g_N = f_N$ a.e. and $g_N$ is $\mathcal{B}_\theta$ – measurable. We may thus conclude that $\tilde{g} := \limsup_{N\to\infty} g_N$ is $\mathcal{B}_\theta$ – measurable. It now follows that $g := \tilde{g}1_{|\tilde{g}|<\infty}$ is $\mathcal{B}_\theta$ – measurable function such that $g = f$ a.e. ■

**Theorem 18.6.** *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a $\sigma$ – finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_*\mu = \mu$. Then;*

1. *$U : L^2(\mu) \to L^2(\mu)$ defined by $Uf := f \circ \theta$ is an isometry. The isometry $U$ is unitary if $\theta^{-1}$ exists as a measurable map.*
2. *The map,*
$$L^2(\Omega, \mathcal{B}_\theta, \mu) \ni f \to f \in \operatorname{Nul}(U - I)$$
*is unitary. In other words, $Uf = f$ iff there exists $g \in L^2(\Omega, \mathcal{B}_\theta, \mu)$ such that $f = g$ a.e.*

*3. For every $f \in L^2(\mu)$ we have,*

$$L^2(\mu) - \lim_{n\to\infty} \frac{f + f\circ\theta + \dots + f\circ\theta^{n-1}}{n} = \mathbb{E}_{\mathcal{B}_\theta}[f]$$

*where $\mathbb{E}_{\mathcal{B}_\theta}$ denotes orthogonal projection from $L^2(\Omega,\mathcal{B},\mu)$ onto $L^2(\Omega,\mathcal{B}_\theta,\mu)$, i.e. $\mathbb{E}_{\mathcal{B}_\theta}$ is conditional expectation.*

**Proof.** 1. To see that $U$ is an isometry observe that

$$\|Uf\|^2 = \int_\Omega |f\circ\theta|^2 d\mu = \int_\Omega |f|^2 d(\theta_*\mu) = \int_\Omega |f|^2 d\mu = \|f\|^2$$

for all $f \in L^2(\mu)$.

2. $f \in \mathrm{Nul}(U-I)$ iff $f\circ\theta = Uf = f$ a.e., i.e. iff $f$ is almost invariant. According to Lemma 18.5 this happen iff there exists a $\mathcal{B}_\theta$ – measurable function, $g$, such that $f = g$ a.e. Necessarily, $g \in L^2(\mu)$ so that $g \in L^2(\Omega,\mathcal{B}_\theta,\mu)$ as required.

3. The last assertion now follows from items 1. and 2. and the mean ergodic Theorem 18.1. ■

**Assumption 1** *From now on we will assume that $\mu = P$ is a probability measure such that $P\theta^{-1} = \theta$.*

**Exercise 18.1.** For every $Z \in L^1(P)$, show that $\mathbb{E}[Z\circ\theta|\mathcal{B}_\theta] = \mathbb{E}[Z|\mathcal{B}_\theta]$ a.s. More generally, show for sub – $\sigma$ -algebra, $\mathcal{G} \subset \mathcal{B}$, that

$$\mathbb{E}\left[Z\circ\theta|\theta^{-1}\mathcal{G}\right] = \mathbb{E}[Z|\mathcal{G}]\circ\theta \text{ a.s.}$$

**Exercise 18.2.** Let $1 \le p < \infty$. Following the ideas introduced in Exercises 16.9 and 16.10, show

$$L^p(P) - \lim_{n\to\infty} \frac{f + f\circ\theta + \dots + f\circ\theta^{n-1}}{n} = \mathbb{E}_{\mathcal{B}_\theta}[f] \text{ for all } f \in L^p(\Omega,\mathcal{B},P).$$

(Some of these ideas will again be used in the proof of Theorem 18.9 below.)

**Definition 18.7.** *A sequence of random variables $\xi = \{\xi_k\}_{k=1}^\infty$ is a* ***stationary*** *if $(\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$.*

If we temporarily let

$$\theta(x_1,x_2,x_3,\dots) = \theta(x_2,x_3,\dots) \text{ for } (x_1,x_2,x_3,\dots) \in \mathbb{R}^{\mathbb{N}}, \tag{18.2}$$

the stationarity condition states that $\theta\xi \overset{d}{=} \xi$. Equivalently if $\mu = \mathrm{Law}_P(\xi_1,\xi_2,\dots)$ on $\left(\mathbb{R}^{\mathbb{N}},\mathcal{B}_{\mathbb{R}}^{\otimes\mathbb{N}}\right)$, then $\xi = \{\xi_k\}_{k=1}^\infty$ is stationary iff $\mu\circ\theta^{-1} = \mu$. Let us also observe that $\xi$ is stationary implies $\theta^2\xi \overset{d}{=} \theta\xi \overset{d}{=} \xi$ and $\theta^3\xi \overset{d}{=} \theta\xi \overset{d}{=} \xi$, etc. so that $\theta^n\xi \overset{d}{=} \xi$ for all $n \in \mathbb{N}$.[1] In what follows for $x \in (x_1,x_2,x_3,\dots) \in \mathbb{R}^{\mathbb{N}}$ we will let $S_0(x) = 0$,

$$S_n(x) = x_1 + x_2 + \dots + x_n, \text{ and}$$
$$S_n^* := \max(S_1,S_2,\dots,S_n)$$

for all $n \in \mathbb{N}$.

**Lemma 18.8 (Maximal Ergodic Lemma).** *Suppose $\xi := \{\xi_k\}_{k=1}^\infty$ is a stationary sequence and $S_n(\xi) = \xi_1 + \dots + \xi_n$ as above, then*

$$\mathbb{E}\left[\xi_1 : \sup_n S_n(\xi) > 0\right] \ge 0. \tag{18.3}$$

**Proof.** In this proof, $\theta$ will be as in Eq. (18.2). If $1 \le k \le n$, then

$$S_k(\xi) = \xi_1 + S_{k-1}(\theta\xi) \le \xi_1 + S_{k-1}^*(\theta\xi) \le \xi_1 + S_n^*(\theta\xi) = \xi_1 + [S_n^*(\theta\xi)]_+$$

and therefore, $S_n^*(\xi) \le \xi_1 + [S_n^*(\theta\xi)]_+$. So we may conclude that

$$\begin{aligned}\mathbb{E}[\xi_1 : S_n^*(\xi) > 0] &\ge \mathbb{E}\left[S_n^*(\xi) - [S_n^*(\theta\xi)]_+ : S_n^* > 0\right] \\ &= \mathbb{E}\left[[S_n^*(\xi)]_+ - [S_n^*(\theta\xi)]_+ 1_{S_n^*>0}\right] \\ &\ge \mathbb{E}\left[[S_n^*(\xi)]_+ - [S_n^*(\theta\xi)]_+\right] = \mathbb{E}[S_n^*(\xi)]_+ - \mathbb{E}[S_n^*(\theta\xi)]_+ = 0,\end{aligned}$$

wherein we used $\xi \overset{d}{=} \theta\xi$ for the last equality. Letting $n \to \infty$ making use of the MCT and the observation that $\{S_n^*(\xi) > 0\} \uparrow \{\sup_n S_n(\xi) > 0\}$ gives Eq. (18.3). ■

**Theorem 18.9 (Birkoff's Ergodic Theorem).** *Suppose that $f \in L^1(\Omega,\mathcal{B},P)$ or $f \ge 0$ and is $\mathcal{B}$ – measurable, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1} = \mathbb{E}[f|\mathcal{B}_\theta] \text{ a.s.}. \tag{18.4}$$

*Moreover if $f \in L^p(\Omega,\mathcal{B},P)$ for some $1 \le p < \infty$ then the convergence in Eq. (18.4) holds in $L^p$ as well.*

[1] In other words if $\{\xi_k\}_{k=1}^\infty$ is stationary, then by lopping off the first random variable on each side of the identity, $(\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$, implies that

$$(\xi_3,\xi_4,\dots) \overset{d}{=} (\xi_2,\xi_3,,\dots) \overset{d}{=} (\xi_1,\xi_2,\dots).$$

Continuing this way inductively shows that stationarity is equivalent to $(\xi_n,\xi_{n+1},\xi_{n+2},\dots) \overset{d}{=} (\xi_1,\xi_2,\dots)$ for all $n \in \mathbb{N}$.

**Proof.** Let us begin with the general observation that if $\xi = (\xi_1, \xi_2, \dots)$ is a sequence of random variables such that $\xi_i \circ \theta = \xi_{i+1}$ for $i = 1, 2, \dots$, then $\xi$ is stationary. This is because,

$$(\xi_1, \xi_2, \dots) \stackrel{d}{=} (\xi_1, \xi_2, \dots) \circ \theta = (\xi_1 \circ \theta, \xi_2 \circ \theta, \dots) = (\xi_2, \xi_3, \dots).$$

We will first prove Eq. (18.4) under the assumption that $f \in L^1(P)$. We now let $g := \mathbb{E}[f|\mathcal{B}_\theta]$ and $\xi_k := f \circ \theta^{k-1} - g$ for all $k \in \mathbb{N}$. Since $g$ is $\mathcal{B}_\theta$ – measurable we know that $g \circ \theta = g$ and therefore,

$$\xi_k \circ \theta = \left(f \circ \theta^{k-1} - g\right) \circ \theta = f \circ \theta^k - g = \xi_{k+1}$$

and therefore $\xi = (\xi_1, \xi_2, \dots)$ is stationary. To simplify notation let us write $S_n$ for $S_n(\xi) = \xi_1 + \dots + \xi_n$. To finish the proof we need to show that $\lim_{n\to\infty} \frac{S_n}{n} = 0$ a.s. for then,

$$\frac{1}{n} \sum_{k=1}^n f \circ \theta^{k-1} = \frac{1}{n} S_n + g \to g = \mathbb{E}[f|\mathcal{B}_\theta] \text{ a.s.}$$

In order to show $\lim_{n\to\infty} \frac{S_n}{n} = 0$ a.s. it suffices to show

$$M(\xi) := \limsup_{n\to\infty} \frac{S_n(\xi)}{n} \le 0 \text{ a.s.}$$

If we can do this we can also show that $M(-\xi) = \limsup_{n\to\infty} \frac{-S_n(\xi)}{n} \le 0$, i.e.

$$\liminf_{n\to\infty} \frac{S_n(\xi)}{n} \ge 0 \ge \limsup_{n\to\infty} \frac{S_n(\xi)}{n} \text{ a.s.}$$

which shows that $\lim_{n\to\infty} \frac{S_n}{n} = 0$ a.s. Finally in order to prove $M(\xi) \le 0$ a.s. it suffices to show $P(M(\xi) > \varepsilon) = 0$ for all $\varepsilon > 0$. This is what we will do now.

Since $S_n \circ \theta = S_{n+1} - \xi_1$ we have so that

$$M(\xi) \circ \theta = \limsup_{n\to\infty} \frac{1}{n}(S_{n+1} - \xi_1) = \limsup_{n\to\infty} \left[\frac{n+1}{n} \cdot \frac{1}{n+1} S_{n+1}\right] = M(\xi).$$

Thus $M(\xi)$ is an invariant function and therefore $A_\varepsilon := \{M(\xi) > \varepsilon\} \in \mathcal{B}_\theta$. Using $\mathbb{E}[\xi_1|\mathcal{B}_\theta] = \mathbb{E}[f - g|\mathcal{B}_\theta] = g - g = 0$ a.s. it follows that

$$\begin{aligned} 0 &= \mathbb{E}[\mathbb{E}[\xi_1|\mathcal{B}_\theta] : M(\xi) > \varepsilon] = \mathbb{E}[\xi_1 : M(\xi - \varepsilon) > 0] \\ &= \mathbb{E}[\xi_1 - \varepsilon : M(\xi - \varepsilon) > 0] + \varepsilon P(A_\varepsilon). \end{aligned}$$

If we now define $\xi_n^\varepsilon := (\xi_n - \varepsilon) 1_{A_\varepsilon}$, which is still stationary since

$$\xi_n^\varepsilon \circ \theta = (\xi_n \circ \theta - \varepsilon) 1_{A_\varepsilon} \circ \theta = (\xi_{n+1} - \varepsilon) 1_{A_\varepsilon} = \xi_{n+1}^\varepsilon,$$

then it is easily verified[2] that

$$A_\varepsilon = \{M(\xi - \varepsilon) > 0\} = \left\{\sup_n S_n(\xi^\varepsilon) > 0\right\}.$$

Therefore by an application of the maximal ergodic Lemma 18.8 we have,

$$-\varepsilon P(M(\xi) > \varepsilon) = \mathbb{E}[\xi_1 - \varepsilon : A_\varepsilon] = \mathbb{E}\left[\xi_1^\varepsilon : \sup_n S_n(\xi^\varepsilon) > 0\right] \ge 0$$

which shows $P(M(\xi) > \varepsilon) = 0$.

Now suppose that $f \in L^p(P)$. To prove the $L^p$ – convergence of the limit in Eq. (18.4) it suffices by Theorem 14.51 to show $\left\{\left|\frac{1}{n} S_n(\eta)\right|^p\right\}_{n=1}^\infty$ is uniformly integrable. This can be done as in the second solution to Exercise 14.6 (Resnick § **6.7**, #5). Here are the details.

First observe that $\{|\eta_k|^p\}_{k=1}^\infty$ are uniformly integrable. Indeed, by stationarity,

$$\mathbb{E}[|\eta_k|^p : |\eta_k|^p \ge a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \ge a]$$

and therefore

$$\sup_k \mathbb{E}[|\eta_k|^p : |\eta_k|^p \ge a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \ge a] \stackrel{\text{DCT}}{\to} 0 \text{ as } a \to \infty.$$

Thus if $\varepsilon > 0$ is given we may find (see Proposition 14.48) $\delta > 0$ such that $\mathbb{E}[|\eta_k|^p : A] \le \varepsilon$ whenever $A \in \mathcal{B}$ with $P(A) \le \delta$. Then for such an $A$, we have (using Jensen's inequality relative to normalized counting measure on $\{1, 2, \dots, n\}$),

$$\mathbb{E}\left[\left|\frac{1}{n} S_n(\eta)\right|^p : A\right] \le \mathbb{E}\left[\frac{1}{n} S_n(|\eta|^p) : A\right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|\eta_k|^p : A] \le \frac{1}{n} n\varepsilon = \varepsilon.$$

Another application of Proposition 14.48 shows $\left\{\left|\frac{1}{n} S_n(\eta)\right|^p\right\}_{n=1}^\infty$ is uniformly integrable as

$$\sup_n \mathbb{E}\left|\frac{1}{n} S_n(\eta)\right|^p \le \sup_n \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|\eta_k|^p] = \mathbb{E}|\eta_1|^p < \infty.$$

[2] Since $A_\varepsilon \subset \{\sup S_n/n > \varepsilon\}$, it follows that

$$\begin{aligned} A_\varepsilon &= \left\{\sup \frac{S_n}{n} > \varepsilon\right\} \cap A_\varepsilon = \{\sup S_n - n\varepsilon > 0\} \cap A_\varepsilon \\ &= \{\sup S_n(\xi - \varepsilon) > 0\} \cap A_\varepsilon = \{\sup S_n(\xi - \varepsilon) 1_{A_\varepsilon} > 0\} \\ &= \left\{\sup_n S_n(\xi^\varepsilon) > 0\right\}. \end{aligned}$$

Finally we need to consider the case where $f \geq 0$ but $f \notin L^1(P)$. As before, let $g = \mathbb{E}[f|\mathcal{B}_\theta] \geq 0$. For $r \in (0,\infty)$ and let $f_r := f \cdot 1_{g \leq r}$. We then have

$$\mathbb{E}[f_r|\mathcal{B}_\theta] = \mathbb{E}[f \cdot 1_{g\leq r}|\mathcal{B}_\theta] = 1_{g\leq r}\mathbb{E}[f \cdot |\mathcal{B}_\theta] = 1_{g\leq r} \cdot g$$

and in particular, $\mathbb{E}f_r = \mathbb{E}(1_{g\leq r} g) \leq r < \infty$. Thus by the $L^1$ – case already proved,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f_r \circ \theta^{k-1} = 1_{g\leq r} \cdot g \text{ a.s.}$$

On the other hand, since $g$ is $\theta$ –invariant, we see that $f_r \circ \theta^k = f \circ \theta^k \cdot 1_{g\leq r}$ and therefore

$$\frac{1}{n}\sum_{k=1}^n f_r \circ \theta^{k-1} = \left(\frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1}\right) 1_{g\leq r}.$$

Using these identities and the fact that $r < \infty$ was arbitrary we may conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = g \text{ a.s. on } \{g < \infty\}. \tag{18.5}$$

To take care of the set where $\{g = \infty\}$, again let $r \in (0,\infty)$ but now take $f_r = f \wedge r \leq f$. It then follows that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \left[f_r \circ \theta^{k-1}\right] = \mathbb{E}[f \wedge r|\mathcal{B}_\theta].$$

Letting $r \uparrow \infty$ and using the cMCT implies,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} \geq \mathbb{E}[f|\mathcal{B}_\theta] = g$$

and therefore $\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = \infty$ a.s. on $\{g = \infty\}$. This then shows that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = \infty = g \text{ a.s. on } \{g = \infty\}.$$

which combined with Eq. (18.5) completes the proof. ■

As a corollary we have the following version of the strong law of large numbers, also see Theorems 22.31 and Example 20.82 below for other proofs.

**Theorem 18.10 (Kolmogorov's Strong Law of Large Numbers).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. If $X_n$ are integrable or $X_n \geq 0$, then*

$$\lim_{n\to\infty} \frac{1}{n} S_n = \mathbb{E}X_1 \text{ a.s.}$$

*and $\frac{1}{n}S_n \to \mathbb{E}X_1$ in $L^1(P)$ when $\mathbb{E}|X_n| < \infty$.*

**Proof.** We may assume that $\Omega = \mathbb{R}^{\mathbb{N}}$, $\mathcal{B}$ is the product $\sigma$ – algebra, and $P = \mu^{\otimes\mathbb{N}}$ where $\mu = \mathrm{Law}_P(X_1)$. In this model, $X_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and we take $\theta : \Omega \to \Omega$ as in Eq. (18.2). Wit this notation we have $X_n = X_1 \circ \theta^{n-1}$ and therefore, $S_n = \sum_{k=1}^n X_1 \circ \theta^{k-1}$. So by Birkoff's ergodic theorem $\lim_{n\to\infty} \frac{1}{n}S_n = \mathbb{E}[X_1|\mathcal{B}_\theta] =: g$ a.s.

If $A \in \mathcal{B}_\theta$, then $A = \theta^{-n}(A) \in \sigma(X_{n+1}, X_{n+2}, \dots)$ and therefore $A \in \mathcal{T} = \cap_n \sigma(X_{n+1}, X_{n+2}, \dots)$ – the tail $\sigma$ – algebra. However by Kolmogorov's 0 - 1 law (Proposition 12.53), we know that $\mathcal{T}$ is almost trivial and therefore so is $\mathcal{B}_\theta$. Hence we may conclude that $g = c$ a.s. where $c \in [0,\infty]$ is a constant, see Lemma 12.52.

If $X_1 \geq 0$ a.s. and $\mathbb{E}X_1 = \infty$ then we must $c = \mathbb{E}[X_1|\mathcal{B}_\theta] = \infty$ a.s. for if $c < \infty$, then $\mathbb{E}X_1 = \mathbb{E}[\mathbb{E}[X_1|\mathcal{B}_\theta]] = \mathbb{E}[c] < \infty$. When $X_1 \in L^1(P)$, the convergence in Birkoff's ergodic theorem is also in $L^1$ and therefore we may conclude that

$$c = \mathbb{E}c = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}S_n\right] = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[S_n] = \mathbb{E}X_1.$$

Thus we have shown in all cases that $\lim_{n\to\infty} \frac{1}{n}S_n = \mathbb{E}[X_1|\mathcal{B}_\theta] = \mathbb{E}X_1$ a.s. ■

# Part III

## Stochastic Processes I

In the sequel $(\Omega, \mathcal{B}, P)$ will be a probability space and $(S, \mathcal{S})$ will denote a measurable space which we refer to as **state space**. If we say that $f : \Omega \to S$ is a function we will always assume that it is $\mathcal{B}/\mathcal{S}$ – measurable. We also let $\mathcal{S}_b$ denote the bounded $\mathcal{S}/\mathcal{B}_{\mathbb{R}}$ – measurable functions from $S$ to $\mathbb{R}$. On occasion we will assume that $(S, \mathcal{S})$ is a standard Borel space in order to have available to us the existence of regular conditional distributions (see Remark 16.21 and Theorem 16.41) and the use of Kolmogorov's extension Theorem 19.68 for proving the existence of Markov processes.

In the rest of this book we will devote most of our time to studying **stochastic processes**, i.e. a collection of random variables or more generally random functions, $X := \{X_t : \Omega \to S\}_{t \in T}$, indexed by some parameter space, $T$. The weakest description of such a stochastic process will be through its "finite dimensional distributions."

**Definition 18.11.** *Given a stochastic process, $X := \{X_t : \Omega \to S\}_{t \in T}$, and a finite subset, $\Lambda \subset T$, we say that $\nu_\Lambda := \mathrm{Law}_P\left(\{X_t\}_{t \in \Lambda}\right)$ on $\left(S^\Lambda, \mathcal{S}^{\otimes \Lambda}\right)$ is a finite dimensional distribution of $X$.*

Unless $T$ is a countable or finite set or $X_t$ has some continuity properties in $t$, knowledge of the finite dimensional distributions alone is not going to be adequate for our purposes, however it is a starting point. For now we are going to restrict our attention to the case where $T = \mathbb{N}_0$ or $T = \mathbb{R}_+ := [0, \infty)$ ($t \in T$ is typically interpreted as a time). Later in this part we will further restrict attention to stochastic processes indexed by $\mathbb{N}_0$ leaving the technically more complicated case where $T = \mathbb{R}_+$ to later parts of the book.

**Definition 18.12.** *An increasing (i.e. non-decreasing) sequence $\{\mathcal{B}_t\}_{t \in T}$ of sub-$\sigma$-algebras of $\mathcal{B}$ is called a **filtration**. We will let $\mathcal{B}_\infty := \vee_{t \in T} \mathcal{B}_t := \sigma\left(\cup_{t \in T} \mathcal{B}_t\right)$. A four-tuple, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in T}, P\right)$, where $(\Omega, \mathcal{B}, P)$ is a probability space and $\{\mathcal{B}_t\}_{t \in T}$ is a filtration is called a **filtered probability space.** We say that a stochastic process, $\{X_t\}_{t \in T}$, of random functions from $\Omega \to S$ is **adapted** to the filtration if $X_t$ is $\mathcal{B}_t/\mathcal{S}$ – measurable for every $t \in T$.*

A typical way to make a filtration is to start with a stochastic process $\{X_t\}_{t \in T}$ and then define $\mathcal{B}_t^X := \sigma\left(X_s : s \le t\right)$. Clearly $\{X_t\}_{t \in T}$ will always be adapted to this filtration.

In this part of the book we are going to study stochastic processes with certain dependency structures. This will take us to the notion of Markov processes and martingales. Before starting our study of Markov processes it will be helpful to record a few more facts about probability kernels.

Given a probability kernel, $Q$, on $S \times S$ (so $Q : S \times \mathcal{S} \to [0, 1]$), we may associate a linear transformation, $T = T_Q : \mathcal{S}_b \to \mathcal{S}_b$ defined by

$$(Tf)(x) = Q(x, f) = \int_S Q(x, dy)\, f(y) \text{ for all } f \in \mathcal{S}_b. \tag{18.6}$$

It is easy to check that $T$ satisfies;

1. $T1 = 1$,
2. $Tf \ge 0$ if $0 \le f \in \mathcal{S}_b$,
3. if $f_n \in \mathcal{S}_b$ and $f_n \to f$ boundedly then $Tf_n \to Tf$ boundedly as well.

Notice that an operator $T : \mathcal{S}_b \to \mathcal{S}_b$ satisfying conditions 1. and 2. above also satisfies $Tf \le Tg$ and $Tf$ is real is $f$. Indeed if $f = f_+ - f_-$ is real then

$$Tf = T(f_+ - f_-) = Tf_+ - Tf_-$$

with $0 \le Tf_\pm \in \mathbb{R}$ and if $f \le g$ then $0 \le f - g$ which implies

$$Tf - Tg = T(f - g) \ge 0.$$

As $\pm f \le |f|$ when $f$ is real, we have $\pm Tf \le T|f|$ and therefore $|Tf| \le T|f|$. More generally if $f$ is complex and $x \in S$, we may choose $\theta \in \mathbb{R}$ such that $e^{i\theta}(Tf)(x) \ge 0$ and therefore,

$$\begin{aligned} |(Tf)(x)| &= e^{i\theta}(Tf)(x) = \left(T\left[e^{i\theta} f\right]\right)(x) \\ &= \left(T\,\mathrm{Re}\left[e^{i\theta} f\right]\right)(x) + i\left(T\,\mathrm{Im}\left[e^{i\theta} f\right]\right)(x). \end{aligned}$$

Furthermore we must have $\left(T\,\mathrm{Im}\left[e^{i\theta} f\right]\right)(x) = 0$ and using $\mathrm{Re}\left[e^{i\theta} f\right] \le |f|$ we find,

$$|(Tf)(x)| = \left(T\,\mathrm{Re}\left[e^{i\theta} f\right]\right)(x) \le (T|f|)(x).$$

As $x \in S$ was arbitrary we have shown that $|Tf| \le T|f|$. Thus if $|f| \le M$ for some $0 \le M < \infty$ we may conclude,

$$|Tf| \le T|f| \le T(M \cdot 1) = MT1 = M.$$

**Proposition 18.13.** *If $T : \mathcal{S}_b \to \mathcal{S}_b$ is a linear transformation satisfying the three properties listed after Eq. (18.6), then $Q(x, A) := (T1_A)(x)$ for all $A \in \mathcal{S}$ and $x \in S$ is a probability kernel such that Eq. (18.6) holds.*

The proof of this proposition is straightforward and will be left to the reader. Let me just remark that if $Q(x, A) := (T1_A)(x)$ for all $x \in S$ and $A \in \mathcal{S}$ then $Tf = Q(\cdot, f)$ for all simple functions in $\mathcal{S}_b$ and then by approximation for all $f \in \mathcal{S}_b$.

**Corollary 18.14.** *If $Q_1$ and $Q_2$ are two probability kernels on $(S, \mathcal{S}) \times (S, \mathcal{S})$, then $T_{Q_1} T_{Q_2} = T_Q$ where $Q$ is the probability kernel given by*

$$\begin{aligned} Q(x, A) &= \left(T_{Q_1} T_{Q_2} 1_A\right)(x) = Q_1\left(x, Q_2(\cdot, A)\right) \\ &= \int_S Q_1(x, dy)\, Q_2(y, A) \end{aligned}$$

*for all $A \in \mathcal{S}$ and $x \in S$. We will denote $Q$ by $Q_1 Q_2$.*

From now on we will identify the probability kernel $Q : S \times \mathcal{S} \to [0,1]$ with the linear transformation $T = T_Q$ and simply write $Qf$ for $Q(\cdot, f)$. The last construction that we need involving probability kernels is the following extension of the notion of product measure.

**Proposition 18.15.** *Suppose that $\nu$ is a probability measure on $(S, \mathcal{S})$ and $Q_k : S \times \mathcal{S} \to [0,1]$ are probability kernels on $(S, \mathcal{S}) \times (S, \mathcal{S})$ for $1 \leq k \leq n$. Then there exists a probability measure $\mu$ on $\left(S^{n+1}, \mathcal{S}^{\otimes(n+1)}\right)$ such that for all $f \in \mathcal{S}_b^{\otimes(n+1)}$ we have*

$$\mu(f) = \int_S d\nu(x_0) \int_S Q_1(x_0, dx_1) \int_S Q_2(x_1, dx_2) \cdot \\ \dots \cdot \int_S Q_n(x_{n-1}, dx_n) f(x_0, \dots, x_n). \quad (18.7)$$

*Part of the assertion here is that all functions appearing are bounded and measurable so that all of the above integrals make sense. We will denote $\mu$ in the future by,*

$$d\mu(x_0, \dots, x_n) = d\nu(x_0) Q_1(x_0, dx_1) Q_2(x_1, dx_2) \dots Q_n(x_{n-1}, dx_n).$$

**Proof.** The fact that all of the iterated integrals make sense in Eq. (18.7) follows from Exercise 16.5, the measurability statements in Fubini's theorem, and induction. The measure $\mu$ is defined by setting $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{S}^{\otimes(n+1)}$. It is a simple matter to check that $\mu$ is a measure on $\left(S^{n+1}, \mathcal{S}^{\otimes(n+1)}\right)$ and that $\int_S f d\mu$ agrees with the right side of Eq. (18.7) for all $f \in \mathcal{S}_b^{\otimes(n+1)}$. ■

*Remark 18.16.* As usual the measure $\mu$ is determined by its value on product functions of the form $f(x_0, \dots, x_n) = \prod_{i=0}^n f_i(x_i)$ with $f_i \in \mathcal{S}_b$. For such a function we have

$$\mu(f) = \mathbb{E}_\nu\left[f_0 Q_1 M_{f_1} Q_2 M_{f_2} \dots Q_{n-1} M_{f_{n-1}} Q_n f_n\right]$$

where $M_f : \mathcal{S}_b \to \mathcal{S}_b$ is defined by $M_f g = fg$, i.e. $M_f$ is multiplication by $f$.

# 19

# The Markov Property

For purposes of this section, $T = \mathbb{N}_0$ or $\mathbb{R}_+$, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in T}, P\right)$ is a filtered probability space, and $(S, \mathcal{S})$ be a measurable space. We will often write $t \geq 0$ to mean that $t \in T$. Thus we will often denote a stochastic process by $\{X_t\}_{t\geq 0}$ instead of $\{X_t\}_{t\in T}$.

**Definition 19.1 (The Markov Property).** *A stochastic process $\{X_t : \Omega \to S\}_{t\in T}$ is said to satisfy the **Markov property** if $X_t$ is adapted and*

$$\mathbb{E}_{\mathcal{B}_s} f(X_t) := \mathbb{E}[f(X_t)|\mathcal{B}_s] = \mathbb{E}[f(X_t)|X_s] \text{ a.s. for all } 0 \leq s < t \tag{19.1}$$

*and for every $f \in \mathcal{S}_b$.*

If Eq. (19.1) holds then by the factorization Lemma 8.40 there exists $F \in \mathcal{S}_b$ such that $F(X_s) = \mathbb{E}[f(X_t)|X_s]$. Conversely if we want to verify Eq. (19.1) it suffices to find an $F \in \mathcal{S}_b$ such that $\mathbb{E}_{\mathcal{B}_s} f(X_t) = F(X_s)$ a.s. This is because, by the tower property of conditional expectation,

$$\mathbb{E}_{\mathcal{B}_s} f(X_t) = F(X_s) = \mathbb{E}[F(X_s)|X_s] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_s} f(X_t)|X_s] = \mathbb{E}[f(X_t)|X_s] \text{ a.s.} \tag{19.2}$$

Poetically speaking as stochastic process with the Markov property is forgetful in the sense that knowing the positions of the process up to some time $s \leq t$ does not give any more information about the position of the process, $X_t$, at time $t$ than knowing where the process was at time $s$. We will in fact show (Theorem 19.4 below) that given $X_s$ what the process did before time $s$ is independent of the what it will do after time $s$.

**Lemma 19.2.** *If $\{X_t\}_{t\geq 0}$ satisfies the Markov property relative to the filtration $\{\mathcal{B}_t\}_{t\geq 0}$ it also satisfies the Markov property relative to $\left\{\mathcal{B}_t^X = \sigma(X_s : s \leq t)\right\}_{t\geq 0}$.*

**Proof.** It is clear that $\{X_t\}_{t\in T}$ is $\mathcal{B}_t^X$ – adapted and that $\sigma(X_s) \subset \mathcal{B}_s^X \subset \mathcal{B}_s$ for all $s \in T$. Therefore using the tower property of conditional expectation we have,

$$\mathbb{E}_{\mathcal{B}_s^X} f(X_t) = \mathbb{E}_{\mathcal{B}_s^X} \mathbb{E}_{\mathcal{B}_s} f(X_t) = \mathbb{E}_{\mathcal{B}_s^X} \mathbb{E}_{\sigma(X_s)} f(X_t) = \mathbb{E}_{\sigma(X_s)} f(X_t).$$

■

*Remark 19.3.* If $T = \mathbb{N}_0$, a stochastic process $\{X_n\}_{n\geq 0}$ is Markov iff for all $f \in \mathcal{S}_b$,

$$\mathbb{E}[f(X_{m+1})|\mathcal{B}_m] = \mathbb{E}[f(X_{m+1})|X_m] \text{ a.s. for all } m \geq 0 \tag{19.3}$$

Indeed if Eq. (19.3) holds for all $m$, we may use induction on $n$ to show

$$\mathbb{E}[f(X_n)|\mathcal{B}_m] = \mathbb{E}[f(X_n)|X_m] \text{ a.s. for all } n \geq m. \tag{19.4}$$

It is clear that Eq. (19.4) holds for $n = m$ and $n = m + 1$. So now suppose Eq. (19.4) holds for a given $n \geq m$. Using Eq. (19.3) with $m = n$ implies

$$\mathbb{E}_{\mathcal{B}_n} f(X_{n+1}) = \mathbb{E}[f(X_{n+1})|X_n] = F(X_n)$$

for some $F \in \mathcal{S}_b$. Thus by the tower property of conditional expectations and the induction hypothesis,

$$\begin{aligned} \mathbb{E}_{\mathcal{B}_m} f(X_{n+1}) &= \mathbb{E}_{\mathcal{B}_m} \mathbb{E}_{\mathcal{B}_n} f(X_{n+1}) = \mathbb{E}_{\mathcal{B}_m} F(X_n) = \mathbb{E}[F(X_n)|X_m] \\ &= \mathbb{E}[\mathbb{E}_{\mathcal{B}_n} f(X_{n+1})|X_m] = \mathbb{E}[f(X_{n+1})|X_m]. \end{aligned}$$

The next theorem and Exercise 19.1 shows that a stochastic process has the Markov property iff it has the property that; the future of time $s$ only depends on the past states $\{X_r : 0 \leq r \leq s\}$ through the state $(X_s)$ of the process at time $s$.

**Theorem 19.4 (Markov Independence).** *Suppose that $\{X_t\}_{t\in T}$ is an adapted stochastic process with the Markov property and let $\mathcal{F}_s := \sigma(X_t : t \geq s)$ be the future $\sigma$ – algebra. Then for any $s \in T$;*

1. *$\mathbb{E}[G|\mathcal{B}_s] = \mathbb{E}[G|X_s]$ for all $G \in (\mathcal{F}_s)_b$ and*
2. *$\mathcal{B}_s$ is independent[1] of $\mathcal{F}_s$ given $X_s$ which we abbreviate as $\mathcal{B}_s \underset{X_s}{\perp\!\!\!\perp} \mathcal{F}_s$. In more detail, we are asserting;*

$$P(A \cap B|X_s) = P(A|X_s) \cdot P(B|X_s) \text{ a.s.}$$

*for all $A \in \mathcal{B}_s$ and $B \in \mathcal{F}_s$ or equivalently that*

$$\mathbb{E}[FG|X_s] = \mathbb{E}[F|X_s] \cdot \mathbb{E}[G|X_s] \text{ a.s.} \tag{19.5}$$

*for all $F \in (\mathcal{B}_s)_b$ and $G \in (\mathcal{F}_s)_b$ and $s \in T$.*

[1] In words, the future and past are independent given the present.

**Proof.** We take each item in turn.

1. Suppose first that $G = \prod_{i=0}^{n} g_i(X_{t_i})$ with $s = t_0 < t_1 < t_2 < \cdots < t_n$ and $g_i \in \mathcal{S}_b$. Then by the Markov property and the tower property of conditional expectations,
$$\begin{aligned}\mathbb{E}_{\mathcal{B}_s}[G] &= \mathbb{E}_{\mathcal{B}_s}\left[\prod_{i=0}^{n} g_i(X_{t_i})\right] = \mathbb{E}_{\mathcal{B}_s}\mathbb{E}_{\mathcal{B}_{t_{n-1}}}\left[\prod_{i=0}^{n} g_i(X_{t_i})\right] \\ &= \mathbb{E}_{\mathcal{B}_s}\prod_{i=0}^{n-1} g_i(X_{t_i}) \cdot \mathbb{E}_{\mathcal{B}_{t_{n-1}}} g_n(X_{t_n}) \\ &= \mathbb{E}_{\mathcal{B}_s}\prod_{i=0}^{n-1} g_i(X_{t_i}) \cdot \mathbb{E}\left[g_n(X_{t_n})|X_{t_{n-1}}\right] = \mathbb{E}_{\mathcal{B}_s}\left[\tilde{G}\right]\end{aligned}$$
where
$$\tilde{G} := \left(\prod_{i=0}^{n-2} g_i(X_{t_i})\right) \cdot (g_{n-1}g)\left(X_{t_{n-1}}\right)$$
where $g \in \mathcal{S}_b$ is chosen so that $\mathbb{E}\left[g_n(X_{t_n})|X_{t_{n-1}}\right] = g\left(X_{t_{n-1}}\right)$ a.s.. Continuing this way inductively we learn that $\mathbb{E}_{\mathcal{B}_s}[G] = F(X_s)$ a.s. for some $F \in \mathcal{S}_b$ and therefore,
$$\mathbb{E}[G|X_s] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_s}G|X_s] = \mathbb{E}[F(X_s)|X_s] = F(X_s) = \mathbb{E}_{\mathcal{B}_s}[G] \text{ a.s.}$$
An application of the multiplicative system Theorem 10.2 may now be used to show item 1. holds for $G \in (\mathcal{F}_s)_b$
2. If $G \in (\mathcal{F}_s)$ and $F \in (\mathcal{B}_s)_b$, then by item 1. and the pull out property of conditional expectations,
$$\mathbb{E}_{\mathcal{B}_s}[FG] = F\mathbb{E}_{\mathcal{B}_s}[G] = F \cdot \mathbb{E}[G|X_s].$$
Applying $\mathbb{E}_{\sigma(X_s)}$ to this last equation while making use of the tower and pull out properties property of conditional expectations gives Eq. (19.5).

■

**Exercise 19.1.** Suppose that $\{X_t\}_{t\geq 0}$ is an adapted stochastic process such that Eq. (19.5) holds for all $F \in (\mathcal{B}_s)_b$ and $G \in (\mathcal{F}_s)_b$ and $s \in T$. Show that $\{X_t\}_{t\in T}$ has the Markov property.

## 19.1 Markov Processes

If $S$ is a standard Borel space (i.e. $S$ is isomorphic to a Borel subset of $[0,1]$), we may a find regular conditional probability kernels, $Q_{s,t}$ on $(S,\mathcal{S})\times(S,\mathcal{S}) \to [0,1]$ for all $0 \leq s < t$ such that
$$\mathbb{E}[f(X_t)|X_s] = Q_{s,t}(X_s; f) = (Q_{s,t}f)(X_s) \text{ a.s.} \tag{19.6}$$
Moreover by the Markov property if $0 \leq \sigma < s < t$, then
$$\begin{aligned}(Q_{\sigma,t}f)(X_\sigma) &= \mathbb{E}[f(X_t)|X_\sigma] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_s}f(X_t)|X_\sigma] \\ &= \mathbb{E}[(Q_{st}f)(X_s)|X_\sigma] = (Q_{\sigma,s}Q_{st}f)(X_\sigma) \;\; P-\text{a.s.} \\ &= Q_{\sigma,s}(X_\sigma; Q_{s,t}(\cdot; f)),\; P-\text{a.s.}\end{aligned}$$
If we let $\mu_t := \mathrm{Law}_P(X_t) : \mathcal{S} \to [0,1]$ for all $t \in T$, we have just shown that for every $f \in \mathcal{S}_b$, that
$$Q_{\sigma,t}(\cdot; f) = Q_{\sigma,s}(\cdot; Q_{s,t}(\cdot; f)) \qquad \mu_\sigma - \text{a.s.} \tag{19.7}$$
In the sequel we want to assume that such kernels exists and that Eq. (19.7) holds for everywhere not just $\mu_\sigma$ – a.s. Thus we make the following definitions.

**Definition 19.5 (Markov transition kernels).** *We say a collection of probability kernels, $\{Q_{s,t}\}_{0\leq s\leq t<\infty}$, on $S \times S$ are **Markov transition kernels** if $Q_{s,s}(x,dy) = \delta_x(dy)$ (as an operator $Q_{s,s} = I_{\mathcal{S}_b}$) for all $s \in T$ and the **Chapmann-Kolmogorov equations** hold;*
$$Q_{\sigma,t} = Q_{\sigma,s}Q_{s,t} \text{ for all } 0 \leq \sigma \leq s \leq t. \tag{19.8}$$

Recall that Eq. (19.8) is equivalent to
$$Q_{\sigma,t}(x,A) = \int_S Q_{\sigma,s}(x,dy)\,Q_{s,t}(y,A) \text{ for all } x \in S \text{ and } A \in \mathcal{S} \tag{19.9}$$
or
$$Q_{\sigma,t}(x; f) = Q_{\sigma,s}(x; Q_{s,t}(\cdot; f)) \text{ for all } x \in S \text{ and } f \in \mathcal{S}_b. \tag{19.10}$$
Thus Markov transition kernels should satisfy Eq. (19.7) everywhere not just almost everywhere.

The reader should keep in mind that $Q_{\sigma,t}(x,A)$ represents the jump probability of starting at $x$ at time $\sigma$ and ending up in $A \in \mathcal{S}$ at time $t$. With this in mind, $Q_{\sigma,s}(x,dy)\,Q_{s,t}(y,A)$ intuitively is the probability of jumping from $x$ at time $s$ to $y$ at time $t$ followed by a jump into $A$ at time $u$. Thus Eq. (19.9) states that averaging these probabilities over the intermediate location ($y$) of the particle at time $t$ gives the jump probability of starting at $x$ at time $s$ and ending up in $A \in \mathcal{S}$ at time $t$. This interpretation is rigorously true when $S$ is a finite or countable set.

**Definition 19.6 (Markov process).** *A **Markov process** is an adapted stochastic process, $\{X_t : \Omega \to S\}_{t\geq 0}$, with the Markov property such that there are Markov transition kernels $\{Q_{s,t}\}_{0\leq s\leq t<\infty}$, on $S \times S$ such that Eq. (19.6) holds.*

**Definition 19.7.** *A stochastic process, $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$, has **independent increments** if for all finite subsets, $\Lambda = \{0 \le t_0 < t_1 < \dots < t_n\} \subset T$ the random variables $\{X_0\} \cup \left\{X_{t_k} - X_{t_{k-1}}\right\}_{k=1}^n$ are independent. We refer to $X_t - X_s$ for $s < t$ as an **increment** of $X$.*

**Exercise 19.2.** Suppose that $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$ is a stochastic process with independent increments and let $\mathcal{B}_t := \mathcal{B}_t^X$ for all $t \in T$. Show, for all $0 \le s < t$, that $(X_t - X_s)$ is independent of $\mathcal{B}_s^X$ and then use this to show $\{X_t\}_{t\in T}$ is a Markov process with transition kernels defined by $0 \le s \le t$,

$$Q_{s,t}(x, A) := \mathbb{E}\left[1_A(x + X_t - X_s)\right] \text{ for all } A \in \mathcal{S} \text{ and } x \in \mathbb{R}^d. \tag{19.11}$$

You should verify that $\{Q_{s,t}\}_{0\le s\le t}$ are indeed Markov transition kernels, i.e. satisfy the Chapmann-Kolmogorov equations.

*Example 19.8 (Random Walks).* Suppose that $\left\{\xi_n : \Omega \to S := \mathbb{R}^d\right\}_{n=0}^\infty$ are independent random vectors and $X_m := \sum_{k=0}^m \xi_k$ and $\mathcal{B}_m := \sigma(\xi_0, \dots, \xi_m)$ for each $m \in T = \mathbb{N}_0$. Then $\{X_m\}_{m\ge 0}$ has independent increments and therefore has the Markov property with Markov transition kernels being given by

$$\begin{aligned} Q_{s,t}(x, f) &= \mathbb{E}\left[f(x + X_t - X_s)\right] \\ &= \mathbb{E}\left[f\left(x + \sum_{s<k\le t} \xi_k\right)\right] \end{aligned}$$

or in other words,

$$Q_{s,t}(x, \cdot) = \mathrm{Law}_P\left(x + \sum_{s<k\le t} \xi_k\right).$$

The one step transition kernels are determined by

$$(Q_{n,n+1} f)(x) = \mathbb{E}\left[f(x + \xi_{n+1})\right] \text{ for } n \in \mathbb{N}_0.$$

**Exercise 19.3.** Let us now suppose that $\{\xi_n : \Omega \to S\}_{n=0}^\infty$ are independent random functions where $(S, \mathcal{S})$ is a general a measurable space, $\mathcal{B}_n := \sigma(\xi_0, \xi_1, \dots, \xi_n)$ for $n \ge 0$, $u_n : S \times S \to S$ are measurable functions for $n \ge 1$, and $X_n : \Omega \to S$ for $n \in \mathbb{N}_0$ are defined by $X_0 = \xi_0$ and then inductively for $n \ge 1$ by

$$X_{n+1} = u_{n+1}(X_n, \xi_{n+1}) \text{ for } n \ge 0.$$

Convince yourself that for $0 \le m < n$ there is a measurable function, $\varphi_{n,m} : S^{n-m+1} \to S$ determined by the $\{u_k\}$ such that $X_n = \varphi_{n,m}(X_m, \xi_{m+1}, \dots, \xi_n)$. (You need not write the proof of this assertion in your solution.) In particular, $X_n = \varphi_{n,0}(\xi_0, \dots, \xi_n)$ is $\mathcal{B}_n/\mathcal{S}$ – measurable so that $X = \{X_n\}_{n\ge 0}$ is adapted. Show $\{X_n\}_{n\ge 0}$ is a Markov process with transition kernels,

$$Q_{m,n}(x, \cdot) = \mathrm{Law}_P\left(\varphi_{n,m}(x, \xi_{m+1}, \dots, \xi_n)\right) \text{ for all } 0 \le m \le n$$

where (by definition) $Q_{m,m}(x, \cdot) = \delta_x(\cdot)$. Please explicitly verify that $\{Q_{m,n}\}_{0\le m\le n}$ are Markov transition kernels, i.e. satisfy the Chapmann-Kolmogorov equations.

*Remark 19.9.* Suppose that $T = \mathbb{N}_0$ and $\{Q_{m,n} : 0 \le m \le n\}$ are Markov transition kernels on $S \times S$. Since

$$Q_{m,n} = Q_{m,m+1} Q_{m+1,m+2} \dots Q_{n-1,n}, \tag{19.12}$$

it follows that the $Q_{m,n}$ are uniquely determined by knowing the one step transition kernels, $\{Q_{n,n+1}\}_{n=0}^\infty$. Conversely if $\{Q_{n,n+1}\}_{n=0}^\infty$ are arbitrarily given probability kernels on $S \times S$ and $Q_{m,n}$ are defined as in Eq. (19.12), then the resulting $\{Q_{m,n} : 0 \le m \le n\}$ are Markov transition kernels on $S \times S$. Moreover if $S$ is a countable set, then we may let

$$q_{m,n}(x, y) := Q_{m,n}(x, \{y\}) = P(X_n = y | X_m = x) \text{ for all } x, y \in S \tag{19.13}$$

so that

$$Q_{m,n}(x, A) = \sum_{y\in A} q_{m,n}(x, y).$$

In this case it is easily checked that

$$\begin{aligned} &q_{m,n}(x, y) \\ &\quad = \sum_{x_i \in S : m < i < n} q_{m,m+1}(x, x_{m+1})\, q_{m+1,m+2}(x_{m+1}, x_{m+2}) \dots q_{n-1,n}(x_{n-1}, y). \end{aligned} \tag{19.14}$$

The reader should observe that this is simply matrix multiplication!

**Exercise 19.4 (Polya's Urn).** Suppose that an urn contains $r$ red balls and $g$ green balls. At each time $(t \in T = \mathbb{N}_0)$ we draw a ball out, then replace it and add $c$ more balls of the color drawn. It is reasonable to model this as a Markov process with $S := \mathbb{N}_0 \times \mathbb{N}_0$ and $X_n := (r_n, g_n) \in S$ being the number of red and green balls respectively in the urn at time $n$. Find

$$q_{n,n+1}((r, g), (r', g')) = P(X_{n+1} = (r', g') | X_n = (r, g))$$

for this model.

**Theorem 19.10 (Finite Dimensional Distributions).** *Suppose that* $X = \{X_t\}_{t\geq 0}$ *is a Markov process with Markov transition kernels* $\{Q_{s,t}\}_{0\leq s\leq t}$. *Further let* $\nu := \mathrm{Law}_P(X_0)$, *then for all* $0 = t_0 < t_1 < t_2 < \cdots < t_n$ *we have*

$$\mathrm{Law}_P(X_{t_0}, X_{t_1}, \dots, X_{t_n})(dx_0, dx_1, \dots, dx_n) = d\nu(x_0) \prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \tag{19.15}$$

*or equivalently,*

$$\mathbb{E}[f(X_{t_0}, X_{t_1}, \dots, X_{t_n})] = \int_{S^{n+1}} f(x_0, x_1, \dots, x_n)\, d\nu(x_0) \prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \tag{19.16}$$

*for all* $f \in \mathcal{S}_b^{\otimes(n+1)}$.

**Proof.** Because of the multiplicative system Theorem 10.2, it suffices to prove Eq. (19.16) for functions of the form $f(x_1, \dots, x_n) = \prod_{i=0}^n f_i(x_i)$ where $f_i \in \mathcal{S}_b$. The proof is now easily completed by induction on $n$. It is true for $n = 0$ by definition of $\nu$. Now assume it is true for some $n - 1 \geq 0$. We then have, making use of the inductive hypothesis, that

$$\begin{aligned}
&\mathbb{E}[f(X_{t_0}, X_{t_1}, \dots, X_{t_n})] \\
&= \mathbb{E}\,\mathbb{E}_{\mathcal{B}_{t_{n-1}}}\left[\prod_{i=0}^n f_i(X_{t_i})\right] \\
&= \mathbb{E}\left[Q_{t_{n-1},t_n}(X_{t_{n-1}}, f_n) \cdot \prod_{i=0}^{n-1} f_i(X_{t_i})\right] \\
&= \int_{S^n} Q_{t_{n-1},t_n}(x_{n-1}, f_n) \cdot \prod_{i=0}^{n-1} f_i(x_i)\, d\nu(x_0) \prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^n} \left[\int_S Q_{t_{n-1},t_n}(x_{n-1}, dx_n) f_n(x_n)\right] \cdot \prod_{i=0}^{n-1} f_i(x_i)\, d\nu(x_0) \prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^{n+1}} f(x_0, x_1, \dots, x_n)\, d\nu(x_0) \prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i)
\end{aligned}$$

as desired. ■

**Theorem 19.11 (Existence of Markov processes).** *Suppose that* $\{Q_{s,t}\}_{0\leq s\leq t}$ *are Markov transition kernels on a standard Borel space,* $(S, \mathcal{S})$. *Let* $\Omega := S^T$, $X_t : \Omega \to S$ *be the projection map,* $X_t(\omega) = \omega(t)$ *and* $\mathcal{B}_t = \mathcal{B}_t^X = \sigma(X_s : s \leq t)$ *for all* $t \in T$ *and* $\mathcal{B} := \mathcal{S}^{\otimes T} = \sigma(X_t : t \in T)$. *Then to each probability measure,* $\nu$, *on* $(S, \mathcal{S})$ *there exists a unique probability measure* $P_\nu$ *on* $(\Omega, \mathcal{B})$ *such that 1)* $\mathrm{Law}_{P_\nu}(X_0) = \nu$ *and 2)* $\{X_t\}_{t\geq 0}$ *is a Markov process having* $\{Q_{s,t}\}_{0\leq s\leq t}$ *as its Markov transition kernels.*

**Proof.** This is mainly an exercise in applying Kolmogorov's extension Theorem 19.68 as described in the appendix to this chapter. I will only briefly sketch the proof here.

For each $\Lambda = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$, let $P_\Lambda$ be the measure on $(S^{n+1}, \mathcal{S}^{\otimes(n+1)})$ defined by

$$dP_\Lambda(x_0, x_1, \dots, x_n) = d\nu(x_0) \prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i).$$

Using the Chapman-Kolmogorov equations one shows that the $\{P_\Lambda\}_{\Lambda\subset_f T}$ ($\Lambda \subset_f T$ denotes a finite subset of $T$) are consistently defined measures as described in the statement of Theorem 19.68. Therefore it follows by an application of that theorem that there exists a unique measure $P_\nu$ on $(\Omega, \mathcal{B})$ such that

$$\mathrm{Law}_{P_\nu}(X|_\Lambda) = P_\Lambda \text{ for all } \Lambda \subset_f T. \tag{19.17}$$

In light of Theorem 19.10, in order to finish the proof we need only show that $\{X_t\}_{t\geq 0}$ is a Markov process having $\{Q_{s,t}\}_{0\leq s\leq t}$ as its Markov transition kernels. Since if this is this case it finite dimensional distributions must be given as in Eq. (19.17) and therefore $P_\nu$ is uniquely determined. So let us now verify the desired Markov property.

Again let $\Lambda = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$ with $t_{n-1} = s < t = t_n$ and suppose that $f(x_0, \dots, x_n) = h(x_0, \dots, x_{n-1})\, g(x_n)$ with $h \in \mathcal{S}_b^{\otimes n}$ and $f \in \mathcal{S}_b$. By the definition of $P_\nu$ we then have (writing $\mathbb{E}_\nu$ for $\mathbb{E}_{P_\nu}$),

$$\begin{aligned}
&\mathbb{E}_\nu\left[h(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}})\, g(X_t)\right] \\
&= \int_{S^{n+1}} h(x_0, x_1, \dots, x_{n-1})\, g(x_n)\, d\nu(x_0) \prod_{i=1}^n Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \int_{S^n} h(x_0, x_1, \dots, x_{n-1})\, Q_{t_{n-1},t_n}(x_{n-1}, g)\, d\nu(x_0) \prod_{i=1}^{n-1} Q_{t_{i-1},t_i}(x_{i-1}, dx_i) \\
&= \mathbb{E}_\nu\left[h(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}})\, Q_{t_{n-1},t_n}(X_{t_{n-1}}, g)\right] \\
&= \mathbb{E}_\nu\left[h(X_{t_0}, X_{t_1}, \dots, X_{t_{n-1}})\, Q_{s,t}(X_s, g)\right]
\end{aligned}$$

It then follows by an application of the multiplicative system theorem that

$$\mathbb{E}_\nu[H g(X_t)] = \mathbb{E}_\nu[H Q_{s,t}(X_s, g)] \text{ for all } H \in (\mathcal{B}_s)_b$$

and therefore that

$$\mathbb{E}_\nu[g(X_t)|\mathcal{B}_s] = Q_{s,t}(X_s, g) \text{ a.s.}$$

■

We are now going to specialize to the more manageable class of "time homogeneous" Markov processes.

**Definition 19.12.** *We say that a collection of Markov transition kernels, $\{Q_{s,t}\}_{0\le s\le t}$ are* ***time homogeneous*** *if $Q_{s,t} = Q_{0,t-s}$ for all $0 \le s \le t$. In this case we usually let $Q_t := Q_{0,t-s}$. The condition that $Q_{s,s}(x,\cdot) = \delta_x$ now reduces to $Q_0(x,\cdot) = \delta_x$ and the Chapmann-Kolmogorov equations reduce to*

$$Q_sQ_t = Q_{s+t} \text{ for all } s,t \ge 0, \tag{19.18}$$

*i.e.*

$$\int_S Q_s(x,dy)\,Q_t(y,A) = Q_{s+t}(x,A) \text{ for all } s,t\ge 0,\ x\in S, \text{ and } A\in\mathcal{S}. \tag{19.19}$$

*A collection of operators $\{Q_t\}_{t\ge 0}$ with $Q_0 = Id$ satisfying Eq. (19.18) is called a* ***one parameter semi-group.***

**Definition 19.13.** *A Markov process is* ***time homogeneous*** *if it has time homogeneous Markov transition kernels. In this case we will have,*

$$\mathbb{E}[f(X_t)|\mathcal{B}_s] = Q_{t-s}(X_s,f) = (Q_{t-s}f)(X_s) \text{ a.s.} \tag{19.20}$$

*for all $0\le s\le t$ and $f\in\mathcal{S}_b$.*

**Theorem 19.14 (The time homogeneous Markov property).** *Suppose that $(S,\mathcal{S})$ is a measurable space, $Q_t : S\times\mathcal{S}\to[0,1]$ are time homogeneous Markov transition kernels, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0}\right)$ is a filtered measure space, $\{X_t:\Omega\to S\}_{t\ge0}$ are adapted functions, and for each $x\in S$ there exists a probability measure, $P_x$ on $(\Omega,\mathcal{B})$ such that;*

1. *$X_0(\omega) = x$ for $P_x$ – a.e. $\omega$ and*
2. *$\{X_t\}_{t\ge0}$ is a time homogeneous Markov process with transition kernels $\{Q_t\}_{t\ge0}$ relative to $P_x$.*

*Under these assumptions we have the following conclusions.*

1. *If $F\in\mathcal{S}_b^{\otimes T}$, then $S\ni x\to\mathbb{E}_xF(X)$ is $\mathcal{S}/\mathcal{B}_{\mathbb{R}}$ – measurable.*
2. *If $P$ is any probability measure on $(\Omega,\mathcal{B})$ such that $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0},\{X_t\}_{t\ge0},\{Q_t\}_{t\ge0},P\right)$ is a time homogeneous Markov process (with transition kernels being $\{Q_t\}$), then for all $t\ge0$,*

$$\mathbb{E}_P[F(X_{t+\cdot})|\mathcal{B}_t] = \mathbb{E}_P[F(X_{t+\cdot})|X_t] = \mathbb{E}_{X_t}[F(X)] \quad P\text{ – a.s.} \tag{19.21}$$

***Warning:*** *In this equation $\mathbb{E}_{X_t}$ does not denote $\mathbb{E}_{\sigma(X_t)} = \mathbb{E}[\cdot|X_t]$ but instead[2] it means the composition of $X_t$ with the function $S\ni x\to\mathbb{E}_x[F(X)]\in\mathbb{R}$. In more detail we are saying,*

$$\begin{aligned}\mathbb{E}_P[F(X_{t+\cdot})|\mathcal{B}_t](\omega) &= \mathbb{E}_{X_t(\omega)}[F(X)]\\ &= \int_\Omega F(X(\omega'))\;P_{X_t(\omega)}(d\omega').\end{aligned}$$

**Proof.** Let $F(X) := f(X_{t_0},\dots,X_{t_n})$ for some $f\in\mathcal{S}_b^{\otimes(n+1)}$.

1. For this $F$ we have
$$\mathbb{E}_x[F(X)] = \int Q_{0,t_1}(x,dx_1)\dots Q_{t_{n-1},t_n}(x_{n-1},dx_n)\,f(x_0,\dots,x_n)$$
which is $\mathcal{S}/\mathcal{B}_{\mathbb{R}}$ – measurable by a multiplicative systems theorem argument based on taking $f$ to be a product function. Another application of the multiplicative systems theorem then may be used to show that $x\to\mathbb{E}_x[F(X)]$ is measurable in general.
2. Since $F(X_{t+\cdot})\in(\mathcal{F}_t)_b$, Theorem 19.4 already implies
$$\mathbb{E}_P[F(X_{t+\cdot})|\mathcal{B}_t] = \mathbb{E}_P[F(X_{t+\cdot})|X_t] \quad P\text{ – a.s.}$$
To compute the last conditional expectation, let us first assume that $F(X) := f(X_{t_0},\dots,X_{t_n})$ and $g\in\mathcal{S}_b$, and let $\nu_t = \mathrm{Law}_P(X_t) = \nu$. Then using Theorem 19.10 twice we learn,
$$\begin{aligned}&\mathbb{E}_P[g(X_t)F(X_{t+\cdot})]\\ &\quad= \mathbb{E}_P[g(X_t)f(X_{t_0+t},\dots,X_{t_n+t})]\\ &\quad= \int g(x_0)f(x_0,\dots,x_n)\,d\nu_0(y)\,Q_t(y,dx_0)\prod_{j=1}^n Q_{t_j-t_{j-1}}(x_{j-1},dx_j)\\ &\quad= \int g(x_0)f(x_0,\dots,x_n)\,d\nu_t(x_0)\prod_{j=1}^n Q_{t_j-t_{j-1}}(x_{j-1},dx_j)\\ &\quad= \int d\nu_t(x_0)\,g(x_0)\,\mathbb{E}_{x_0}[f(X_{t_0},\dots,X_{t_n})]\\ &\quad= \int d\nu_t(x_0)\,g(x_0)\,\mathbb{E}_{x_0}F(X) = \mathbb{E}_P[g(X_t)\,\mathbb{E}_{X_t}F(X)].\end{aligned}$$

[2] Unfortunately we now have a lot of different meanings for $\mathbb{E}_\xi$ depending on what $\xi$ happens to be. So if $\xi = P$ is a measure then $\mathbb{E}_P$ stands for expectation relative to $P$. If $\xi = \mathcal{G}$ is a $\sigma$ – algebra it stands for conditional expectation relative to $\mathcal{G}$ and a given probability measure which not indicated in the notation. Finally if $x\in S$ we are writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$.

An application of the multiplicative systems Theorem 10.2 shows this equation is valid for all $F \in \mathcal{S}_b^{\otimes T}$ and this tells us that

$$\mathbb{E}_P \left[ F \left( X_{t+\cdot} \right) | X_t \right] = \mathbb{E}_{X_t} \left[ F \left( X \right) \right] \quad P_x - \text{a.s.}$$

for all $F \in \mathcal{S}_b^{\otimes T}$.

■

*Remark 19.15.* Admittedly Theorem 19.14 is a bit hard to parse on first reading. Therefore it is useful to rephrase what it says in the case that the state space, $S$, is finite or countable and $x \in S$ and $t > 0$ are such that $P(X_t = x) > 0$. Under these additional hypothesis we may combine Theorems 19.4 and 19.14 to find $\{X_s\}_{s \le t}$ and $\{X_s\}_{s \ge t}$ are $P(\cdot | X_t = x)$ – independent and moreover,

$$\operatorname{Law}_{P(\cdot | X_t = x)} \left( X_{t+\cdot} \right) = \operatorname{Law}_{P_x} \left( X_\cdot \right). \tag{19.22}$$

Last assertion simply states that given $X_t = x$ the process, $X$, after time $t$ behaves just like the process starting afresh from $x$.

## 19.2 Discrete Time Homogeneous Markov Processes

The proof of the following easy lemma is left to the reader.

**Lemma 19.16.** *If $Q_n : S \times \mathcal{S} \to [0, 1]$ for $n \in \mathbb{N}_0$ are time homogeneous Markov kernels then $Q_n = Q^n$ where $Q := Q_1$ and $Q^0 := I$. Conversely if $Q$ is a probability kernel on $S \times S$ then $Q_n := Q^n$ for $n \in \mathbb{N}_0$ are time homogeneous Markov kernels.*

*Example 19.17 (Random Walks Revisited).* Suppose that $\xi_0 : \Omega \to S := \mathbb{R}^d$ is independent of $\left\{ \xi_n : \Omega \to S := \mathbb{R}^d \right\}_{n=1}^\infty$ which are now assumed to be i.i.d. If $X_m = \sum_{k=0}^m \xi_k$ is as in Example 19.8, then $\{X_m\}_{m \ge 0}$ is a time homogeneous Markov process with

$$Q_m(x, \cdot) = \operatorname{Law}_P \left( X_m - X_0 \right)$$

and the one step transition kernel, $Q = Q_1$, is given by

$$Qf(x) = Q(x, f) = \mathbb{E}\left[ f(x + \xi_1) \right] = \int_S f(x + y)\, d\rho(y)$$

where $\rho := \operatorname{Law}_P(\xi_1)$. For example if $d = 1$ and $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q := 1 - p$ for some $0 \le p \le 1$, then we may take $S = \mathbb{Z}$ and we then have

$$Qf(x) = Q(x, f) = pf(x + 1) + qf(x - 1).$$

*Example 19.18 (Ehrenfest Urn Model).* Let a beaker filled with a particle fluid mixture be divided into two parts $A$ and $B$ by a semipermeable membrane. Let $X_n = (\#\text{ of particles in } A)$ which we assume evolves by choosing a particle at random from $A \cup B$ and then replacing this particle in the opposite bin from which it was found. Modeling $\{X_n\}$ as a Markov process we find,

$$P(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i-1, i+1\} \\ \frac{i}{N} & \text{if } \quad j = i - 1 \\ \frac{N-i}{N} & \text{if } \quad j = i + 1 \end{cases} =: q(i, j)$$

As these probabilities do not depend on $n$, $\{X_n\}$ is a time homogeneous Markov chain.

**Exercise 19.5.** Consider a rat in a maze consisting of 7 rooms which is laid out as in the following figure.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & & \end{bmatrix}$$

In this figure rooms are connected by either vertical or horizontal adjacent passages only, so that 1 is connected to 2 and 4 but not to 5 and 7 is only connected to 4. At each time $t \in \mathbb{N}_0$ the rat moves from her current room to one of the adjacent rooms with equal probability (the rat always changes rooms at each time step). Find the one step $7 \times 7$ transition matrix, $q$, with entries given by $q(i, j) := P(X_{n+1} = j | X_n = i)$, where $X_n$ denotes the room the rat is in at time $n$.

**Exercise 19.6 (2 - step MC).** Consider the following simple (i.e. no-brainer) two state "game" consisting of moving between two sites labeled 1 and 2. At each site you find a coin with sides labeled 1 and 2. The probability of flipping a 2 at site 1 is $a \in (0, 1)$ and a 1 at site 2 is $b \in (0, 1)$. If you are at site $i$ at time $n$, then you flip the coin at this site and move or stay at the current site as indicated by coin toss. We summarize this scheme by the "jump diagram" of Figure 19.1. It is reasonable to suppose that your location, $X_n$, at time $n$ is modeled by a

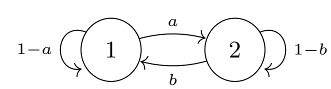



**Fig. 19.1.** The generic jump diagram for a two state Markov chain.

Markov process with state space, $S = \{1, 2\}$. Explain (briefly) why this is a time homogeneous chain and find the one step transition probabilities,

$$q(i,j) = P(X_{n+1} = j | X_n = i) \text{ for } i,j \in S.$$

Use your result and basic linear (matrix) algebra to compute, $\lim_{n\to\infty} P(X_n = 1)$. Your answer should be independent of the possible starting distributions, $\nu = (\nu_1, \nu_2)$ for $X_0$ where $\nu_i := P(X_0 = i)$.

The next exercise deals with how to describe a "lazy" Markov chain. We will say a chain is lazy if $q(x,x) > 0$ for some $x \in S$. The point being that if $q(x,x) > 0$, then the chain starting at $x$ may be lazy and stay at $x$ for some period of time before deciding to jump to a new site. The next exercise describes lazy chains in terms of a non-lazy chain and the random times that the lazy chain will spend lounging at each site $x \in S$. We will refer to this as the jump-hold description of the chain. We will give a similar description of chains on $S$ in the context of continuous time in Theorem 19.36 below.

**Exercise 19.7 (Jump - Hold Description I).** Let $S$ be a countable set $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P, \{Y_n\}_{n=0}^\infty)$ be a Markov chain with transition kernel, $\{q(x,y)\}_{x,y\in S}$ and let $\nu(x) := P(Y_0 = x)$ for all $x \in S$. For simplicity let us assume there are no **absorbing states**,[3] (i.e. $q(x,x) < 1$ for all $x \in S$) and then define,

$$\tilde{q}(x,y) := \begin{cases} \frac{q(x,y)}{1-q(x,x)} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}.$$

Let $\mathbf{j}_k$ denote the time of the $k^{\text{th}}$ – jump of the chain $\{Y_n\}_{n=0}^\infty$ so that

$$\mathbf{j}_1 := \inf\{n > 0 : Y_n \neq Y_0\} \text{ and}$$
$$\mathbf{j}_{k+1} := \inf\{n > \mathbf{j}_k : Y_n \neq Y_{\mathbf{j}_k}\}$$

with the convention that $\mathbf{j}_0 = 0$. Further let $\sigma_k := \mathbf{j}_k - \mathbf{j}_{k-1}$ denote the time spent between the $(k-1)^{\text{st}}$ and $k^{\text{th}}$ jump of the chain $\{Y_n\}_{n=0}^\infty$. Show;

1. For $\{x_k\}_{k=0}^n \subset S$ with $x_k \neq x_{k-1}$ for $k = 1, \dots, n$ and $m_1, \dots, m_k \in \mathbb{N}$, show

$$P\left([\cap_{k=0}^n \{Y_{\mathbf{j}_k} = x_k\}] \cap [\cap_{k=1}^n \{\sigma_k = m_k\}]\right)$$
$$= \nu(x_0) \prod_{k=1}^n q(x_{k-1}, x_{k-1})^{m_k - 1} (1 - q(x_{k-1}, x_{k-1})) \cdot \tilde{q}(x_{k-1}, x_k). \tag{19.23}$$

2. Summing the previous formula on $m_1, \dots, m_k \in \mathbb{N}$, conclude

$$P\left([\cap_{k=0}^n \{Y_{\mathbf{j}_k} = x_k\}]\right) = \nu(x_0) \cdot \prod_{k=1}^n \tilde{q}(x_{k-1}, x_k),$$

i.e. this shows $\{Y_{\mathbf{j}_k}\}_{k=0}^\infty$ is a Markov chain with transition kernel, $\tilde{q}$.

3. Conclude, relative to the conditional probability measure, $P(\cdot | [\cap_{k=0}^n \{Y_{\mathbf{j}_k} = x_k\}])$, that $\{\sigma_k\}_{k=1}^n$ are independent geometric $\sigma_k \overset{d}{=} Geo(1 - q(x_{k-1}, x_{k-1}))$ for $1 \leq k \leq n$, see Exercises 9.5 and 19.8.

[3] A state $x$ is absorbing if $q(x,x) = 1$ since in this case there is no chance for the chain to leave $x$ once it hits $x$.



**Exercise 19.8.** Let $\sigma$ be a geometric random variable with parameter $p \in (0,1]$, i.e. $P(\sigma = n) = (1-p)^{n-1} p$ for all $n \in \mathbb{N}$. Show, for all $n \in \mathbb{N}$ that

$$P(\sigma > n) = (1-p)^n \text{ for all } n \in \mathbb{N}$$

and then use this to conclude that

$$P(\sigma > m + n | \sigma > n) = P(\sigma > m) \ \forall\, m, n \in \mathbb{N}.$$

[This shows that the geometric distributions are the discrete analogue of the exponential distributions.]

## 19.3 Continuous time homogeneous Markov processes

An analogous (to Lemma 19.16) "infinitesimal description" of time homogeneous Markov kernels in the continuous time case can involve a considerable number of technicalities. Nevertheless, in this section we are going to ignore these difficulties in order to give a general impression of how the story goes. We will cover more precisely the missing details later.

So let $\{Q_t\}_{t\in\mathbb{R}_+}$ be time homogeneous collection of Markov transition kernels. We define the **infinitesimal generator** of $\{Q_t\}_{t\geq 0}$ by,

$$Af := \frac{d}{dt}|_{0^+} Q_t f = \lim_{t\downarrow 0} \frac{Q_t f - f}{t}. \tag{19.24}$$

For now we make the (often unreasonable assumption) that the limit in Eq. (19.24) holds for all $f \in \mathcal{S}_b$. This assumption is OK when $S$ is a finite (see Remark 19.19) or sometimes even when $S$ is a countable state space. For more complicated states spaces we will have to restrict the set of $f \in \mathcal{S}_b$ that we consider when computing $Af$ by Eq. ( 19.24). You should get a feeling for this issue by working through Exercise 19.10 which involves "Brownian motion."

*Remark 19.19 (L. Gårding's trick).* If $Q_t$ is a Markov-Semi group on $S = \{1, 2, \dots, n\}$ depending continuously on $t$ then for $\varepsilon > 0$ sufficiently small,

$$B_\varepsilon := \frac{1}{\varepsilon}\int_0^\varepsilon Q_s ds$$

is invertible. This is because

$$\|(I - B_\varepsilon)\| = \left\|\frac{1}{\varepsilon}\int_0^\varepsilon (I - Q_s)\, ds\right\| \le \frac{1}{\varepsilon}\int_0^\varepsilon \|I - Q_s\|\, ds \to 0 \text{ as } \varepsilon \downarrow 0$$

and as is well know $B_\varepsilon = I - (I - B_\varepsilon)$ is invertible provided $\|I - B_\varepsilon\| < 1$. Since

$$Q_t B_\varepsilon = \frac{1}{\varepsilon}\int_0^\varepsilon Q_t Q_s ds = \frac{1}{\varepsilon}\int_0^\varepsilon Q_{t+s} ds = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} Q_\tau d\tau$$

we may conclude by the fundamental theorem of calculus that

$$\frac{d}{dt} Q_t B_\varepsilon = \frac{1}{\varepsilon}\left[Q_{t+\varepsilon} - Q_t\right].$$

This shows $Q_t B_\varepsilon$ is differentiable and hence $Q_t = [Q_t B_\varepsilon] B_\varepsilon^{-1}$ is also differentiable in $t$.

We now just assume that $\lim_{t\downarrow 0} Q_t f = f$ and $\frac{d}{dt}|_{0^+} Q_t f$ exists. Under these assumptions, if $t, h > 0$, then using the semi-group property we have

$$Q_{t+h} - Q_t = (Q_h - I)\, Q_t = Q_t\,(Q_h - I). \tag{19.25}$$

Therefore,

$$Q_{t+h} f - Q_t f = (Q_h - I)\, Q_t f \to 0 \text{ as } h \downarrow 0$$

so that $Q_t f$ is right continuous. Similarly,

$$Q_{t-h} - Q_t = -(Q_h - I)\, Q_{t-h} = -Q_{t-h}\,(Q_h - I) \tag{19.26}$$

and

$$|Q_{t-h} f - Q_t f| = |Q_{t-h}\,(Q_h - I)\, f| \le |Q_{t-h}\,|(Q_h - I)\, f|| \le \sup_S |(Q_h - I)\, f|$$

which will tend to zero as $h \downarrow 0$ provided $Q_h f \to f$ uniformly (another fantasy in general). With this as "justification" we will assume that $t \to Q_t f$ is continuous in $t$.

Taking Eq. (19.25) divided by $h$ and Eq. (19.26) divided by $-h$ and then letting $h \downarrow 0$ implies,

$$\left(\frac{d}{dt}\right)_+ Q_t f = A Q_t f = Q_t A f$$

and

$$\left(\frac{d}{dt}\right)_- Q_t = A Q_t = Q_t A.$$

where $\left(\frac{d}{dt}\right)_+$ and $\left(\frac{d}{dt}\right)_-$ denote the right and left derivatives at $t$. So in principle we can expect that $\{Q_t\}_{t\ge 0}$ is uniquely determined by its infinitesimal generator $A$ by solving the differential equation,

$$\frac{d}{dt} Q_t = A Q_t = Q_t A \text{ with } Q_0 = Id. \tag{19.27}$$

Assuming all of this works out as sketched, it is now reasonable to denote $Q_t$ by $e^{tA}$. Let us now give a few examples to illustrate the discussion above.

*Example 19.20.* Suppose that $S = \{1, 2, \dots, n\}$ and $Q_t$ is a Markov-semi-group with infinitesimal generator, $A$, so that $\frac{d}{dt} Q_t = A Q_t = Q_t A$. By assumption $Q_t\,(i,j) \ge 0$ for all $i, j \in S$ and $\sum_{j=1}^n Q_t\,(i,j) = 1$ for all $i \in S$. We may write this last condition as $Q_t \mathbf{1} = \mathbf{1}$ for all $t \ge 0$ where $\mathbf{1}$ denotes the vector in $\mathbb{R}^n$ with all entries being 1. Differentiating $Q_t \mathbf{1} = \mathbf{1}$ at $t = 0$ shows that $A\mathbf{1} = 0$, i.e. $\sum_{j=1}^n A_{ij} = 0$ for all $i \in S$. Since

$$A_{ij} = \lim_{t\downarrow 0} \frac{Q_t\,(i,j) - \delta_{ij}}{t}$$

if $i \ne j$ we will have,

$$A_{ij} = \lim_{t\downarrow 0} \frac{Q_t\,(i,j)}{t} \ge 0.$$

Thus we have shown the infinitesimal generator, $A$, of $Q_t$ must satisfy $A_{ij} \ge 0$ for all $i \ne j$ and $\sum_{j=1}^n A_{ij} = 0$ for all $i \in S$. In words, $A$ is an $n \times n$ – matrix with non-negative off diagonal entries with all row sums being zero. You are asked to prove the converse in Exercise 19.9. So an explicit example of an infinitesimal generator when $S = \{1, 2, 3\}$ is

$$A = \begin{pmatrix} -3 & 1 & 2 \\ 4 & -6 & 2 \\ 7 & 1 & -8 \end{pmatrix}.$$

**Exercise 19.9.** Suppose that $S = \{1, 2, \dots, n\}$ and $A$ is a matrix such that $A_{ij} \ge 0$ for $i \ne j$ and $\sum_{j=1}^n A_{ij} = 0$ for all $i$. Show

$$Q_t = e^{tA} := \sum_{n=0}^\infty \frac{t^n}{n!} A^n \tag{19.28}$$

is a time homogeneous Markov kernel.

**Hints:** 1. To show $Q_t\,(i,j) \ge 0$ for all $t \ge 0$ and $i, j \in S$, write $Q_t = e^{-t\lambda} e^{t(\lambda I + A)}$ where $\lambda > 0$ is chosen so that $\lambda I + A$ has only non-negative entries. 2. To show $\sum_{j\in S} Q_t\,(i,j) = 1$, compute $\frac{d}{dt} Q_t \mathbf{1}$.

**Theorem 19.21 (Feynmann-Kac Formula).** *Continue the notation in Exercise 19.9 and let* $\left(\Omega, \{\mathcal{B}_t\}_{t\geq 0}, P_x, \{X_t\}_{t\geq 0}\right)$ *be a time homogeneous Markov process (assumed to be right continuous) with transition kernels,* $\{Q_t\}_{t\geq 0}$. *Given* $V : S \to \mathbb{R}$, *let* $T_t := T_t^V$ *be defined by*

$$(T_t g)(x) = \mathbb{E}_x\left[\exp\left(\int_0^t V(X_s)\,ds\right) g(X_t)\right] \tag{19.29}$$

*for all* $g : S \to \mathbb{R}$. *Then* $T_t$ *satisfies,*

$$\frac{d}{dt} T_t = T_t (A + M_V) \ \text{ with } T_0 = I \tag{19.30}$$

*where* $M_V g := Vg$ *for all* $g : S \to \mathbb{R}$, *i.e.* $M_V$ *is the diagonal matrix with* $V(1), \dots, V(n)$ *being placed in the diagonal entries. We may summarize this result as,*

$$\mathbb{E}_x\left[\exp\left(\int_0^t V(X_s)\,ds\right) g(X_t)\right] = \left(e^{t(A+M_V)} g\right)(x). \tag{19.31}$$

**Proof.** To see what is going on let us first assume that $\frac{d}{dt} T_t g$ exists in which case we may compute it as, $\frac{d}{dt}(T_t g)(x) = \frac{d}{dh}|_{0^+}(T_{t+h} g)(x)$. Then by the chain rule, the fundamental theorem of calculus, and the Markov property, we find

$$\begin{aligned}
\frac{d}{dt}(T_t g)(x) &= \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t+h} V(X_s)\,ds\right) g(X_t)\right] \\
&+ \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t} V(X_s)\,ds\right) g(X_{t+h})\right] \\
&= \mathbb{E}_x\left[\frac{d}{dh}|_{0^+}\exp\left(\int_0^{t+h} V(X_s)\,ds\right) g(X_t)\right] \\
&+ \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t} V(X_s)\,ds\right) \left(e^{hA} g\right)(X_t)\right] \\
&= \mathbb{E}_x\left[\exp\left(\int_0^{t} V(X_s)\,ds\right) V(X_t)\, g(X_t)\right] \\
&+ \mathbb{E}_x\left[\exp\left(\int_0^{t} V(X_s)\,ds\right) (Ag)(X_t)\right] \\
&= T_t(Vg)(x) + T_t(Ag)(x)
\end{aligned}$$

which gives Eq. (19.30). [It should be clear that $(T_0 g)(x) = g(x)$.]

We now give a rigorous proof. For $0 \leq \tau \leq t < \infty$, let $Z_{\tau,t} := \exp\left(\int_\tau^t V(X_s)\,ds\right)$ and let $Z_t := Z_{0,t}$. For $h > 0$,

$$\begin{aligned}
(T_{t+h} g)(x) &= \mathbb{E}_x\left[Z_{t+h}\, g(X_{t+h})\right] = \mathbb{E}_x\left[Z_t Z_{t,t+h}\, g(X_{t+h})\right] \\
&= \mathbb{E}_x\left[Z_t\, g(X_{t+h})\right] + \mathbb{E}_x\left[Z_t\left[Z_{t,t+h} - 1\right] g(X_{t+h})\right] \\
&= \mathbb{E}_x\left[Z_t\left(e^{hA} g\right)(X_t)\right] + \mathbb{E}_x\left[Z_t\left[Z_{t,t+h} - 1\right] g(X_{t+h})\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{(T_{t+h} g)(x) - (T_t g)(x)}{h} &= \mathbb{E}_x\left[Z_t \frac{\left(e^{hA} g\right)(X_t) - g(X_t)}{h}\right] \\
&+ \mathbb{E}_x\left[Z_t\left[\frac{Z_{t,t+h} - 1}{h}\right] g(X_{t+h})\right]
\end{aligned}$$

and then letting $h \downarrow 0$ in this equation implies,

$$\frac{d}{dh}|_{0^+}(T_{t+h} g)(x) = \mathbb{E}_x\left[Z_t (Ag)(X_t)\right] + \mathbb{E}_x\left[Z_t V(X_t)\, g(X_{t+h})\right].$$

This shows that $T_t$ is one sided differentiable and this one sided derivatives is given as in Eq. (19.30).

On the other hand for $s, t > 0$, using Theorem 19.14,

$$\begin{aligned}
T_{t+s} g(x) &= \mathbb{E}_x\left[Z_{t+s}\, g(X_{t+s})\right] = \mathbb{E}_x\left[Z_t Z_{t,t+s}\, g(X_{t+s})\right] \\
&= \mathbb{E}_x\left[Z_t \mathbb{E}_{\mathcal{B}_t}\left(Z_{t,t+s}\, g(X_{t+s})\right)\right] = \mathbb{E}_x\left[Z_t \mathbb{E}_{X_t}\left(Z_{0,s}\, g(X_s)\right)\right] \\
&= (T_t T_s g)(x),
\end{aligned}$$

i.e. $\{T_t\}_{t>0}$ still has the semi-group property. So for $h > 0$,

$$T_{t-h} - T_t = T_{t-h} - T_{t-h} T_h = T_{t-h}(I - T_h)$$

and hence

$$\frac{T_{t-h} - T_t}{-h} = T_{t-h} \frac{T_h - I}{h} \to T_t (A + V) \ \text{ as } h \downarrow 0$$

using $T_t$ is continuous in $t$ and the result we have already proved. This shows $T_t$ is differentiable in $t$ and Eq. (19.30) is valid. ∎

*Example 19.22 (Poisson Process).* By Exercise 19.2, it follows that Poisson process, $\{N_t \in S := \mathbb{N}_0\}_{t\geq 0}$ with intensity $\lambda$ has the Markov property. For all $0 \leq s \leq t$ we have,

$$\begin{aligned}
P(N_t = y | N_s = x) &= P(N_s + N_t - N_s = y | N_s = x) \\
&= P(N_s + N_t - N_s = y | N_s = x) \\
&= P(N_t - N_s = y - x | N_s = x) \\
&= 1_{y\geq x} \frac{(\lambda(t-s))^{y-x}}{(y-x)!} e^{-\lambda(t-s)} =: q_{t-s}(x, y).
\end{aligned}$$

With this notation it follows that

$$P(f(N_t)|N_s) = (Q_{t-s}f)(N_s)$$

where

$$\begin{aligned}Q_t f(x) &= \sum_{y\in S} q_t(x,y) f(y) \\ &= \sum_{y\in S} 1_{y\ge x}\frac{(\lambda t)^{y-x}}{(y-x)!}e^{-\lambda t} f(y) \\ &= \sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-\lambda t} f(x+n). \qquad (19.32)\end{aligned}$$

In particular $\{N_t\}_{t\ge0}$ is a time homogeneous Markov process. It is easy (but technically unnecessary) to directly verify the semi-group property;

$$(q_t q_s)(x,z) := \sum_{y\in S} q_t(x,y) q_s(y,z) = q_{s+t}(x,z). \qquad (19.33)$$

This can be done using the binomial theorem as follows;

$$\begin{aligned}\sum_{y\in S} q_t(x,y) q_s(y,z) &= \sum_{z\in S} 1_{y\ge x}\frac{(\lambda t)^{y-x}}{(y-x)!}e^{-\lambda t}\cdot 1_{z\ge y}\frac{(\lambda s)^{z-y}}{(z-y)!}e^{-\lambda s} \\ &= \sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-\lambda t}\cdot 1_{z\ge x+n}\frac{(\lambda s)^{z-x-n}}{(z-x-n)!}e^{-\lambda s} \\ &= 1_{z\ge x}e^{-\lambda(t+s)}\sum_{n=0}^{z-x}\frac{(\lambda t)^n}{n!}\frac{(\lambda s)^{z-x-n}}{(z-x-n)!} \\ &= 1_{z\ge x}e^{-\lambda(t+s)}\frac{(\lambda(t+s))^{z-x}}{(z-x)!} = q_{s+t}(x,z).\end{aligned}$$

To identify infinitesimal generator, $A = \frac{d}{dt}|_{0^+}Q_t$, in this example observe that

$$\begin{aligned}\frac{d}{dt}Q_t f(x) &= \frac{d}{dt}\left[e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}f(x+n)\right] \\ &= -\lambda Q_t f(x) + \lambda e^{-\lambda t}\sum_{n=0}^{\infty}\frac{n(\lambda t)^{n-1}}{n!}f(x+n) \\ &= -\lambda Q_t f(x) + \lambda e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}f(x+n+1) \\ &= -\lambda Q_t f(x) + \lambda (Q_t f)(x+1) \\ &= Q_t[\lambda f(\cdot+1) - \lambda f(\cdot)](x)\end{aligned}$$

and hence

$$Af(x) = \lambda(f(x+1) - f(x)).$$

Finally let us try to solve Eq. (19.27) in order to recover $Q_t$ from $A$. Formally we can hope that $Q_t = e^{tA}$ where $e^{tA}$ is given as its power series expansion. To simplify the computation it convenient to write $A = \lambda(T - I)$ where $If = f$ and $Tf = f(\cdot+1)$. Since $I$ and $T$ commute we further expect

$$e^{tA} = e^{\lambda t(T-I)} = e^{-\lambda t I}e^{\lambda t T} = e^{-\lambda t}e^{\lambda t T}$$

where

$$\begin{aligned}\left(e^{\lambda t T}f\right)(x) &= \sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}(T^n f)(x) \\ &= \sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}f(x+n).\end{aligned}$$

Putting this all together we find

$$\left(e^{tA}f\right)(x) = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}f(x+n)$$

which is indeed in agreement with $Q_t f(x)$ as we saw in Eq. (19.32).

*Remark 19.23 (Convolution Semi-Groups).* Here is an alternative explanation of Eq. (19.33). Let $\mu_t := \sum_{n\in\mathbb{N}_0}\frac{(\lambda t)^n}{n!}e^{-\lambda t}\delta_n$ and notice that

$$\tilde{\mu}_t(z) := \sum_{n\in\mathbb{Z}} z^n \mu_t(\{n\}) = \sum_{n=0}^{\infty} z^n\frac{(\lambda t)^n}{n!}e^{-\lambda t} = e^{\lambda t(1-z)}$$

and

$$\begin{aligned}\widetilde{\mu_t * \mu_s}(z) &= \sum_{n=0}^{\infty} z^n \mu_t * \mu_s(\{n\}) = \sum_{n=0}^{\infty} z^n \sum_{k=0}^{\infty}\mu_t(n-k)\mu_s(k) \\ &= \sum_{0\le k\le n<\infty} z^{n-k}\mu_t(n-k) z^k \mu_s(k) = \sum_{0\le k<\infty}\tilde{\mu}_t(z) z^k\mu_s(k) \\ &= \tilde{\mu}_t(z)\tilde{\mu}_s(z) = e^{\lambda t(1-z)}e^{\lambda s(1-z)} = e^{\lambda(s+t)(1-z)} \\ &= \tilde{\mu}_{s+t}(z)\end{aligned}$$

from which it follows that $\{\mu_t\}_{t>0}$ is a convolution semi - group, i.e. $\mu_t * \mu_s = \mu_{s+t}$ for all $s, t > 0$. Using

$$(q_t f)(x) = \sum_{n=0}^{\infty} \mu_t(n) f(x+n) = \sum_{n \in \mathbb{Z}} \mu_t(n) f(x+n),$$

it follows that

$$\begin{aligned}(q_s q_t f)(x) &= \sum_{k \in \mathbb{Z}} \mu_s(k) \sum_{n \in \mathbb{Z}} \mu_t(n) f(x+k+n) \\ &= \sum_{k \in \mathbb{Z}} \mu_s(k-n) \sum_{n \in \mathbb{Z}} \mu_t(n) f(x+k) \\ &= \sum_{k \in \mathbb{Z}} (\mu_s * \mu_t)(k) f(x+k) = \sum_{k} \mu_{s+t}(k) f(x+k) \\ &= (q_{s+t} f)(x).\end{aligned}$$

**Definition 19.24 (Brownian Motion).** *Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+}, P\right)$ be a filtered probability space. A real valued adapted process, $\{B_t : \Omega \to S = \mathbb{R}\}_{t \in \mathbb{R}_+}$, is called a **Brownian motion** if;*

1. *$\{B_t\}_{t \in \mathbb{R}_+}$ has independent increments with increments $B_t - B_s$ being independnet of $\mathcal{B}_s$ for all $0 \le s < t < \infty$.*
2. *for $0 \le s < t$, $B_t - B_s \overset{d}{=} N(0, t-s)$, i.e. $B_t - B_s$ is a normal mean zero random variable with variance $(t-s)$,*
3. *$t \to B_t(\omega)$ is continuous for all $\omega \in \Omega$.*

**Exercise 19.10 (Brownian Motion).** Assuming a Brownian motion $\{B_t\}_{t \ge 0}$ exists as described in Definition 19.24 show;

1. The process is a time homogeneous Markov process with transition kernels given by;
$$Q_t(x, dy) = q_t(x, y)\, dy \tag{19.34}$$
where
$$q_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2}. \tag{19.35}$$
2. Show by direct computation that $Q_t Q_s = Q_{t+s}$ for all $s, t > 0$. **Hint:** one of the many ways to do this it to use basic facts you have already proved about sums of independent Gaussian random variables along with the identity,
$$(Q_t f)(x) = \mathbb{E}\left[f\left(x + \sqrt{t} Z\right)\right],$$
where $Z \overset{d}{=} N(0, 1)$.
3. Show by direct computation that $q_t(x, y)$ satisfies the **heat equation**,
$$\frac{d}{dt} q_t(x, y) = \frac{1}{2} \frac{d^2}{dx^2} q_t(x, y) = \frac{1}{2} \frac{d^2}{dy^2} q_t(x, y) \text{ for } t > 0.$$
4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function with compact support. Show
$$\frac{d}{dt} Q_t f = A Q_t f = Q_t A f \text{ for all } t > 0,$$
where
$$A f(x) = \frac{1}{2} f''(x).$$

By combining Exercise 19.10 with Theorem 19.11 proves the following corollary.

**Corollary 19.25.** *There exits a Markov process $\{B_t\}_{t \ge 0}$ satisfying properties 1. and 2. of Definition 19.24.*

To get the path continuity property of Brownian motion requires additional arguments which we will do in a number of ways later, see Theorems 28.3, 28.7, and **??**. Modulo technical details, Exercise 19.10 shows that $A = \frac{1}{2}\frac{d^2}{dx^2}$ is the infinitesimal generator of Brownian motion, i.e. of $Q_t$ in Eqs. (19.34) and (19.35). The technical details we have ignored involve the proper function spaces in which to carry out these computations along with a proper description of the domain of the operator $A$. We will have to postpone these somewhat delicate issues until later. By the way, it is no longer necessarily a good idea to try to recover $Q_t$ as $\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$ in this example since in order for $\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n f$ to make sense one needs to assume that $f$ is a least $C^\infty$ and even this will not guarantee convergence of the sum!

*Remark 19.26.* A Lévy process is a general class of processes which contains both Brownian motions and Poisson processes as example. A **Lévy process** is a process with independent stationary increments which has right continuous paths. For a Levy process we must have $(Q_t f)(x) = (\mu_t * f)(x)$ where $\{\mu_t\}_{t > 0}$ is a one parameter convolution semi-group which is necessarily satisfies, $\hat{\mu}_t(\lambda) = e^{t\psi(\lambda)}$, where $\psi(\lambda)$ is a Levy exponent as described in Eq. (25.4) of Theorem 25.7 below. In more detail $\psi(\lambda)$ is any one of the functions of the form,

$$\psi(\lambda) = i\lambda b - \frac{1}{2} a \lambda^2 + \int_{\mathbb{R} \setminus \{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x| \le 1}\right) d\nu(x),$$

where $b \in \mathbb{R}$, $a \ge 0$, and $\nu$ is a measure on $\mathbb{R} \setminus \{0\}$ such that $\int_{\mathbb{R} \setminus \{0\}} \left(x^2 \wedge 1\right) d\nu(x) < \infty$.

*Example 19.27.* If $\{N_t\}_{t \ge 0}$ is a Poisson process and $\{B_t\}_{t \ge 0}$ is a Brownian motion which is independent of $\{N_t\}_{t \ge 0}$, then $X_t = B_t + N_t$ is a Lévy process, i.e. has independent stationary increments and is right continuous. The process

$X$ is a time homogeneous Markov process with Markov transition kernels given by;

$$\begin{aligned}(Q_t f)(x) &= \mathbb{E} f(x+N_t+B_t) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y|^2} \mathbb{E}\left[f(x+y+N_t)\right] dy \\ &= \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2t}|y|^2} \frac{(\lambda t)^n}{n!} f(x+y+n)\, dy \\ &= \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2t}|y-n|^2} \frac{(\lambda t)^n}{n!} f(x+y)\, dy \\ &= \int_{\mathbb{R}} q_t(y-x) f(y)\, dy\end{aligned}$$

where

$$q_t(y) := \frac{e^{-\lambda t}}{\sqrt{2\pi t}} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\frac{1}{2t}|y-n|^2}.$$

The infinitesimal generator, $A = \frac{d}{dt}|_{0+} Q_t$ of this process satisfies,

$$(Af)(x) = \frac{1}{2} f''(x) + \lambda \left(f(x+1) - f(x)\right)$$

at least for all $f \in C_c^2(\mathbb{R})$. This example will be significantly generalized in Theorem 19.30 below.

In order to continue with giving examples in the continuous time case we will need a simple measure theoretic result.

**Lemma 19.28.** *If $\Omega$ is a set, $\Omega_0 \subset \Omega$, and $\mathcal{B}_0$ is a $\sigma$ – algebra on $\Omega_0$, then $\tilde{\mathcal{B}}_0 := \{A \subset \Omega : A \cap \Omega_0 \in \mathcal{B}_0\}$ is a $\sigma$ – algebra on $\Omega$. Moreover, $f : \Omega \to \mathbb{R}$ is $\tilde{\mathcal{B}}_0$ – measurable iff $f|_{\Omega_0}$ is $\mathcal{B}_0$ measurable.*

**Proof.** It is clear that $\emptyset, \Omega \in \tilde{\mathcal{B}}_0$ and that $\tilde{\mathcal{B}}_0$ is closed under countable unions since $A_n \in \tilde{\mathcal{B}}_0$ iff $A_n \cap \Omega_0 \in \mathcal{B}_0$ which implies $[\cup A_n] \cap \Omega_0 = \cup [A_n \cap \Omega_0] \in \mathcal{B}_0$ and this implies that $[\cup A_n] \in \tilde{\mathcal{B}}_0$. Lastly if $A \in \tilde{\mathcal{B}}_0$ then $A \cap \Omega_0 \in \tilde{\mathcal{B}}_0$ implies that

$$A^c \cap \Omega_0 = \Omega_0 \setminus A = \Omega_0 \setminus [A \cap \Omega_0] \in \mathcal{B}_0$$

and therefore $A \in \tilde{\mathcal{B}}_0$.

For the second assertion, let us observe that for $W \in \mathcal{B}_{\mathbb{R}}$ we have

$$f^{-1}(W) \cap \Omega_0 = f|_{\Omega_0}^{-1}(W)$$

so that $f^{-1}(W) \in \tilde{\mathcal{B}}_0$ iff $f^{-1}(W) \cap \Omega_0 \in \mathcal{B}_0$ iff $f|_{\Omega_0}^{-1}(W) \in \mathcal{B}_0$. It now clearly follows that $f : \Omega \to \mathbb{R}$ is $\tilde{\mathcal{B}}_0$ – measurable iff $f|_{\Omega_0}$ is $\mathcal{B}_0$ measurable. ∎

**Definition 19.29.** *Suppose that $\Omega = \sum_{n=0}^{\infty} \Omega_n$ and $\mathcal{B}_n$ is a $\sigma$ – algebra on $\Omega_n$ for all $n$. Then we let $\oplus_{n=0}^{\infty} \mathcal{B}_n =: \mathcal{B}$ be the $\sigma$ – algebra on $\Omega$ such that $A \subset \Omega$ is measurable iff $A \cap \Omega_n \in \mathcal{B}_n$ for all $n$. That is to say $\mathcal{B} = \cap_{n=0}^{\infty} \tilde{\mathcal{B}}_n$ with $\tilde{\mathcal{B}}_n = \{A \subset \Omega : A \cap \Omega_n \in \mathcal{B}_n\}$.*

From Lemma 19.28 it follows that $f : \Omega \to \mathbb{R}$ is $\oplus_{n=0}^{\infty} \mathcal{B}_n$ – measurable iff $f^{-1}(W) \in \tilde{\mathcal{B}}_n$ for all $n$ iff $f|_{\Omega_n}^{-1}(W) \in \mathcal{B}_n$ for all $n$ iff $f|_{\Omega_n}$ is $\mathcal{B}_n$ – measurable for all $n$. We in fact do not really use any properties of $\Omega_n$ for these statements it is not even necessary for $n$ to run over a countable index set!

The compound Poisson process in the next theorem gives another example of a Lévy process an example of the construction in Theorem 19.32. (The reader should compare the following result with Theorem 25.39 below.)

**Theorem 19.30 (Compound Poisson Process).** *Suppose that $\{Z_i\}_{i=1}^{\infty}$ are i.i.d. random vectors in $\mathbb{R}^d$ and $\{N_t\}_{t \geq 0}$ be an independent Poisson process with intensity $\lambda$. Further let $Z_0 : \Omega \to \mathbb{R}^d$ be independent of $\{N_t\}_{t \geq 0}$ and the $\{Z_i\}_{i=1}^{\infty}$ and then define, for $t \in \mathbb{R}_+$,*

$$\mathcal{B}_t = \oplus_{n=0}^{\infty} \left[\sigma(N_s : s \leq t, Z_0, \dots, Z_n)\right]_{\{N_t = n\}}$$

*and $X_t := S_{N_t}$ where $S_n := Z_0 + Z_1 + \dots + Z_n$. Then $\{\mathcal{B}_t\}_{t \geq 0}$ is a filtration (i.e. it is increasing), $\{X_t\}_{t \geq 0}$ is a $\mathcal{B}_t$ – adapted process such that for all $0 \leq s < t$, $X_t - X_s$ is independent of $\mathcal{B}_s$. The increments are stationary and therefore $\{X_t\}_{t \geq 0}$ is a Lévy process. The time homogeneous transition kernel is given by*

$$\begin{aligned}(Q_t f)(x) &= \mathbb{E}\left[f(x + Z_1 + \dots + Z_{N_t})\right] \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \mathbb{E}\left[f(x + Z_1 + \dots + Z_n)\right].\end{aligned}$$

*If we define $\left(\tilde{Q} f\right)(x) := \mathbb{E}\left[f(x + Z_1)\right]$, the above equation may be written as,*

$$Q_t = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \tilde{Q}^n = e^{\lambda t (\tilde{Q} - I)}.$$

**Proof.** Let us begin by showing that $\mathcal{B}_t$ is increasing. First observe that

$$\left[\sigma(N_s : s \leq t, Z_0, \dots, Z_n)\right]_{\{N_t = n\}} = \{A \subset \{N_t = n\} : A \in \sigma(N_s : s \leq t, Z_0, \dots, Z_n)\}.$$

If $0 \leq s < t$ and $A \in \mathcal{B}_s$ then $A \cap \{N_s = m\} \in \sigma(N_r : r \leq s, Z_0, \dots, Z_m)$ therefore, for $n \geq m$, we have

$$A \cap \{N_s = m\} \cap \{N_t = n\} \in \sigma(N_r : r \leq t, Z_0, \dots, Z_m) \subset \sigma(N_r : r \leq t, Z_0, \dots, Z_n)$$

and we may conclude that

$$A \cap \{N_t = n\} = \cup_{m \le n} [A \cap \{N_s = m\} \cap \{N_t = n\}] \in \sigma (N_r : r \le t, Z_0, \dots, Z_n)$$

for all $n \in \mathbb{N}_0$. Thus we have shown $A \in \mathcal{B}_t$ and therefore $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \le t$. Since

$$X_t|_{N_t=n} = [Z_0 + Z_1 + \dots + Z_n]\,|_{N_t=n}$$

is $[\sigma (N_s : s \le t, Z_0, \dots, Z_n)]_{\{N_t=n\}}$ – measurable for all $n \in \mathbb{N}_0$, it follows by Lemma 19.28 that $X_t$ is $\mathcal{B}_t$ – measurable.

We now show that $X_t - X_s$ is independent of $\mathcal{B}_s$ for all $t \ge s$. To this end, let $f \in (\mathcal{B}_{\mathbb{R}^d})_b$ and $g \in (\mathcal{B}_s)_b$. Then for each $n \in \mathbb{N}_0$, we have

$$g|_{N_s=n} = G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right)$$

while

$$(X_t - X_s)\,|_{N_s=n} = Z_{n+1} + \dots + Z_{N_t}.$$

Therefore we have,

$$\mathbb{E}[f(X_t - X_s) \cdot g] = \sum_{m,n=0}^{\infty} \mathbb{E}[f(X_t - X_s) \cdot g : N_s = n,\ N_t - N_s = m]$$

and if we let $a_{m,n}$ be the summand on the right side of this equation we have,

$$\begin{aligned}
a_{m,n} &= \mathbb{E}\left[ f(Z_{n+1} + \dots + Z_{n+m}) \cdot G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right) : N_s = n,\ N_t - N_s = m \right] \\
&= \mathbb{E}\left[ f(Z_{n+1} + \dots + Z_{n+m})\, 1_{N_t - N_s = m} \cdot G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right) 1_{N_s = n} \right] \\
&= \mathbb{E}[f(Z_{n+1} + \dots + Z_{n+m})\, 1_{N_t - N_s = m}] \cdot \mathbb{E}\left[ G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right) 1_{N_s = n} \right] \\
&= e^{-\lambda(t-s)} \frac{(\lambda (t-s))^m}{m!} \mathbb{E}[f(Z_1 + \dots + Z_m)] \cdot \mathbb{E}\left[ G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right) 1_{N_s = n} \right].
\end{aligned}$$

Therefore it follows that

$$\begin{aligned}
\mathbb{E}[f(X_t - X_s) \cdot g] &= \sum_{m,n=0}^{\infty} a_{m,n} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{m,n} \\
&= (Q_{t-s} f)(0) \cdot \sum_{n=0}^{\infty} \mathbb{E}\left[ G_n \left( \{N_r\}_{r \le s}, Z_0, \dots, Z_n \right) 1_{N_s = n} \right] \\
&= (Q_{t-s} f)(0) \cdot \mathbb{E}[g]
\end{aligned}$$

from which it follows that $X_t - X_s$ is independent of $\mathcal{B}_s$ and

$$\mathbb{E}[f(X_t - X_s)] = (Q_{t-s} f)(0).$$

(This equation shows that the distribution of the increments is stationary.) We now know by Exercise 19.2 that $\{X_t\}$ is a Markov process and the transition kernel is given by

$$(Q_{t-s} f)(x) = \mathbb{E}[f(x + X_t - X_s)] = (Q_{t-s} f(x + \cdot))(0)$$

as described above. ■

*Remark 19.31.* If $Q_t f$ is as in Theorem 19.30, then (formally),

$$\frac{d}{dt} Q_t = \frac{d}{dt} e^{\lambda t (\tilde{Q} - I)} = \lambda \left( \tilde{Q} - I \right) Q_t = Q_t \lambda \left( \tilde{Q} - I \right)$$

where

$$\lambda \left( \tilde{Q} - I \right) f(x) = \lambda \mathbb{E}[f(x + Z_1) - f(x)] = \lambda \int_{\mathbb{R}^d} [f(x + \xi) - f(x)]\, d\mu(\xi)$$

and $\mu = \mathrm{Law}(Z_1)$. This a process which jumps at rate $\lambda$ from position $x$ to position $x + \xi$ with distribution $\mu$.

## 19.4 Continuous Time Markov Chains on Denumerable State Spaces

Our goal in this section is to give an introduction to the general story of continuous time homogeneous Markov chains on countable state spaces, $S$. For this section and the rest of this chapter we will be assuming that $S$ is now at most countable.

**Theorem 19.32.** *Suppose that* $\left( \Omega, \left\{ \tilde{\mathcal{B}}_n \right\}_{n \in \mathbb{N}_0}, \mathcal{B}, P, \{Y_n : \Omega \to S\}_{n \in \mathbb{N}_0} \right)$ *is a time homogeneous Markov chain with one step transition kernel,* $\tilde{Q}$*. Further suppose that* $\{N_t\}_{t \ge 0}$ *is a Poisson process with parameter* $\lambda$ *which is independent of* $\tilde{\mathcal{B}}_\infty := \vee_n \tilde{\mathcal{B}}_n$*. Let* $\mathcal{B}_s$ *be the* $\sigma$ *– algebra on* $\Omega$ *such that*

$$[\mathcal{B}_s]_{N_s = n} := \left[ \sigma \left( N_r : r \le s\ \&\ \tilde{\mathcal{B}}_n \right) \right]_{N_s = n}.$$

*To be more explicit,* $A \subset \Omega$ *is in* $\mathcal{B}_s$ *iff*

$$A \cap \{N_s = n\} \in \sigma \left( N_r : r \le s\ \&\ \tilde{\mathcal{B}}_n \right) \text{ for all } n \in \mathbb{N}_0.$$

*Alternatively, we may describe* $\mathcal{B}_s$ *as the* $\sigma$ *– algebra generated by sets of the form*

$$A = B \cap C \ \ni\ B := [\cap_{i=1}^n \{N_{s_i} = k_i\}] \text{ and } C \in \tilde{\mathcal{B}}_{k_n} \tag{19.36}$$

*where* $n \in \mathbb{N}$, $0 = s_0 < s_1 < \dots < s_n = s$, $k_i \in \mathbb{N}$ *are such that* $k_1 \le k_2 \le \dots \le k_n$.

*If* $X_t := Y_{N_t}$ *for* $t \in \mathbb{R}_+$, *then* $\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}$ *is a filtration,* $\{X_t\}_{t\ge 0}$ *is adapted to this filtration and is a time homogeneous Markov process with transition semi-group given by*

$$Q_t = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\tilde{Q}^n = e^{-\lambda t}e^{\lambda t\tilde{Q}} = e^{t\lambda\left(\tilde{Q}-I\right)}. \tag{19.37}$$

**Proof.** Let us begin by showing that $\mathcal{B}_t$ is increasing. If $0 \le s < t$ and $A \in \mathcal{B}_s$ then $A \cap \{N_s = m\} \in \sigma\left(N_r : r \le s \ \&\ \tilde{\mathcal{B}}_m\right)$ therefore, for $n \ge m$, we have

$$A \cap \{N_s = m\} \cap \{N_t = n\} \in \sigma\left(N_r : r \le t \ \&\ \tilde{\mathcal{B}}_m\right) \subset \sigma\left(N_r : r \le t \ \&\ \tilde{\mathcal{B}}_n\right)$$

and therefore,

$$A \cap \{N_t = n\} = \cup_{m\le n}\left[A \cap \{N_s = m\} \cap \{N_t = n\}\right] \in \sigma\left(N_r : r \le t \ \&\ \tilde{\mathcal{B}}_n\right)$$

for all $n \in \mathbb{N}_0$. Thus we have shown $A \in \mathcal{B}_t$ and therefore $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \le t$. Since $X_t|_{N_t=n} = Y_n|_{N_t=n}$ is $\sigma\left(N_r : r \le t \ \&\ \tilde{\mathcal{B}}_n\right)$ – measurable for all $n \in \mathbb{N}_0$, it follows by Lemma 19.28 that $X_t$ is $\mathcal{B}_t$ – measurable.

We now need to show that Markov property. To this end let $t \ge s$, $f \in \mathcal{S}_b$ and $g \in (\mathcal{B}_s)_b$. Then for each $n \in \mathbb{N}_0$, we have $g|_{N_s=n} \in \sigma\left(N_r : r \le s \ \&\ \tilde{\mathcal{B}}_n\right)$ and therefore,

$$\begin{aligned}
\mathbb{E}\left[f(X_t)\, g\right] &= \sum_{m,n=0}^{\infty}\mathbb{E}\left[f(X_t)\cdot g : N_s = n,\ N_t - N_s = m\right] \\
&= \sum_{m,n=0}^{\infty}\mathbb{E}\left[f(Y_{n+m})\, 1_{N_t-N_s=m}\cdot g\cdot 1_{N_s=n}\right] \\
&= \sum_{m,n=0}^{\infty}P(N_t - N_s = m)\,\mathbb{E}\left[f(Y_{n+m})\cdot g\cdot 1_{N_s=n}\right] \\
&= \sum_{m,n=0}^{\infty}e^{-\lambda(t-s)}\frac{(\lambda(t-s))^m}{m!}\mathbb{E}\left[\left(\tilde{Q}^m f\right)(Y_n)\cdot g\cdot 1_{N_s=n}\right] \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[(Q_{t-s}f)(Y_n)\cdot g\cdot 1_{N_s=n}\right] = \mathbb{E}\left[(Q_{t-s}f)(Y_{N_s})\cdot g\right] \\
&= \mathbb{E}\left[(Q_{t-s}f)(X_s)\cdot g\right].
\end{aligned}$$

This shows that

$$\mathbb{E}\left[f(X_t)\,|\mathcal{B}_s\right] = (Q_{t-s}f)(X_s) \ \ P-\text{a.s.} \tag{19.38}$$

which completes the proof.

**Second proof using Eq. (19.36).** If $A = B\cap C \in \mathcal{B}_s$ be as in Eq. (19.36), $f \in \mathcal{S}_b$, and $t > s$, then

$$\begin{aligned}
\mathbb{E}\left[f(X_t) : A\right] &= \sum_{m\ge k_n}\mathbb{E}\left[f(Y_m) : A, N_t = m\right] \\
&= \sum_{m=0}^{\infty}\mathbb{E}\left[f(Y_{k_n+m}) : B\cap C\cap\{N_t - N_s = m\}\right] \\
&= \sum_{m=0}^{\infty}\mathbb{E}\left[f(Y_{k_n+m}) : C\right]P(B)\,P(N_t - N_s = m) \\
&= \sum_{m=0}^{\infty}\mathbb{E}\left[\left(\tilde{Q}^m f\right)(Y_{k_n}) : C\right]P(B)\,e^{-\lambda(t-s)}\frac{(\lambda(t-s))^m}{m!} \\
&= e^{-\lambda(t-s)}\mathbb{E}\left[\sum_{m=0}^{\infty}\frac{(\lambda(t-s))^m}{m!}\left(\tilde{Q}^m f\right)(X_s) : B\cap C\right] \\
&= \mathbb{E}\left[\left(e^{\lambda(t-s)\left(\tilde{Q}-I\right)}f\right)(X_s) : A\right].
\end{aligned}$$

Since this equation holds for a multiplicative system of sets, $A$ of the form in Eq. (19.36) which generate $\mathcal{B}_s$, we may conclude again that Eq. (19.38) holds. ■

Suppose $S$ is countable or finite set and $a : S\times S \to \mathbb{R}$ is a function such that

$$a(x,y) \ge 0 \ \forall\ x \ne y \text{ and } \sum_{y\in S}a(x,y) = 0 \ \forall\ x \in S. \tag{19.39}$$

We further set and

$$a_x := \sum_{y\ne x}a(x,y) = -a(x,x) \in [0,\infty) \text{ and } \lambda := \sup_{x\in S}a_x. \tag{19.40}$$

Given a bounded function $f$ on $S$ we let

$$Af(x) = \sum_{y\in S}a(x,y)\,f(y) = \sum_{y\ne x}a(x,y)\left[f(y) - f(x)\right] \ \forall\ x \in S.$$

Notice that

$$|Af(x)| \le \sum_{y\in S}|a(x,y)|\,|f(y)| \le \sum_{y\in S}|a(x,y)|\,\|f\|_\infty = 2a_x\,\|f\|_\infty \le 2\lambda\,\|f\|_\infty.$$

Thus we see that $A$ takes bounded functions to bounded functions provided $\lambda < \infty$. If $\lambda < \infty$ it is readily verified that

$$Q_t := e^{tA} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}$$

is a well defined operator on $\mathcal{S}_b$ for all $t \in \mathbb{R}$.

We would now like to apply Theorem 19.32 in order to construct an associated time homogeneous Markov process with generator $A$. To do this we must find a Markov matrix, $\tilde{Q}$, on $S$ such that $A = \lambda\left(\tilde{Q} - I\right)$ for some $\lambda > 0$. If this is going to work we must take

$$\tilde{Q} := I + \frac{1}{\lambda} A. \tag{19.41}$$

Notice that no matter how we choose $\lambda$, $\tilde{Q}(x,y) \geq 0$ for $x \neq y$ and

$$\sum_{y \in S} \tilde{Q}(x,y) = \sum_{y \in S} \left[\delta_x(y) + \frac{1}{\lambda} a(x,y)\right] = 1 + 0 = 1.$$

So it only remains to check $\tilde{Q}(x,x) \geq 0$ for all $x$ and this will be the case iff

$$0 \geq 1 + \frac{1}{\lambda} a(x,x) = 1 - \frac{1}{\lambda} a_x \iff a_x \leq \lambda \text{ for all } x.$$

Thus the $\lambda$ we have defined in Eq. (19.40) will do the trick provided it is finite. We summarize the result in the next corollary.

**Corollary 19.33.** *Let $S$ be a countable or finite set and $a : S \times S \to \mathbb{R}$ be a function satisfying Eq. (19.39) and $\lambda := \sup_{x \in S} a_x < \infty$, see Eq. (19.40). Let $\tilde{Q} = I + \frac{1}{\lambda} A$ be the Markov matrix as in Eq. (19.41) so that $\tilde{Q}$ has matrix elements,*

$$\tilde{q}(x,y) := \begin{cases} \lambda^{-1} a(x,y) & \text{if } x \neq y \\ 1 - \lambda^{-1} a_x & \text{if } x = y. \end{cases}$$

*If $\{Y_n\}_{n=0}^{\infty}$ is a Markov chain with one step transition kernel, $\tilde{Q}$ and $\{N_t\}_{t \geq 0}$ is an independent Poisson process with intensity $\lambda$, then $X_t := Y_{N_t}$ is a time homogeneous Markov process with transition kernels,*

$$Q_t = e^{tA} = e^{-t\lambda} e^{t\lambda \tilde{Q}}.$$

*In particular, $A$ is the infinitesimal generator of a Markov transition semigroup.*

*Remark 19.34.* As in Exercise 19.9 we may directly show $Q_t(x,y) \geq 0$ for all $x, y \in S$ and that $\sum_{y \in S} Q_t(x,y) = 1$. [It is not really necessary to do this here since we have already constructed the process.] Indeed,

$$e^{t\lambda \tilde{Q}}(x,y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \tilde{Q}^n(x,y) \geq 0$$

and

$$\begin{aligned} \sum_{y \in S} e^{t\lambda \tilde{Q}}(x,y) &= \sum_{y \in S} \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \tilde{Q}^n(x,y) \\ &= \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \sum_{y \in S} \tilde{Q}^n(x,y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} 1 = e^{t\lambda}. \end{aligned}$$

Therefore $Q_t(x,y) = e^{tA}(x,y) = e^{-t\lambda} e^{t\lambda \tilde{Q}}(x,y) \geq 0$ and

$$\sum_{y \in S} Q_t(x,y) = \sum_{y \in S} e^{-t\lambda} e^{t\lambda \tilde{Q}}(x,y) = e^{-t\lambda} e^{t\lambda} = 1.$$

*Remark 19.35.* Let us pause to describe paths of the Markov process constructed in Corollary 19.33. To do this let $\{T_k\}_{k=1}^{\infty}$ be i.i.d. exponential random variables with intensity $\lambda$ such that $\{T_k\}$ are independent of the chain $\{Y_n\}_{n=0}^{\infty}$. We may then (see Definition 13.8) realize the Poisson process $\{N_t\}_{t \geq 0}$ with intensity $\lambda$ as the counting process,

$$N_t = \sum_{k=1}^{\infty} 1_{[0,t]}(W_k)$$

where

$$W_0 := 0 \text{ and } W_k := T_1 + \cdots + T_k \ \forall \ k \in \mathbb{N}.$$

Further let $\{\mathbf{j}_k\}_{k=0}^{\infty}$ be the jump times of the lazy chain $\{Y_n\}_{n=0}^{\infty}$ as described in Exercise 19.7. Then $J_k := W_{\mathbf{j}_k}$ is the time of the $k^{\text{th}}$ – jump of the $\{X_t := Y_{N_t}\}_{t \geq 0}$, see Figure 19.2 below. We further let $S_k := J_k - J_{k-1}$ for $k \in \mathbb{N}$ which denote the **sojourn time** that the process $\{X_t\}$ "lounges" at location $X_{J_{k-1}}$. It is now clear that the $\{X_t\}_{t \geq 0}$ may be reconstructed from the non - lazy chains $\{Y_{\mathbf{j}_k}\}_{k=0}^{\infty}$ and the corresponding "sojourn" times $\{S_k\}_{k=1}^{\infty}$. The next theorem describes the joint distributions $\{Y_0\} \cup \{Y_{\mathbf{j}_k}, S_k\}_{k=1}^{\infty}$.

**Theorem 19.36 (Jump-Hold Description II).** *Let $\{X_t\}_{t \geq 0}$ be the Markov process as constructed in Theorem 19.32 and Corollary 19.33. To avoid inessential notational complications, let us further suppose that $a_x > 0$ for all $x \in S$, i.e. no sites of $S$ are* ***absorbing.*** *As in Remark 19.35, let $J_0 = 0$ and for $k \in \mathbb{N}$ let $J_k$ be the time of the $k^{th}$ jump of $\{X_t\}_{t \geq 0}$ and $S_k := J_k - J_{k-1}$ be the $k^{th}$ "sojourn" time of the process. Then;*

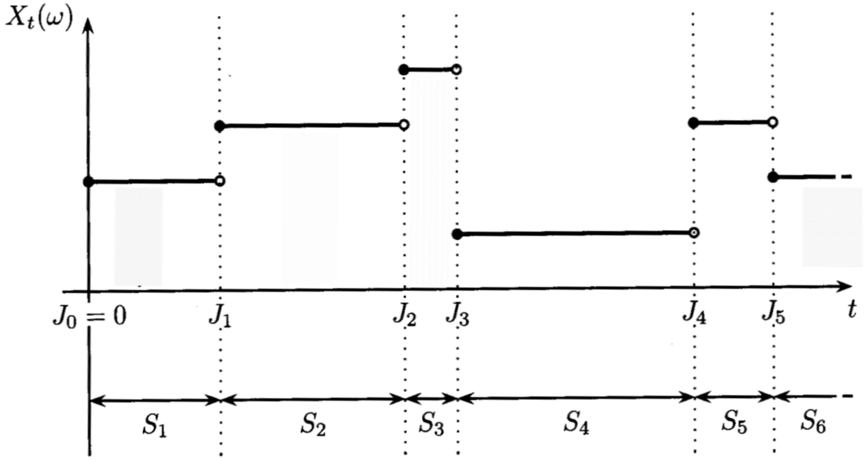



**Fig. 19.2.** Typical sample paths of a continuous time Markov chain in a discrete state space.

1. *$\{X_{J_k}\}_{k=0}^\infty$ is a discrete time Markov chain with Markov transition matrix, $\tilde{q}(x,y)$ defined by*

$$\tilde{q}(x,y) := \begin{cases} a(x,y)/a_x & \text{if } y \neq x \\ 0 & \text{if } y = x \end{cases}. \tag{19.42}$$

2. *If $\{x_k\}_{k=0}^n \subset S$ with $x_{k-1} \neq x_k$ for $1 \le k \le n$, then relative to $P(\cdot | \cap_{k=0}^n \{X_{J_k} = x_k\})$ the sojourn times, $\{S_k\}_{k=1}^n$, are independent exponential random variables with $S_k \overset{d}{=} \exp(a_{x_{k-1}})$.*

**Proof.** Recall that $\{Y_n\}_{n=0}^\infty$ was the Markov chain with transition kernels defined by (for some sufficiently large $\lambda > 0$)

$$\hat{q}(x,y) := \begin{cases} \lambda^{-1} a(x,y) & \text{if } x \neq y \\ 1 - \lambda^{-1} a_x & \text{if } x = y \end{cases}.$$

We also continue the notation set up in Remark 19.35.

1. By Exercise 19.7, $\{X_{J_k} = Y_{\mathbf{j}_k}\}_{k=0}^\infty$ is a Markov chain with Markov transition kernel given by $\tilde{q}(x,x) = 0$ for all $x \in S$ and for $x \neq y$,

$$\tilde{q}(x,y) := \frac{\hat{q}(x,y)}{1 - \hat{q}(x,x)} = \frac{\lambda^{-1} a(x,y)}{\lambda^{-1} a_x} = \frac{a(x,y)}{a_x}.$$

2. For $k \ge 1$, let $\sigma_k := \mathbf{j}_k - \mathbf{j}_{k-1}$ and given $\{x_k\}_{k=0}^n \subset S$ with $x_{k-1} \neq x_k$ for $1 \le k \le n$, let

$$A := \cap_{k=0}^n \{X_{J_k} = x_k\} = \cap_{k=0}^n \{Y_{\mathbf{j}_k} = x_k\}.$$

According to Exercise 19.7, relative to the probability measure $P(\cdot|A) = P(\cdot | [\cap_{k=0}^n \{Y_{\mathbf{j}_k} = x_k\}])$ the random $\mathbb{N}$ – valued times, $\{\sigma_k\}_{k=1}^n$, are independent with $\sigma_k \overset{d}{=} Geo(1 - \hat{q}(x_{k-1}, x_{k-1}))$ for $1 \le k \le n$. Combining this result with Exercise 19.11 below shows $\left\{S_k = W_{\mathbf{j}_k} - W_{\mathbf{j}_{k-1}}\right\}_{k=1}^n$ are are independent exponential random variables with

$$S_k \overset{d}{=} \exp((1 - \hat{q}(x_{k-1}, x_{k-1}))\lambda) \overset{d}{=} \exp(a_{x_{k-1}}).$$

■

**Exercise 19.11.** Let $\{T_k\}_{k=1}^\infty$ be i.i.d. exponential random variables with intensity $\lambda$ and $\{\sigma_\ell\}_{\ell=1}^n$ be independent geometric random variables with $\sigma_\ell = Geo(b_\ell)$ for some $b_\ell \in (0,1]$. Further assume that $\{\sigma_\ell\}_{\ell=1}^n \cup \{T_k\}_{k=1}^\infty$ are independent. We also let

$$W_0 = 0, \quad W_n := T_1 + \ldots + T_n,$$
$$\mathbf{j}_0 = 0, \ \mathbf{j}_\ell := \sigma_1 + \cdots + \sigma_\ell,$$
$$S_\ell := W_{\mathbf{j}_\ell} - W_{\mathbf{j}_{\ell-1}} \text{ for } 1 \le \ell \le n.$$

Show $\{S_\ell\}_{\ell=1}^n$ are independent exponential random variables with $S_\ell \overset{d}{=} \exp(b_\ell \lambda)$ for all $1 \le \ell \le n$.

## 19.5 First Step Analysis and Hitting Probabilities

In this section we suppose that $T = \mathbb{N}_0$, $(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in T})$ is a filtered measures space, $X_t : \Omega \to S$ is a $\mathcal{B}_t/\mathcal{S}$ – measurable function for all $t \in T$, $Q : S \times \mathcal{S} \to [0,1]$ is a Markov-transition kernel, and for each $x \in S$ there exists a probability, $P_x$, on $(\Omega, \mathcal{B})$ such that $P_x(X_0 = x) = 1$ and $\{X_t\}_{t \ge 0}$ is a time homogenous Markov process with $Q$ as its one step Markov transition kernel. To shorten notation we will write $\mathbb{E}_x$ for the expectation relative to the measure $P_x$.

**Definition 19.37 (Hitting times).** *For $B \in \mathcal{S}$, let*

$$T_B(X) := \min\{n \ge 0 : X_n \in B\}$$

*with the convention that $\min \emptyset = \infty$. We call $T_B(X) = T_B(X_0, X_1, \ldots)$ the **first hitting time** of $B$ by $X = \{X_n\}_n$.*

**Notation 19.38** *For $A \in \mathcal{S}$, let $Q_A : A \times \mathcal{S}_A \to [0,1]$ be the restriction of $Q$ to $A$, so that $Q_A(x,C) := Q(x,C)$ for all $x \in A$ and $C \in \mathcal{S}_A$. As with probability kernels we may identify $Q_A$ with an operator from $(\mathcal{S}_A)_b$ to itself via,*

$$(Q_A f)(x) = \int_A Q(x,dy) f(y) \text{ for all } x \in A \text{ and } f \in (\mathcal{S}_A)_b.$$

**Lemma 19.39 (A $T_B < \infty$ criteria).** *Let $B \in \mathcal{S}$, $A := B^c$, and*

$$\alpha := \sup_{y \in A} P_y (T_B = \infty). \tag{19.43}$$

*Then*

$$P_x (T_B = \infty) = \lim_{n \to \infty} (Q_A^n 1)(x) \ \forall \ x \in A \tag{19.44}$$

*and if $\alpha < 1$, then $P_x (T_B = 0) = 0$ for all $x \in A$. [This lemma is generalized in Theorem 19.52 below.]*

**Proof.** Let $u(x) := P_x (T_B = \infty)$ for $x \in A := B^c$, and observe that

$$P_x (T_B > n) = P_x (X_1 \in A, \dots, X_n \in A) = (Q_A^n 1)(x). \tag{19.45}$$

Letting $n \uparrow \infty$ shows that

$$u(x) := P_x (T_B = \infty) = \lim_{n \to \infty} (Q_A^n 1)(x)$$

which proves Eq. (19.44). Passing to the limit as $n \to \infty$ in the identity,

$$(Q_A^n 1)(x) = \int_A Q(x, dy) \left(Q_A^{n-1} 1\right)(y)$$

shows (using DCT) that

$$u(x) = \lim_{n \to \infty} (Q_A^n 1)(x) = \int_A Q(x, dy) \lim_{n \to \infty} \left(Q_A^{n-1} 1\right)(y) = \int_A Q(x, dy)\, u(y),$$

i.e. $u = Q_A u$. Iterating the equation $u = Q_A u$ then implies, $u = Q_A^n u$. However this then implies

$$u(x) = \int_A Q_A^n (x, dy)\, u(y) \le \int_A Q_A^n (x, dy)\, \alpha = \alpha (Q_A^n 1)(x).$$

Letting $n \to \infty$ then shows $u(x) \le \alpha u(x)$ from which it follows that $u(x) = 0$ if $\alpha < 1$. ■

**Exercise 19.12.** Suppose that $T$ is a $\mathbb{N}_0$ – valued random variable and $n \in \mathbb{N}$. Show

$$\mathbb{E} T \le n \sum_{k=0}^{\infty} P(T > nk). \tag{19.46}$$

**Lemma 19.40 (An $\mathbb{E} T_B < \infty$ criteria).** *Let $B \in \mathcal{S}$, $A := B^c$, and for $n \in \mathbb{N}$ let*

$$\alpha_n := \sup_{y \in A} P_y (T_B > n). \tag{19.47}$$

*Then*

$$\mathbb{E}_x T_B \le \frac{n}{1 - \alpha_n} \ \forall \ x \in A \tag{19.48}$$

*and in particular $\mathbb{E}_x T_B < \infty$ for all $x \in A$ if $\alpha_n < 1$ for some $n \in \mathbb{N}$.*

**Proof.** From Eq. (19.45), we have

$$\alpha_n = \sup_{y \in A} P_y (T_B > n) = \sup_{y \in A} (Q_A^n 1)(y).$$

For $k \in \mathbb{N}_0$,

$$\begin{aligned} Q_A^{nk} 1 &= [Q_A^n]^{k-1} Q_A^n 1 \le [Q_A^n]^{k-1} \alpha_n \\ &\le [Q_A^n]^{k-2} \alpha_n^2 \le \dots \le \alpha_n^k. \end{aligned}$$

Therefore, using Eq. (19.46),

$$\begin{aligned} \mathbb{E}_x T_B &\le n \sum_{k=0}^{\infty} P_x (T_B > nk) = n \sum_{k=0}^{\infty} \left(Q_A^{nk} 1\right)(x) \\ &\le n \sum_{k=0}^{\infty} \alpha_n^k = \frac{n}{1 - \alpha_n} \ \forall \ x \in A. \end{aligned}$$

■

**Theorem 19.41.** *Let $n$ denote a non-negative integer, $B \in \mathcal{S}$, and $A := B^c$. If $h : B \to \mathbb{R}$ is measurable and either bounded or non-negative, then*

$$\mathbb{E}_x [h(X_n) : T_B = n] = \left(Q_A^{n-1} Q [1_B h]\right)(x)$$

*and*

$$\mathbb{E}_x [h(X_{T_B}) : T_B < \infty] = \left(\sum_{n=0}^{\infty} Q_A^n Q [1_B h]\right)(x). \tag{19.49}$$

*If $g : A \to \mathbb{R}_+$ is a measurable function, then for all $x \in A$ and $n \in \mathbb{N}_0$,*

$$\mathbb{E}_x [g(X_n) 1_{n < T_B}] = (Q_A^n g)(x).$$

*In particular we have*

$$\mathbb{E}_x \left[\sum_{n < T_B} g(X_n)\right] = \sum_{n=0}^{\infty} (Q_A^n g)(x) =: u(x), \tag{19.50}$$

*where by convention, $\sum_{n < T_B} g(X_n) = 0$ when $T_B = 0$.*

**Proof.** Let $x \in A$. In computing each of these quantities we will use;

$$\begin{aligned} \{T_B > n\} &= \{X_i \in A \text{ for } 0 \le i \le n\} \text{ and} \\ \{T_B = n\} &= \{X_i \in A \text{ for } 0 \le i \le n-1\} \cap \{X_n \in B\}. \end{aligned}$$

From the second identity above it follows that for

$$\begin{aligned}\mathbb{E}_x\left[h\left(X_n\right):T_B=n\right]&=\mathbb{E}_x\left[h\left(X_n\right):\left(X_1,\dots,X_{n-1}\right)\in A^{n-1},\,X_n\in B\right]\\&=\sum_{n=1}^{\infty}\int_{A^{n-1}\times B}\prod_{j=1}^{n}Q\left(x_{j-1},dx_j\right)h\left(x_n\right)\\&=\left(Q_A^{n-1}Q\left[1_Bh\right]\right)(x)\end{aligned}$$

and therefore

$$\begin{aligned}\mathbb{E}_x\left[h\left(X_{T_B}\right):T_B<\infty\right]&=\sum_{n=1}^{\infty}\mathbb{E}_x\left[h\left(X_n\right):T_B=n\right]\\&=\sum_{n=1}^{\infty}Q_A^{n-1}Q\left[1_Bh\right]=\sum_{n=0}^{\infty}Q_A^nQ\left[1_Bh\right].\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}_x\left[g\left(X_n\right)1_{n<T_B}\right]&=\int_{A^n}Q\left(x,dx_1\right)Q\left(x_1,dx_2\right)\dots Q\left(x_{n-1},dx_n\right)g\left(x_n\right)\\&=\left(Q_A^ng\right)(x)\end{aligned}$$

and therefore,

$$\begin{aligned}\mathbb{E}_x\left[\sum_{n=0}^{\infty}g\left(X_n\right)1_{n<T_B}\right]&=\sum_{n=0}^{\infty}\mathbb{E}_x\left[g\left(X_n\right)1_{n<T_B}\right]\\&=\sum_{n=0}^{\infty}\left(Q_A^ng\right)(x).\end{aligned}$$

■

In practice it is not so easy to sum the series in Eqs. (19.49) and (19.50). Thus we would like to have another way to compute these quantities. Since $\sum_{n=0}^{\infty}Q_A^n$ is a geometric series, we expect that

$$\sum_{n=0}^{\infty}Q_A^n=\left(I-Q_A\right)^{-1}$$

which is basically correct at least when $(I-Q_A)$ is invertible. This suggests that if $u(x)=\mathbb{E}_x\left[h\left(X_{T_B}\right):T_B<\infty\right]$, then (see Eq. (19.49))

$$u=Q_Au+Q\left[1_Bh\right]\ \text{ on } A, \tag{19.51}$$

and if $u(x)=\mathbb{E}_x\left[\sum_{n<T_B}g\left(X_n\right)\right]$, then (see Eq. (19.50))

$$u=Q_Au+g \text{ on } A. \tag{19.52}$$

That these equations are valid is the content of Corollaries 19.44 and 19.45 below which we will prove using the "first step" analysis in the next theorem. We will give another direct proof in Theorem 19.52 below as well.

**Theorem 19.42 (First step analysis).** *Let us keep the assumptions in Theorem 19.14 and add the further assumption that $T=\mathbb{N}_0$. Then for all $F\in\mathcal{S}_b^{\otimes\mathbb{N}}$ or $F:S^{\mathbb{N}_0}\to[0,\infty]$ measurable;*

$$\mathbb{E}_x\left[F\left(X_0,X_1,\dots\right)\right]=\int_SQ\left(x,dy\right)\mathbb{E}_yF\left(x,X_0,X_1,\dots\right). \tag{19.53}$$

*This equation can be iterated to show more generally that*

$$\mathbb{E}_x\left[F\left(X_0,X_1,\dots\right)\right]=\int_{S^n}\prod_{j=1}^{n}Q\left(x_{j-1},dx_j\right)\mathbb{E}_{x_n}\left[F\left(x_0,x_1,\dots,x_{n-1},X_0,X_1,\dots\right)\right] \tag{19.54}$$

*where $x_0:=x$.*

**Proof.** Since $X_0(\omega)=x$ for $P_x$ – a.e. $\omega$, we have $F\left(X_0,X_1,\dots\right)=F\left(x,X_1,X_2,\dots\right)$ a.s. Therefore by Theorem 19.14 we know that

$$\mathbb{E}_x\left[F\left(X_0,X_1,\dots\right)|\mathcal{B}_1\right]=\mathbb{E}_x\left[F\left(x,X_1,X_2,\dots\right)|\mathcal{B}_1\right]=\mathbb{E}_{X_1}F\left(x,X_0,X_1,\dots\right).$$

Taking expectations of this equation shows,

$$\begin{aligned}\mathbb{E}_x\left[F\left(X_0,X_1,\dots\right)\right]&=\mathbb{E}_x\left[\mathbb{E}_{X_1}F\left(x,X_0,X_1,\dots\right)\right]\\&=\int_SQ\left(x,dy\right)\mathbb{E}_yF\left(x,X_0,X_1,\dots\right).\end{aligned}$$

■

*Remark 19.43.* Let $T_B\left(x_0,x_1,x_2,\dots\right)$ be the first hitting time of $B$ of the sequence $\left(x_0,x_1,x_2,\dots\right)$ and when when $T_B\left(x_0,x_1,x_2,\dots\right)<\infty$, let $L_B\left(x_0,x_1,x_2,\dots\right)=x_{T_B\left(x_0,x_1,x_2,\dots\right)}$ be the hitting location. In more detail, $T_B\left(x_0,x_1,x_2,\dots\right)=n$ and $L_B\left(x_0,x_1,x_2,\dots\right)=x_n$ iff $x_n\in B$ while $x_k\in A=B^c$ for $0\le k<n$. As usual we say $T_B\left(x_0,x_1,x_2,\dots\right)=\infty$ if $x_k\in A$ for all $k$ and in this case we leave $L_B\left(x_0,x_1,x_2,\dots\right)$ undefined. In using Theorem 19.42 we will often implicitly make use of the following two simple observations.

1. If $x_0\in A$ and $T_B\left(x_0,x_1,x_2,\dots\right)<\infty$ then
$$L_B\left(x_0,x_1,x_2,\dots\right)=L_B\left(x_1,x_2,\dots\right).$$
2. If $x_0\in A$, then
$$T_B\left(x_0,x_1,x_2,\dots\right)=1+1_A\left(x_1\right)T_B\left(x_1,x_2,\dots\right).$$

3. More generally if $g : A \to [0,\infty]$ is a function,

$$F(x_0, x_1, x_2, \dots) = \sum_{0 \le n < T_B(x_0, x_1, x_2, \dots)} g(x_n), \tag{19.55}$$

and $x_0 \in A$, then

$$F(x_0, x_1, x_2, \dots) = g(x_0) + 1_A(x_1) F(x_1, x_2, \dots).$$

[Taking $g = 1$ gives item 2.]

The reader might interpret $F(x_0, x_1, x_2, \dots)$ in Eq. (19.55) as money collected in $A$ by a traveler on the path $(x_0, x_1, x_2, \dots)$ up to its first exit from $A$ where $g(x)$ is the amount of money at location $x \in A$.

**Corollary 19.44.** *Suppose that $B \in \mathcal{S}$, $A := B^c \in \mathcal{S}$, $h : B \to \mathbb{R}$ is a measurable function which is either bounded or non-negative, and*

$$u(x) := \mathbb{E}_x[h(X_{T_B}) : T_B < \infty] \text{ for } x \in S.$$

*Then $u : S \to \mathbb{R}$ satisfies $u = h$ on $B$ and*

$$u = Qu = Q_A u + Q_B h \text{ on } A$$

*as in Eq. (19.51). In more detail*

$$u(x) = \int_A Q(x, dy) u(y) + \int_B Q(x, dy) h(y) \text{ for all } x \in A.$$

*In particular, when $h \equiv 1$, $u(x) = P_x(T_B < \infty)$ is a solution to the equation,*

$$u = Q_A u + Q 1_B \text{ on } A. \tag{19.56}$$

**Proof.** To shorten the notation we will use the convention that $h(X_{T_B}) = 0$ if $T_B = \infty$ so that we may simply write $u(x) := \mathbb{E}_x[h(X_{T_B})]$. Let

$$F(X_0, X_1, \dots) = h\left(X_{T_B(X)}\right) = h\left(X_{T_B(X)}\right) 1_{T_B(X) < \infty},$$

then for $x \in A$ we have $F(x, X_0, X_1, \dots) = F(X_0, X_1, \dots)$, see Remark 19.43. Therefore by the first step analysis (Theorem 19.42) we learn for $x \in A$ that

$$\begin{aligned} u(x) &= \mathbb{E}_x h\left(X_{T_B(X)}\right) = \mathbb{E}_x F(x, X_1, \dots) = \int_S Q(x, dy) \mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \int_S Q(x, dy) \mathbb{E}_y F(X_0, X_1, \dots) = \int_S Q(x, dy) u(y) \\ &= \int_A Q(x, dy) u(y) + \int_B Q(x, dy) h(y), \end{aligned}$$

i.e.

$$u = Qu = Q_A u + Q_B h \text{ on } A.$$

■

**Corollary 19.45.** *Suppose that $B \in \mathcal{S}$, $A := B^c \in \mathcal{S}$, $g : A \to [0,\infty]$ is a measurable function. Further let*

$$u(x) := \mathbb{E}_x\left[\sum_{0 \le n < T_B} g(X_n)\right] \text{ for } x \in S.$$

*Then $u(x) = 0$ if $x \in B$ and $u(x)$ satisfies Eq. (19.52), i.e.*

$$u = Qu + g = Q_A u + g \text{ on } A$$

*or in more detail,*

$$u(x) = \int_A Q(x, dy) u(y) + g(x) \text{ for all } x \in A.$$

*In particular if we take $g \equiv 1$ in this equation we learn that*

$$\mathbb{E}_x T_B = \int_A Q(x, dy) \mathbb{E}_y T_B + 1 \text{ for all } x \in A.$$

**Proof.** Let

$$F(X_0, X_1, \dots) = \sum_{0 \le n < T_B(X_0, X_1, \dots)} g(X_n)$$

be the sum of the values of $g$ along the chain before its first exit from $A$, i.e. entrance into $B$. With this interpretation in mind, if $x \in A$, it is easy to see that

$$\begin{aligned} F(x, X_0, X_1, \dots) &= \begin{cases} g(x) & \text{if } X_0 \in B \\ g(x) + F(X_0, X_1, \dots) & \text{if } X_0 \in A \end{cases} \\ &= g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \dots). \end{aligned}$$

Therefore by the first step analysis (Theorem 19.42) it follows that

$$\begin{aligned} u(x) &= \mathbb{E}_x F(X_0, X_1, \dots) = \int_S Q(x, dy) \mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \int_S Q(x, dy) \mathbb{E}_y [g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \dots)] \\ &= g(x) + \int_A Q(x, dy) \mathbb{E}_y [F(X_0, X_1, \dots)] \\ &= g(x) + \int_A Q(x, dy) u(y). \end{aligned}$$

■

The next corollary is hybrid of the previous two scenarios. You might envision a game show where you win money along the way determined by $g : A \to [0,\infty]$ with the possibility of keeping or loosing some portion of it which is determined by a function $h : B \to [0,\infty]$.

**Corollary 19.46.** *If $B \in \mathcal{S}$, $A := B^c \in \mathcal{S}$, $h : B \to [0,\infty]$ and $g : A \to [0,\infty]$ are measurable functions, then Let*

$$\mathbb{E}_x \left[ h(X_{T_B}) \sum_{0 \le m < T_B} g(X_m) : T_B < \infty \right] = u(x) \text{ for } x \in S,$$

*where $u : S \to \mathbb{R}$ satisfies $u = 0$ on $B$ and*

$$u = (I_A - Q_A)^{-1} g (I_A - Q_A)^{-1} Q_B h \text{ on } A$$

*where (if there is any ambiguity),*

$$(I_A - Q_A)^{-1} := \sum_{n=0}^{\infty} Q_A^n. \tag{19.57}$$

**Proof.** If we let

$$F(X_0, X_1, \dots) = 1_{T_B(X)<\infty} \cdot h(X_{T_B}) \sum_{0 \le m < T_B} g(X_m),$$

then for $x \in A$ we have by Remark 19.43 that

$$\begin{aligned} F(x, X_0, X_1, \dots) &= 1_{T_B(X)<\infty} \cdot h(X_{T_B}) \left[ g(x) + \sum_{0 \le m < T_B} g(X_m) \right] \\ &= g(x) 1_{T_B(X)<\infty} \cdot h(X_{T_B}) + F(X_0, X_1, \dots). \end{aligned}$$

Therefore by the first step analysis (Theorem 19.42) we learn for $x \in A$ that

$$\begin{aligned} u(x) &= \mathbb{E}_x F(x, X_1, \dots) = \int_S Q(x, dy) \mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \int_S Q(x, dy) \mathbb{E}_y \left[ g(x) 1_{T_B(X)<\infty} \cdot h(X_{T_B}) + F(X_0, X_1, \dots) \right] \\ &= \int_S g(x) Q(x, dy) \mathbb{E}_y \left[ 1_{T_B(X)<\infty} \cdot h(X_{T_B}) \right] + \int_S Q(x, dy) u(y) \\ &= g(x)(Qv)(x) + Qu(x) = g(x)(Qv)(x) + Q_A u(x) \end{aligned}$$

where

$$\begin{aligned} v(x) &:= \mathbb{E}_x \left[ 1_{T_B(X)<\infty} \cdot h(X_{T_B}) \right] \\ &= 1_A(x) \left( (I - Q_A)^{-1} Q_B h \right)(x) + 1_B(x) h(x) \end{aligned}$$

and we have used $u = 0$ on $B$. Putting this all together shows,

$$\begin{aligned} u &= g \left[ Q_A (I - Q_A)^{-1} Q_B h + Q_B h \right] + Q_A u \\ &= g (I - Q_A)^{-1} Q_B h + Q_A u \end{aligned}$$

or in other words,

$$u = (I - Q_A)^{-1} \left[ g (I - Q_A)^{-1} Q_B h \right].$$

Alternatively, we may work from first principles,

$$\begin{aligned} u(x) &= \mathbb{E}_x \left[ h(X_{T_B}) \sum_{0 \le m < T_B} g(X_m) : T_B < \infty \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}_x \left[ h(X_n) \sum_{0 \le m < n} g(X_m) : T_B = n \right] \\ &= \sum_{0 \le m < n < \infty} \mathbb{E}_x \left[ g(X_m) h(X_n) : X_1, \dots, X_{n-1} \in A, X_n \in B \right] \\ &= \sum_{0 \le m < n < \infty} \left[ Q_A^m g Q_A^{n-m-1} Q_B h \right](x) \\ &= \sum_{0 \le m < \infty} \left[ Q_A^m g (I_A - Q_A)^{-1} Q_B h \right](x) \\ &= \left[ (I_A - Q_A)^{-1} g (I_A - Q_A)^{-1} Q_B h \right](x). \end{aligned}$$

which is the same result as above. Moreover, this method shows that Eq. (19.57) is the correct way to interpret $(I_A - Q_A)^{-1}$ whenever there is any ambiguity in its meaning.

■

The problem with Corollaries 19.44 and 19.45 is that the solutions to Eqs. (19.51) and (19.52) may not be unique as we will see in the next examples. Theorem 19.52 below will explain when these ambiguities may occur and how to deal with them when they do.

*Example 19.47 (Biased random walks 0).* Let $p \in (1/2, 1)$ and consider the biased random walk $\{S_n\}_{n \ge 0}$ on the $S = \mathbb{Z}$ where $S_n = X_0 + X_1 + \dots + X_n$, $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with $P(X_i = 1) = p \in (0,1)$ and $P(X_i = -1) = q := 1 - p$, and $X_0 = x$ for some $x \in \mathbb{Z}$. This is time homogeneous Markov chain with transition matrix,

$$Q(x,y) = p1_{y=x+1} + q1_{y=x-1}.$$

For any $a \in \mathbb{Z}$ let $T_a$ be the first hitting time of $a$. Given $a < 0 < b$ we would like to compute $P_0(T_b < T_a) = \mathbb{E}_0 h\left(X_{T_{\{a,b\}}}\right)$ where $h(b) = 1$ and $h(a) = 0$. To do this we will compute

$$u(x) := P_x(T_b < T_a) = \mathbb{E}_x h\left(X_{T_{\{a,b\}}}\right) \text{ for } a \le x \le b$$

where $u(b) = 1$ and $u(a) = 0$. By the first step analysis (using $A = (a,b) \cap \mathbb{Z}$ and $Q_A(x,y) = (p1_{y=x+1} + q1_{y=x-1}) 1_{a<y<b}$) shows[4]

$$\begin{aligned} u(x) &= pu(x+1) + qu(x-1) \text{ for } a < x < b \text{ with} \\ u(a) &= 0 \text{ and } u(b) = 1. \end{aligned}$$

From Exercise 19.15 below, we know that the general solution to Eq. (19.59) is of the form

$$u(x) = a\lambda_+^x + b\lambda_-^x$$

where $\lambda_\pm$ are the roots for the characteristic polynomial, $p\lambda^2 - \lambda + q = 0$. since constants solve Eq. (19.59) we know that one root is 1 as is easily verified. The other root[5] is $\rho := q/p < 1$. Thus the general solution is of the form, $u(x) = \alpha + \beta\rho^{(x-a)}$. We now need to choose $\alpha$ and $\beta$ so that the boundary conditions $u(a) = 0$ and $u(b) = 1$ are satisfied, i.e. $\alpha + \beta = 0$ or $\beta = -\alpha$ and

$$1 = \alpha + \beta\rho^{(b-a)} = \alpha\left(1 - \rho^{(b-a)}\right) \implies \alpha = \frac{1}{1 - \rho^{(b-a)}}.$$

Thus we have

$$P_x(T_b < T_a) = \frac{1 - \rho^{(x-a)}}{1 - \rho^{(b-a)}} \implies P_0(T_b < T_a) = \frac{1 - \rho^{-a}}{1 - \rho^{(b-a)}}.$$

Letting $a \downarrow -\infty$ (keep in mind $\rho \in (0,1)$) then shows,

$$P_x(T_b < \infty) = 1 \text{ for all } x < b.$$

Also as $b \uparrow \infty$ $\{T_b < T_a\} \downarrow \{T_a = \infty\}$ and so

[4] For $x = b - 1$, we have $(Q_A u)(x) = qu(x-1)$ while $(Q_B h)(x) = ph(x+1) = p$ and so $(Q_A u)(x) + (Q_B h)(x) = qu(x-1) + pu(x+1)$ provided we remember that $u(x+1) = h(b) = 1$ when $x = b - 1$.

[5] Indeed,

$$p\left(\frac{q}{p}\right)^2 - \frac{q}{p} + q = \frac{q}{p}[q - 1 + p] = 0.$$



$$P_0(T_a = \infty) = \lim_{b\uparrow\infty} P_0(T_b < T_a) = \lim_{b\uparrow\infty} \frac{1 - \rho^{-a}}{1 - \rho^{(b-a)}} = 1 - \left(\frac{q}{p}\right)^{-a}.$$

Next let us try to find $u(x) := \mathbb{E}_x\left[T_{\{a,b\}}\right]$ which we know to be finite by Lemma 19.40. By the first step analysis we have,

$$\begin{aligned} u(x) &= p[u(x+1) + 1] + q[u(x-1) + 1] \\ &= pu(x+1) + qu(x-1) + 1 \text{ for } a < x < b \qquad (19.58) \\ u(a) &= 0 = u(b). \end{aligned}$$

A particular solution to Eq. (19.58) may be found by trying $u(x) = cx$. Plugging this into Eq. (19.58) then shows,

$$cx = cx + c(p - q) + 1 \implies c = \frac{1}{q - p}.$$

Let us notice that $v(x) = \rho^{(x-a)} - 1$ and $w(x) = \rho^{(x-b)} - 1$ are solutions to the homogeneous equations such that $v(a) = 0$ and $w(b) = 0$ and so in general

$$u(x) = \frac{x}{q - p} + \alpha\left(\rho^{(x-a)} - 1\right) + \beta\left(\rho^{(x-b)} - 1\right).$$

The coefficients $\alpha$ and $\beta$ are found by requiring

$$\begin{aligned} 0 = u(a) &= \frac{a}{q-p} + \beta\left(\rho^{(a-b)} - 1\right) \implies \beta = \frac{a}{p-q}\frac{1}{\rho^{(a-b)} - 1} \\ 0 = u(b) &= \frac{b}{q-p} + \alpha\left(\rho^{(b-a)} - 1\right) \implies \alpha = \frac{b}{p-q}\frac{1}{\rho^{(b-a)} - 1} \end{aligned}$$

and so we have shown,

$$\mathbb{E}_x T_{\{a,b\}} = \frac{1}{p-q}\left[\frac{\rho^{(x-a)} - 1}{\rho^{(b-a)} - 1}b + \frac{\rho^{(x-b)} - 1}{\rho^{(a-b)} - 1}a - x\right].$$

Letting $b \uparrow \infty$ shows $\mathbb{E}_x T_a = \infty$ and letting $a \downarrow -\infty$ shows

$$\mathbb{E}_x T_b = \frac{1}{p-q}[b - x] \text{ for } x < b.$$

*Example 19.48 (Biased random walks I).* Let $p \in (1/2, 1)$ and consider the biased random walk $\{S_n\}_{n\ge0}$ on the $S = \mathbb{Z}$ where $S_n = X_0 + X_1 + \cdots + X_n$, $\{X_i\}_{i=1}^\infty$ are i.i.d. with $P(X_i = 1) = p \in (0,1)$ and $P(X_i = -1) = q := 1 - p$, and $X_0 = x$ for some $x \in \mathbb{Z}$. Let $B := \{0\}$ and $u(x) := P_x(T_B < \infty)$. Clearly $u(0) = 0$ and by the first step analysis,

$$u(x) = pu(x+1) + qu(x-1) \text{ for } x \neq 0. \qquad (19.59)$$

From Exercise 19.15 below, we know that the general solution to Eq. (19.59) is of the form

$$u(x) = a\lambda_+^x + b\lambda_-^x$$

where $\lambda_\pm$ are the roots for the characteristic polynomial, $p\lambda^2 - \lambda + q = 0$. since constants solve Eq. (19.59) we know that one root is 1 as is easily verified. The other root[6] is $q/p$. Thus the general solution is of the form, $w(x) = a + b(q/p)^x$. In all case we are going to choose $a$ and $b$ so that $0 = u(0) = w(0)$ (i.e. $a+b=0$) so that $w(x) = a + (1-a)(q/p)^x$. For $x > 0$ we choose $a = a_+$ so that $w_+(x) := a_+ + (1-a_+)(q/p)^x$ satisfies $w_+(1) = u(1)$ and for $x < 0$ we choose $a = a_-$ so that $w_-(x) := a_- + (1-a_-)(q/p)^x$ satisfies $w_-(-1) = u(-1)$. With these choice we will have $u(x) = w_+(x)$ for $x \geq 0$ and $u(x) = w_-(x)$ for $x \leq 0$ – see Exercise 19.15 and Remark 19.50. Observe that

$$u(1) = a_+ + (1-a_+)(q/p) \implies a_+ = \frac{u(1) - (q/p)}{1-(q/p)}$$

and

$$u(-1) = a_- + (1-a_-)(p/q) \implies a_- = \frac{(p/q) - u(-1)}{(p/q) - 1}.$$

**Case 1.** $x < 0$ : As $x \to -\infty$, we will have $|u(x)| \to \infty$ unless $a_- = 1$. Thus we must take $a_- = 1$ and we have shown,

$$P_x(T_0 < \infty) = w_-(x) = 1 \text{ for all } x < 0.$$

**Case 2.** $x > 0$ : For $n \in \mathbb{N}_0$, let $T_n = \min\{m : X_m = n\}$ be the first time $X$ hits $n$. By the MCT we have,

$$P_x(T_0 < \infty) = \lim_{n\to\infty} P_x(T_0 < T_n).$$

So we will now try to compute $u(x) = P_x(T_0 < T_n)$. By the first step analysis (take $B = \{0, n\}$ and $h(0) = 1$ and $h(n) = 0$ in Corollary 19.44) we will still have that $u(x)$ satisfies Eq. (19.59) for $0 < x < n$ but now the boundary conditions are $u(0) = 1$ and $u(n) = 0$. Accordingly $u(x)$ for $0 \leq x \leq n$ is still of the form given in Eq. (19.59) but we may now determine $a = a_n$ using the boundary condition

$$0 = u(n) = a + (1-a)(q/p)^n = (q/p)^n + a(1-(q/p)^n)$$

from which it follows that

$$a_n = \frac{(q/p)^n}{(q/p)^n - 1} \to 0 \text{ as } n \to \infty.$$

Thus we have shown

$$\begin{aligned} P_x(T_0 < T_n) &= \frac{(q/p)^n}{(q/p)^n - 1} + \left(1 - \frac{(q/p)^n}{(q/p)^n - 1}\right)(q/p)^x \\ &= \frac{(q/p)^n - (q/p)^x}{(q/p)^n - 1} \\ &= \frac{(q/p)^x - (q/p)^n}{1 - (q/p)^n} \to (q/p)^x \text{ as } n \to \infty \end{aligned}$$

and therefore, since $T_n \uparrow \infty$ $P_x$ – a.s. as $n \uparrow \infty$,

$$P_x(T_0 < \infty) = (q/p)^x \text{ for all } x > 0.$$

*Example 19.49 (Biased random walks II).* Continue the notation in Example 19.48. Let us now try to compute $\mathbb{E}_x T_0$. Since $P_x(T_0 = \infty) > 0$ for $x > 0$ we already know that $\mathbb{E}_x T_0 = \infty$ for all $x > 0$. Nevertheless we will deduce this fact again here.

Letting $u(x) = \mathbb{E}_x T_0$ it follows by the first step analysis that, for $x \neq 0$,

$$\begin{aligned} u(x) &= p[1 + u(x+1)] + q[1 + u(x-1)] \\ &= pu(x+1) + qu(x-1) + 1 \end{aligned} \tag{19.60}$$

with $u(0) = 0$. Notice $u(x) = \infty$ is a solution to this equation while if $u(a) < \infty$ for some $a \neq 0$ then Eq. (19.60) implies that $u(x) < \infty$ for all $x \neq 0$ with the same sign as $a$.

A particular solution to this equation may be found by trying $u(x) = \alpha x$ to learn,

$$\alpha x = p\alpha(x+1) + q\alpha(x-1) + 1 = \alpha x + \alpha(p-q) + 1$$

which is valid for all $x$ provided $\alpha = (q-p)^{-1}$. The general **finite** solution to Eq. (19.60) is therefore,

$$u(x) = (q-p)^{-1} x + a + b(q/p)^x. \tag{19.61}$$

Using the boundary condition, $u(0) = 0$ allows us to conclude that $a + b = 0$ and therefore,

$$u(x) = u_a(x) = (q-p)^{-1} x + a[1 - (q/p)^x]. \tag{19.62}$$

Notice that $u_a(x) \to -\infty$ as $x \to +\infty$ no matter how $a$ is chosen and therefore we must conclude that the desired solution to Eq. (19.60) is $u(x) = \infty$ for $x > 0$ as we already mentioned.

[6] Indeed,

$$p\left(\frac{q}{p}\right)^2 - \frac{q}{p} + q = \frac{q}{p}[q - 1 + p] = 0.$$

The question now is for $x<0$. Is it again the case that $u(x)=\infty$ or is $u(x)=u_a(x)$ for some $a\in\mathbb{R}$. Since $\lim_{x\to\infty}u_a(x)=-\infty$ unless $a\le 0$, we may restrict our attention to $a\le 0$. To work out which $a\le 0$ is correct observe by MCT that

$$\mathbb{E}_xT_0=\lim_{n\to-\infty}\mathbb{E}_x\left[T_n\wedge T_0\right]=\lim_{n\to-\infty}\mathbb{E}_x\left[T_{\{n,0\}}\right].$$

So let $n\in\mathbb{Z}$ with $n<0$ be fixed for the moment. By item 8. of Theorem 19.52 we may conclude that $u(x):=\mathbb{E}_x\left[T_{\{n,0\}}\right]<\infty$ for all $n\le x\le 0$. Then by the first step analysis, $u(x)$ satisfies Eq. (19.60) for $n<x<0$ and has boundary conditions $u(n)=0=u(0)$. Using the boundary condition $u(n)=0$ to determine $a=a_n$ in Eq. (19.62) implies,

$$0=u_a(n)=(q-p)^{-1}n+a\left[1-(q/p)^n\right]$$

so that

$$a=a_n=\frac{n}{\left(1-(q/p)^n\right)(p-q)}\to 0\text{ as }n\to-\infty.$$

Thus we conclude that

$$\begin{aligned}\mathbb{E}_xT_0&=\lim_{n\to-\infty}\mathbb{E}_x\left[T_n\wedge T_0\right]=\lim_{n\to-\infty}u_{a_n}(x)\\&=\frac{x}{q-p}=\frac{|x|}{p-q}\text{ for }x<0.\end{aligned}$$

*Remark 19.50 (More on the boundary conditions).* If we were to use Corollary 19.45 directly to derive Eq. (19.60) in the case that $u(x):=\mathbb{E}_x\left[T_{\{n,0\}}\right]<\infty$ we for all $0\le x\le n$. we would find, for $x\neq 0$, that

$$u(x)=\sum_{y\notin\{n,0\}}q(x,y)\,u(y)+1$$

which implies that $u(x)$ satisfies Eq. (19.60) for $n<x<0$ provided $u(n)$ and $u(0)$ are taken to be equal to zero. Let us again choose $a$ and $b$

$$w(x):=(q-p)^{-1}x+a+b\,(q/p)^x$$

satisfies $w(0)=0$ and $w(-1)=u(-1)$. Then both $w$ and $u$ satisfy Eq. (19.60) for $n<x\le 0$ and agree at 0 and $-1$ and therefore are equal[7] for $n\le x\le 0$ and in particular $0=u(n)=w(n)$. Thus correct boundary conditions on $w$ in order for $w=u$ are $w(0)=w(n)=0$ as we have used above.

[7] Observe from Eq. (19.60) we have for $x\neq 0$ that,

$$u(x-1)=q^{-1}\left[u(x)-pu(x+1)-1\right].$$

From this equation it follows easily that $u(x)$ for $x\le 0$ is determined by its values at $x=0$ and $x=-1$.

**Definition 19.51.** *Suppose $(A,\mathcal{A})$ is a measurable space. A* ***sub-probability kernel*** *on $(A,\mathcal{A})$ is a function $\rho:A\times\mathcal{A}\to[0,1]$ such that $\rho(\cdot,C)$ is $\mathcal{A}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $C\in\mathcal{A}$ and $\rho(x,\cdot):\mathcal{A}\to[0,1]$ is a measure for all $x\in A$.*

As with probability kernels we will identify $\rho$ with the linear map, $\rho:\mathcal{A}_b\to\mathcal{A}_b$ given by

$$(\rho f)(x)=\rho(x,f)=\int_A f(y)\,\rho(x,dy).$$

Of course we have in mind that $\mathcal{A}=\mathcal{S}_A$ and $\rho=Q_A$. In the following lemma let $\|g\|_\infty:=\sup_{x\in A}|g(x)|$ for all $g\in\mathcal{A}_b$.

**Theorem 19.52.** *Let $\rho$ be a sub-probability kernel on a measurable space $(A,\mathcal{A})$ and define $u_n(x):=(\rho^n 1)(x)$ for all $x\in A$ and $n\in\mathbb{N}_0$. Then;*

1. *$u_n$ is a decreasing sequence so that $u:=\lim_{n\to\infty}u_n$ exists and is in $\mathcal{A}_b$. (When $\rho=Q_A$, $u_n(x)=P_x(T_B>n)\downarrow u(x)=P(T_B=\infty)$ as $n\to\infty$.)*
2. *The function $u$ satisfies $\rho u=u$.*
3. *If $w\in\mathcal{A}_b$ and $\rho w=w$ then $|w|\le\|w\|_\infty u$. In particular the equation, $\rho w=w$, has a non-zero solution $w\in\mathcal{A}_b$ iff $u\neq 0$.*
4. *If $u=0$ and $g\in\mathcal{A}_b$, then there is at most one $w\in\mathcal{A}_b$ such that $w=\rho w+g$.*
5. *Let*
$$U:=\sum_{n=0}^\infty u_n=\sum_{n=0}^\infty\rho^n 1:A\to[0,\infty]\tag{19.63}$$
*and suppose that $U(x)<\infty$ for all $x\in A$. Then for each $g\in\mathcal{S}_b$,*
$$w=\sum_{n=0}^\infty\rho^n g\tag{19.64}$$
*is absolutely convergent,*
$$|w|\le\|g\|_\infty U,\tag{19.65}$$
*$\rho(x,|w|)<\infty$ for all $x\in A$, and $w$ solves $w=\rho w+g$. Moreover if $v$ also solves $v=\rho v+g$ and $|v|\le CU$ for some $C<\infty$ then $v=w$.*
*Observe that when $\rho=Q_A$,*
$$U(x)=\sum_{n=0}^\infty P_x(T_B>n)=\sum_{n=0}^\infty\mathbb{E}_x\left(1_{T_B>n}\right)=\mathbb{E}_x\left(\sum_{n=0}^\infty 1_{T_B>n}\right)=\mathbb{E}_x\left[T_B\right].$$
6. *If $g:A\to[0,\infty]$ is any measurable function then*
$$w:=\sum_{n=0}^\infty\rho^n g:A\to[0,\infty]$$
*is a solution to $w=\rho w+g$. (It may be that $w\equiv\infty$ though!) Moreover if $v:A\to[0,\infty]$ satisfies $v=\rho v+g$ then $w\le v$. Thus $w$ is the minimal non-negative solution to $v=\rho v+g$.*

7. *If there exists $\alpha < 1$ such that $u \le \alpha$ on $A$ then $u = 0$. (When $\rho = Q_A$, this state that $P_x(T_B = \infty) \le \alpha$ for all $x \in A$ implies $P_x(T_A = \infty) = 0$ for all $x \in A$.)*
8. *If there exists an $\alpha < 1$ and an $n \in \mathbb{N}$ such that $u_n = \rho^n 1 \le \alpha$ on $A$, then there exists $C < \infty$ such that*
$$u_k(x) = (\rho^k 1)(x) \le C\beta^k \text{ for all } x \in A \text{ and } k \in \mathbb{N}_0$$
*where $\beta := \alpha^{1/n} < 1$. In particular, $U \le C(1-\beta)^{-1}$ and $u = 0$ under this assumption.*
*(When $\rho = Q_A$ this assertion states; if $P_x(T_B > n) \le \alpha$ for all $\alpha \in A$, then $P_x(T_B > k) \le C\beta^k$ and $\mathbb{E}_x T_B \le C(1-\beta)^{-1}$ for all $k \in \mathbb{N}_0$.)*

**Proof.** We will prove each item in turn.

1. First observe that $u_1(x) = \rho(x, A) \le 1 = u_0(x)$ and therefore,
$$u_{n+1} = \rho^{n+1} 1 = \rho^n u_1 \le \rho^n 1 = u_n.$$
We now let $u := \lim_{n\to\infty} u_n$ so that $u : A \to [0,1]$.
2. Using DCT we may let $n \to \infty$ in the identity, $\rho u_n = u_{n+1}$ in order to show $\rho u = u$.
3. If $w \in \mathcal{A}_b$ with $\rho w = w$, then
$$|w| = |\rho^n w| \le \rho^n |w| \le \|w\|_\infty \rho^n 1 = \|w\|_\infty \cdot u_n.$$
Letting $n \to \infty$ shows that $|w| \le \|w\|_\infty u$.
4. If $w_i \in \mathcal{A}_b$ solves $w_i = \rho w_i + g$ for $i = 1, 2$ then $w := w_2 - w_1$ satisfies $w = \rho w$ and therefore $|w| \le Cu = 0$.
5. Let $U := \sum_{n=0}^\infty u_n = \sum_{n=0}^\infty \rho^n 1 : A \to [0, \infty]$ and suppose $U(x) < \infty$ for all $x \in A$. Then $u_n(x) \to 0$ as $n \to \infty$ and so bounded solutions to $\rho u = u$ are necessarily zero. Moreover we have, for all $k \in \mathbb{N}_0$, that
$$\rho^k U = \sum_{n=0}^\infty \rho^k u_n = \sum_{n=0}^\infty u_{n+k} = \sum_{n=k}^\infty u_n \le U. \tag{19.66}$$
Since the tails of convergent series tend to zero it follows that $\lim_{k\to\infty} \rho^k U = 0$.
Now if $g \in \mathcal{S}_b$, we have
$$\sum_{n=0}^\infty |\rho^n g| \le \sum_{n=0}^\infty \rho^n |g| \le \sum_{n=0}^\infty \rho^n \|g\|_\infty = \|g\|_\infty \cdot U < \infty \tag{19.67}$$
and therefore $\sum_{n=0}^\infty \rho^n g$ is absolutely convergent. Making use of Eqs. (19.66) and (19.67) we see that
$$\sum_{n=1}^\infty \rho |\rho^n g| \le \|g\|_\infty \cdot \rho U \le \|g\|_\infty U < \infty$$
and therefore (using DCT),
$$w = \sum_{n=0}^\infty \rho^n g = g + \sum_{n=1}^\infty \rho^n g = g + \rho \sum_{n=1}^\infty \rho^{n-1} g = g + \rho w,$$
i.e. $w$ solves $w = g + \rho w$.
If $v : A \to \mathbb{R}$ is measurable such that $|v| \le CU$ and $v = g + \rho v$, then $y := w - v$ solves $y = \rho y$ with $|y| \le (C + \|g\|_\infty) U$. It follows that
$$|y| = |\rho^n y| \le (C + \|g\|_\infty) \rho^n U \to 0 \text{ as } n \to \infty,$$
i.e. $0 = y = w - v$.
6. If $g \ge 0$ we may always define $w$ by Eq. (19.64) allowing for $w(x) = \infty$ for some or even all $x \in A$. As in the proof of the previous item (with DCT being replaced by MCT), it follows that $w = \rho w + g$. If $v \ge 0$ also solves $v = g + \rho v$, then
$$v = g + \rho(g + \rho v) = g + \rho g + \rho^2 v$$
and more generally by induction we have
$$v = \sum_{k=0}^n \rho^k g + \rho^{n+1} v \ge \sum_{k=0}^n \rho^k g.$$
Letting $n \to \infty$ in this last equation shows that $v \ge w$.
7. If $u \le \alpha < 1$ on $A$, then by item 3. with $w = u$ we find that
$$u \le \|u\|_\infty \cdot u \le \alpha u$$
which clearly implies $u = 0$.
8. If $u_n \le \alpha < 1$, then for any $m \in \mathbb{N}$ we have,
$$u_{n+m} = \rho^m u_n \le \alpha \rho^m 1 = \alpha u_m.$$
Taking $m = kn$ in this inequality shows, $u_{(k+1)n} \le \alpha u_{kn}$. Thus a simple induction argument shows $u_{kn} \le \alpha^k$ for all $k \in \mathbb{N}_0$. For general $l \in \mathbb{N}_0$ we write $l = kn + r$ with $0 \le r < n$. We then have,
$$u_l = u_{kn+r} \le u_{kn} \le \alpha^k = \alpha^{\frac{l-r}{n}} = C\alpha^{l/n}$$
where $C = \alpha^{-\frac{n-1}{n}}$.

$\blacksquare$

**Corollary 19.53.** *If $h : B \to [0,\infty]$ is measurable, then $u(x) := \mathbb{E}_x[h(X_{T_B}) : T_B < \infty]$ is the unique minimal non-negative solution to Eq. (19.51) while if $g : A \to [0,\infty]$ is measurable, then $u(x) = \mathbb{E}_x\left[\sum_{n<T_B} g(X_n)\right]$ is the unique minimal non-negative solution to Eq. (19.52).*

**Exercise 19.13.** Keeping the notation of Example 19.48 and 19.49. Use Corollary 19.53 to show again that $P_x(T_B < \infty) = (q/p)^x$ for all $x > 0$ and $\mathbb{E}_x T_0 = x/(q-p)$ for $x < 0$. You should do so without making use of the extraneous hitting times, $T_n$ for $n \neq 0$.

**Corollary 19.54.** *If $P_x(T_B = \infty) = 0$ for all $x \in A$ and $h : B \to \mathbb{R}$ is a bounded measurable function, then $u(x) := \mathbb{E}_x[h(X_{T_B})]$ is the* ***unique*** *solution to Eq. (19.51).*

**Corollary 19.55.** *Suppose now that $A = B^c$ is a finite subset of $S$ and there exists an $\alpha \in (0,1)$ such that $P_x(T_B = \infty) \leq \alpha$ for all $x \in A$. Then there exists $C < \infty$ and $\beta \in (0,1)$ such that $P_x(T_B > n) \leq C\beta^n$. In particular $\mathbb{E}_x T_B < \infty$ for all $x \in A$.*

**Proof.** We know that

$$\lim_{n\to\infty} P_x(T_B > n) = P_x(T_B = \infty) \leq \alpha \text{ for all } x \in A.$$

Therefore if $\tilde{\alpha} \in (\alpha, 1)$, using the fact that $A$ is a finite set, there exists an $n$ sufficiently large such that $P_x(T_B > n) \leq \tilde{\alpha}$ for all $x \in A$. The result now follows from item 8. of Theorem 19.52. $\blacksquare$

**Definition 19.56 (First return time).** *For any $x \in S$, let $\tau_x := \min\{n \geq 1 : X_n = x\}$ where the minimum of the empty set is defined to be $\infty$.*

On the event $\{X_0 \neq x\}$ we have $\tau_x = T_x := \min\{n \geq 0 : X_n = x\}$ – the first hitting time of $x$. So $\tau_x$ is really manufactured for the case where $X_0 = x$ in which case $T_x = 0$ while $\tau_x$ is the *first return time* to $x$.

**Exercise 19.14.** Let $x \in X$. Show;

a for all $n \in \mathbb{N}_0$,

$$P_x(\tau_x > n+1) \leq \sum_{y\neq x} p(x,y) P_y(T_x > n). \tag{19.68}$$

b Use Eq. (19.68) to conclude that if $P_y(T_x = \infty) = 0$ for all $y \neq x$ then $P_x(\tau_x = \infty) = 0$, i.e. $\{X_n\}$ will return to $x$ when started at $x$.



c Sum Eq. (19.68) on $n \in \mathbb{N}_0$ to show

$$\mathbb{E}_x[\tau_x] \leq P_x(\tau_x > 0) + \sum_{y\neq x} p(x,y)\mathbb{E}_y[T_x]. \tag{19.69}$$

d Now suppose that $S$ is a finite set and $P_y(T_x = \infty) < 1$ for all $y \neq x$, i.e. there is a positive chance of hitting $x$ from any $y \neq x$ in $S$. Explain how Eq. (19.69) combined with Corollary 19.55 shows that $\mathbb{E}_x[\tau_x] < \infty$.

## 19.6 Finite state space chains

In this subsection I would like to write out the above theorems in the special case where $S$ is a finite set. In this case we will let $q(x,y) := Q(x,\{y\})$ so that

$$(Qf)(x) = \sum_{y\in S} q(x,y) f(y).$$

Thus if we view $f : S \to \mathbb{R}$ as a column vector and $Q$ to be the matrix with $q(x,y)$ in the $x^{\text{th}}$ – row and $y^{\text{th}}$ – column, then $Qf$ is simply matrix multiplication. As above we now suppose that $S$ is partitioned into two nonempty subsets $B$ and $A = B^c$. We further assume that $P_x(T_B < \infty) > 0$ for all $x \in A$, i.e. it is possible with positive probability for the chain $\{X_n\}_{n=0}^\infty$ to visit $B$ when started from any point in $A$. Because of Corollary 19.55 we know that in fact there exists $C < \infty$ and $\beta \in (0,1)$ such that $P_x(T_B > n) \leq C\beta^n$ for all $n \in \mathbb{N}_0$. In particular it follows that $\mathbb{E}_x T_B < \infty$ and $P_x(T_B < \infty) = 1$ for all $x \in A$.

If we let $Q_A = Q_{A,A}$ be the matrix with entries, $Q_A = (q(x,y))_{x,y\in A}$ and $I$ be the corresponding identity matrix, then $(Q_A - I)^{-1}$ exits according to Theorem 19.52. Let us further let $R = Q_{A,B}$ be the matrix with entries, $(q(x,y))_{x\in A \text{ and } y\in B}$. Thus $Q$ decomposes as

$$Q = \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} Q_A & R \\ * & * \end{bmatrix} \end{array} \begin{array}{c} \\ A \\ B \end{array}.$$

To summarize, $Q_A$ is $Q$ with the rows and columns indexed by $B$ deleted and $R$ is the $Q$ – matrix with the columns indexed by $A$ deleted and rows indexed by $B$ being deleted. Given a function $h : B \to \mathbb{R}$ let $(Rh)(x) = \sum_{y\in B} q(x,y) h(y)$ for all $x \in A$ which again may be thought of as matrix multiplication.

**Theorem 19.57.** *Let us continue to use the notation and assumptions as described above. If $h : B \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are given functions, then for all $x \in A$ we have;*

$$\mathbb{E}_x \left[ h\left(X_{T_B}\right)\right] = \left[ (I - Q_A)^{-1} Rh \right](x) \text{ and}$$
$$\mathbb{E}_x \left[ \sum_{n<T_B} g\left(X_n\right) \right] = \left[ (I - Q_A)^{-1} g \right](x).$$

*Remark 19.58.* Here is a story to go along with the above scenario. Suppose that $g(x)$ is the toll you have to pay for visiting a site $x \in A$ while $h(y)$ is the amount of prize money you get when landing on a point in $B$. Then $\mathbb{E}_x \left[\sum_{0 \le n < T} g(X_n)\right]$ is the expected toll you have to pay before your first exit from $A$ while $\mathbb{E}_x \left[h\left(X_T\right)\right]$ is your expected winnings upon exiting $B$.

Here are some typical choices for $h$ and $g$.

1. If $y \in B$ and $h = \delta_y$, then
$$P_x\left(X_{T_B} = y\right) = \left[ (I - Q_A)^{-1} R\delta_y \right](x) = \left[ (I - Q_A)^{-1} R \right]_{x,y}.$$
2. If $y \in A$ and $g = \delta_y$, then
$$\sum_{n<T_B} g\left(X_n\right) = \sum_{n<T_B} \delta_y\left(X_n\right) = \# \text{ visits to before hitting } B$$
and hence
$$\mathbb{E}_x\left(\# \text{ visits to before hitting } B\right) = \left[ (I - Q_A)^{-1} \delta_y \right](x)$$
$$= (I - Q_A)^{-1}_{xy}.$$
3. If $g = \mathbf{1}$, i.e. $g(y) = 1$ for all $y \in A$, then $\sum_{n<T_B} g\left(X_n\right) = T_B$ and we find,
$$\mathbb{E}_x T_B = \left[ (I - Q_A)^{-1} \mathbf{1} \right]_x = \sum_{y \in A} (I - Q_A)^{-1}_{xy},$$
where $\mathbb{E}_x T_B$ is the expected hitting time of $B$ when starting from $x$.

*Example 19.59.* Let us continue the rat in the maze Exercise 19.5 and now suppose that room 3 contains food while room 7 contains a mouse trap.

$$\begin{bmatrix} 1 & 2 & 3 \text{ (food)} \\ 4 & 5 & 6 \\ 7 \text{ (trap)} & & \end{bmatrix}.$$

We would like to compute the probability that the rat reaches the food before he is trapped. To answer this question we let $A = \{1, 2, 4, 5, 6\}$, $B = \{3, 7\}$, and $T := T_B$ be the first hitting time of $B$. Then deleting the 3 and 7 rows of $q$ in Eq. (??) leaves the matrix,

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix}$$
$$\begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix}.$$

Deleting the 3 and 7 columns from this matrix gives

$$Q_A = \begin{matrix} 1 & 2 & 4 & 5 & 6 \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix}$$

and deleting the $1, 2, 4, 5,$ and 6 columns gives

$$R = Q_{A,B} = \begin{matrix} 3 & 7 \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix}.$$

Therefore,

$$I - Q_A = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix},$$

and using a computer algebra package we find

$$(I - Q_A)^{-1} = \begin{matrix} 1 & 2 & 4 & 5 & 6 \\ \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{matrix}.$$

In particular we may conclude,

$$\begin{bmatrix} \mathbb{E}_1 T \\ \mathbb{E}_2 T \\ \mathbb{E}_4 T \\ \mathbb{E}_5 T \\ \mathbb{E}_6 T \end{bmatrix} = (I - Q_A)^{-1} \mathbf{1} = \begin{bmatrix} \frac{17}{3} \\ \frac{14}{3} \\ \frac{14}{3} \\ \frac{16}{3} \\ \frac{11}{3} \end{bmatrix},$$

and

$$\begin{bmatrix} P_1(X_T = 3) & P_1(X_T = 7) \\ P_2(X_T = 3) & P_2(X_T = 3) \\ P_4(X_T = 3) & P_4(X_T = 3) \\ P_5(X_T = 3) & P_5(X_T = 3) \\ P_6(X_T = 3) & P_6(X_T = 7) \end{bmatrix} = (I - Q_A)^{-1} R = \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

.

Since the event of hitting 3 before 7 is the same as the event $\{X_T = 3\}$, the desired hitting probabilities are

$$\begin{bmatrix} P_1(X_T = 3) \\ P_2(X_T = 3) \\ P_4(X_T = 3) \\ P_5(X_T = 3) \\ P_6(X_T = 3) \end{bmatrix} = \begin{bmatrix} \frac{7}{12} \\ \frac{3}{4} \\ \frac{5}{12} \\ \frac{2}{3} \\ \frac{5}{6} \end{bmatrix}.$$

We can also derive these hitting probabilities from scratch using the first step analysis. In order to do this let

$$h_i = P_i(X_T = 3) = P_i(X_n \text{ hits 3 (food) before 7(trapped)}).$$

By the first step analysis we will have,

$$\begin{aligned} h_i &= \sum_j P_i(X_T = 3 | X_1 = j) P_i(X_1 = j) \\ &= \sum_j q(i,j) P_i(X_T = 3 | X_1 = j) \\ &= \sum_j q(i,j) P_j(X_T = 3) \\ &= \sum_j q(i,j) h_j \end{aligned}$$

where $h_3 = 1$ and $h_7 = 0$. Looking at the jump diagram (Figure 19.3) we easily find



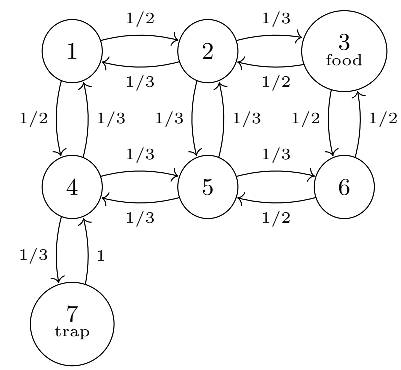



**Fig. 19.3.** The jump diagram for our proverbial rat in the maze.

$$\begin{aligned} h_1 &= \frac{1}{2}(h_2 + h_4) \\ h_2 &= \frac{1}{3}(h_1 + h_3 + h_5) = \frac{1}{3}(h_1 + 1 + h_5) \\ h_4 &= \frac{1}{3}(h_1 + h_5 + h_7) = \frac{1}{3}(h_1 + h_5) \\ h_5 &= \frac{1}{3}(h_2 + h_4 + h_6) \\ h_6 &= \frac{1}{2}(h_3 + h_5) = \frac{1}{2}(1 + h_5) \end{aligned}$$

and the solutions to these equations are (as seen before) given by

$$\left[ h_1 = \frac{7}{12}, h_2 = \frac{3}{4}, h_4 = \frac{5}{12}, h_5 = \frac{2}{3}, h_6 = \frac{5}{6} \right]. \tag{19.70}$$

Similarly, if

$$k_i := P_i(X_T = 7) = P_i(X_n \text{ is trapped before dinner}),$$

we need only use the above equations with $h$ replaced by $k$ and now taking $k_3 = 0$ and $k_7 = 1$ to find,

$$k_1 = \frac{1}{2}(k_2 + k_4)$$
$$k_2 = \frac{1}{3}(k_1 + k_5)$$
$$k_4 = \frac{1}{3}(k_1 + k_5 + 1)$$
$$k_5 = \frac{1}{3}(k_2 + k_4 + k_6)$$
$$k_6 = \frac{1}{2}k_5$$

and then solve to find,

$$\left[k_1 = \frac{5}{12}, k_2 = \frac{1}{4}, k_4 = \frac{7}{12}, k_5 = \frac{1}{3}, k_6 = \frac{1}{6}\right]. \tag{19.71}$$

Notice that the sum of the hitting probabilities in Eqs. (19.70) and (19.71) add up to 1 as they should.

### 19.6.1 Invariant distributions and return times

For this subsection suppose that $S = \{1, 2, \ldots, n\}$ and $Q_{ij}$ is a Markov matrix. To each state $i \in S$, let

$$\tau_i := \min\{n \geq 1 : X_n = i\} \tag{19.72}$$

be the **first passage time of the chain to site** $i$.

**Proposition 19.60.** *The Markov matrix $Q$ has an invariant distribution.*

**Proof.** If $\mathbf{1} := \left[1\ 1\ \ldots\ 1\right]^{\text{tr}}$, then $Q\mathbf{1} = \mathbf{1}$ from which it follows that

$$0 = \det(Q - I) = \det\left(Q^{\text{tr}} - I\right).$$

Therefore there exists a non-zero row vector $\nu$ such that $Q^{\text{tr}}\nu^{\text{tr}} = \nu^{\text{tr}}$ or equivalently that $\nu Q = \nu$. At this point we would be done if we knew that $\nu_i \geq 0$ for all $i$ – but we don't. So let $\pi_i := |\nu_i|$ and observe that

$$\pi_i = |\nu_i| = \left|\sum_{k=1}^n \nu_k Q_{ki}\right| \leq \sum_{k=1}^n |\nu_k| Q_{ki} \leq \sum_{k=1}^n \pi_k Q_{ki}.$$

We now claim that in fact $\pi = \pi Q$. If this were not the case we would have $\pi_i < \sum_{k=1}^n \pi_k Q_{ki}$ for some $i$ and therefore

$$0 < \sum_{i=1}^n \pi_i < \sum_{i=1}^n \sum_{k=1}^n \pi_k Q_{ki} = \sum_{k=1}^n \sum_{i=1}^n \pi_k Q_{ki} = \sum_{k=1}^n \pi_k$$

which is a contradiction. So all that is left to do is normalize $\pi_i$ so $\sum_{i=1}^n \pi_i = 1$ and we are done. ■

We are now going to assume that $Q$ is **irreducible** which means that for all $i \neq j$ there exists $n \in \mathbb{N}$ such that $Q_{ij}^n > 0$. Alternatively put this implies that $P_i(T_j < \infty) = P_i(\tau_j < \infty) > 0$ for all $i \neq j$. By Corollary 19.55 we know that $\mathbb{E}_i[\tau_j] = \mathbb{E}_i T_j < \infty$ for all $i \neq j$ and it is not too hard to see that $\mathbb{E}_i \tau_i < \infty$ also holds. The fact that $\mathbb{E}_i \tau_i < \infty$ for all $i \in S$ will come out of the proof of the next proposition as well.

**Proposition 19.61.** *If $Q$ is irreducible, then there is precisely one invariant distribution, $\pi$, which is given by $\pi_i = 1/(\mathbb{E}_i \tau_i) > 0$ for all $i \in S$.*

**Proof.** We begin by using the first step analysis to write equations for $\mathbb{E}_i[\tau_j]$ as follows:

$$\begin{aligned}\mathbb{E}_i[\tau_j] &= \sum_{k=1}^n \mathbb{E}_i[\tau_j | X_1 = k] Q_{ik} = \sum_{k \neq j} \mathbb{E}_i[\tau_j | X_1 = k] Q_{ik} + Q_{ij} 1 \\ &= \sum_{k \neq j} (\mathbb{E}_k[\tau_j] + 1) Q_{ik} + Q_{ij} 1 = \sum_{k \neq j} \mathbb{E}_k[\tau_j] Q_{ik} + 1.\end{aligned}$$

and therefore,

$$\mathbb{E}_i[\tau_j] = \sum_{k \neq j} Q_{ik} \mathbb{E}_k[\tau_j] + 1. \tag{19.73}$$

Now suppose that $\pi$ is any invariant distribution for $Q$, then multiplying Eq. (19.73) by $\pi_i$ and summing on $i$ shows

$$\begin{aligned}\sum_{i=1}^n \pi_i \mathbb{E}_i[\tau_j] &= \sum_{i=1}^n \pi_i \sum_{k \neq j} Q_{ik} \mathbb{E}_k[\tau_j] + \sum_{i=1}^n \pi_i 1 \\ &= \sum_{k \neq j} \pi_k \mathbb{E}_k[\tau_j] + 1.\end{aligned}$$

Since $\sum_{k \neq j} \pi_k \mathbb{E}_k[\tau_j] < \infty$ we may cancel it from both sides of this equation in order to learn $\pi_j \mathbb{E}_j[\tau_j] = 1$. ■

We may use Eq. (19.73) to compute $\mathbb{E}_i[\tau_j]$ in examples. To do this, fix $j$ and set $v_i := \mathbb{E}_i \tau_j$. Then Eq. (19.73) states that $v = Q^{(j)} v + \mathbf{1}$ where $Q^{(j)}$ denotes $Q$ with the $j^{th}$ – column replaced by all zeros. Thus we have

$$(\mathbb{E}_i \tau_j)_{i=1}^n = \left(I - Q^{(j)}\right)^{-1} \mathbf{1}, \tag{19.74}$$

i.e.

$$\begin{bmatrix} \mathbb{E}_1 \tau_j \\ \vdots \\ \mathbb{E}_n \tau_j \end{bmatrix} = \left(I - Q^{(j)}\right)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{19.75}$$

### 19.6.2 Some worked examples

*Example 19.62.* Let $S = \{1, 2\}$ and $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with jump diagram in Figure 19.4. In this case $Q^{2n} = I$ while $Q^{2n+1} = Q$ and therefore $\lim_{n\to\infty} Q^n$ does not

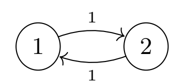



**Fig. 19.4.** A non-random chain.

exist. On the other hand it is easy to see that the invariant distribution, $\pi$, for $Q$ is $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ and, moreover,

$$\frac{Q + Q^2 + \cdots + Q^N}{N} \to \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix}.$$

Let us compute

$$\begin{bmatrix} \mathbb{E}_1\tau_1 \\ \mathbb{E}_2\tau_1 \end{bmatrix} = \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbb{E}_1\tau_2 \\ \mathbb{E}_2\tau_2 \end{bmatrix} = \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

so that indeed, $\pi_1 = 1/\mathbb{E}_1\tau_1$ and $\pi_2 = 1/\mathbb{E}_2\tau_2$. Of course $\tau_1 = 2$ ($P_1$ -a.s.) and $\tau_2 = 2$ ($P_2$ -a.s.) so that it is obvious that $\mathbb{E}_1\tau_1 = \mathbb{E}_2\tau_2 = 2$.

*Example 19.63.* Again let $S = \{1, 2\}$ and $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with jump diagram in Figure 19.5. In this case the chain is not irreducible and every $\pi = [a \ b]$ with


**Fig. 19.5.** A simple non-irreducible chain.

$a + b = 1$ and $a, b \geq 0$ is an invariant distribution.

*Example 19.64.* Suppose that $S = \{1, 2, 3\}$, and

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

has the jump graph given by 19.6. Notice that $Q_{11}^2 > 0$ and $Q_{11}^3 > 0$ that $Q$ is

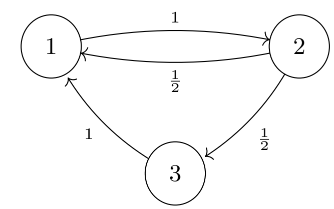



**Fig. 19.6.** A simple 3 state jump diagram.

"aperiodic." We now find the invariant distribution,

$$\mathrm{Nul}\,(Q - I)^{\mathrm{tr}} = \mathrm{Nul}\begin{bmatrix} -1 & \frac{1}{2} & 1 \\ 1 & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{bmatrix} = \mathbb{R}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

Therefore the invariant distribution is given by

$$\pi = \frac{1}{5}\begin{bmatrix} 2 & 2 & 1 \end{bmatrix}.$$

Let us now observe that

$$Q^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$Q^3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$Q^{20} = \begin{bmatrix} \frac{409}{1024} & \frac{205}{512} & \frac{205}{1024} \\ \frac{205}{512} & \frac{409}{1024} & \frac{205}{1024} \\ \frac{205}{512} & \frac{205}{512} & \frac{51}{256} \end{bmatrix} = \begin{bmatrix} 0.399\,41 & 0.400\,39 & 0.200\,20 \\ 0.400\,39 & 0.399\,41 & 0.200\,20 \\ 0.400\,39 & 0.400\,39 & 0.199\,22 \end{bmatrix}.$$

Let us also compute $\mathbb{E}_2\tau_3$ via,

$$\begin{bmatrix} \mathbb{E}_1\tau_3 \\ \mathbb{E}_2\tau_3 \\ \mathbb{E}_3\tau_3 \end{bmatrix} = \left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$$

so that

$$\frac{1}{\mathbb{E}_3\tau_3} = \frac{1}{5} = \pi_3.$$

*Example 19.65.* The transition matrix,

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

is represented by the jump diagram in Figure 19.7. This chain is aperiodic. We

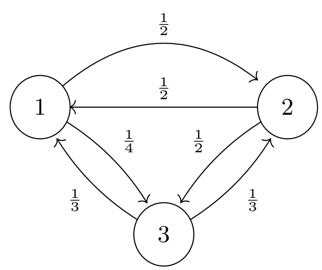



**Fig. 19.7.** In the above diagram there are jumps from 1 to 1 with probability 1/4 and jumps from 3 to 3 with probability 1/3 which are not explicitly shown but must be inferred by conservation of probability.

find the invariant distribution as,

$$\begin{aligned} \mathrm{Nul}\,(Q-I)^{\mathrm{tr}} &= \mathrm{Nul}\left(\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)^{\mathrm{tr}} \\ &= \mathrm{Nul}\left(\begin{bmatrix} -\frac{3}{4} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & -1 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & -\frac{2}{3} \end{bmatrix}\right) = \mathbb{R}\begin{bmatrix} 1 \\ \frac{5}{6} \\ 1 \end{bmatrix} = \mathbb{R}\begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix} \end{aligned}$$

$$\pi = \frac{1}{17}\begin{bmatrix} 6 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 0.352\,94 & 0.294\,12 & 0.352\,94 \end{bmatrix}.$$

In this case

$$Q^{10} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}^{10} = \begin{bmatrix} 0.352\,98 & 0.294\,04 & 0.352\,98 \\ 0.352\,89 & 0.294\,23 & 0.352\,89 \\ 0.352\,95 & 0.294\,1 & 0.352\,95 \end{bmatrix}.$$

Let us also compute

$$\begin{bmatrix} \mathbb{E}_1\tau_2 \\ \mathbb{E}_2\tau_2 \\ \mathbb{E}_3\tau_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 0 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{11}{5} \\ \frac{17}{5} \\ \frac{13}{5} \end{bmatrix}$$

so that

$$1/\mathbb{E}_2\tau_2 = 5/17 = \pi_2.$$

*Example 19.66.* Consider the following Markov matrix,

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 0 & 3/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix}$$

with jump diagram in Figure 19.8. Since this matrix is doubly stochastic (i.e

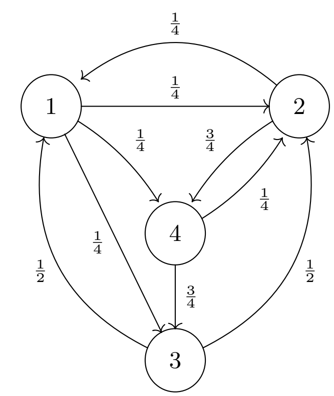



**Fig. 19.8.** The jump diagram for $Q$.

$\sum_{i=1}^{4} Q_{ij} = 1$ for all $j$ as well as $\sum_{j=1}^{4} Q_{ij} = 1$ for all $i$), it is easy to check that $\pi = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. Let us compute $\mathbb{E}_3\tau_3$ as follows

$$\begin{bmatrix} \mathbb{E}_1\tau_3 \\ \mathbb{E}_2\tau_3 \\ \mathbb{E}_3\tau_3 \\ \mathbb{E}_4\tau_3 \end{bmatrix} = \left( \begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{bmatrix} - \begin{bmatrix} 1/4&1/4&0&1/4 \\ 1/4&0&0&3/4 \\ 1/2&1/2&0&0 \\ 0&1/4&0&0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{50}{17} \\ \frac{52}{17} \\ 4 \\ \frac{30}{17} \end{bmatrix}$$

so that $\mathbb{E}_3\tau_3 = 4 = 1/\pi_4$ as it should be. Similarly,

$$\begin{bmatrix} \mathbb{E}_1\tau_2 \\ \mathbb{E}_2\tau_2 \\ \mathbb{E}_3\tau_2 \\ \mathbb{E}_4\tau_2 \end{bmatrix} = \left( \begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{bmatrix} - \begin{bmatrix} 1/4&0&1/4&1/4 \\ 1/4&0&0&3/4 \\ 1/2&0&0&0 \\ 0&0&3/4&0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{54}{17} \\ 4 \\ \frac{44}{17} \\ \frac{50}{17} \end{bmatrix}$$

and again $\mathbb{E}_2\tau_2 = 4 = 1/\pi_2$.

### 19.6.3 Exercises

**Exercise 19.15 (2nd order recurrence relations).** Let $a, b, c$ be real numbers with $a \neq 0 \neq c$, $\alpha, \beta \in \mathbb{Z} \cup \{\pm\infty\}$ with $\alpha < \beta$, and suppose $\{u(x) : x \in [\alpha, \beta] \cap \mathbb{Z}\}$ solves the second order homogeneous recurrence relation:

$$au(x+1) + bu(x) + cu(x-1) = 0 \tag{19.76}$$

for $\alpha < x < \beta$. Show:

1. for any $\lambda \in \mathbb{C}$,
$$a\lambda^{x+1} + b\lambda^x + c\lambda^{x-1} = \lambda^{x-1} p(\lambda) \tag{19.77}$$
where $p(\lambda) = a\lambda^2 + b\lambda + c$ is the **characteristic polynomial** associated to Eq. (19.76).
Let $\lambda_\pm = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ be the roots of $p(\lambda)$ and suppose for the moment that $b^2 - 4ac \neq 0$. From Eq. (19.76) it follows that for any choice of $A_\pm \in \mathbb{R}$, the function,
$$w(x) := A_+\lambda_+^x + A_-\lambda_-^x,$$
solves Eq. (19.76) for all $x \in \mathbb{Z}$.
2. Show there is a unique choice of constants, $A_\pm \in \mathbb{R}$, such that the function $u(x)$ is given by
$$u(x) := A_+\lambda_+^x + A_-\lambda_-^x \text{ for all } \alpha \le x \le \beta.$$
3. Now suppose that $b^2 = 4ac$ and $\lambda_0 := -b/(2a)$ is the double root of $p(\lambda)$. Show for any choice of $A_0$ and $A_1$ in $\mathbb{R}$ that
$$w(x) := (A_0 + A_1 x)\lambda_0^x$$
solves Eq. (19.76) for all $x \in \mathbb{Z}$. **Hint:** Differentiate Eq. (19.77) with respect to $\lambda$ and then set $\lambda = \lambda_0$.
4. Again show that any function $u$ solving Eq. (19.76) is of the form $u(x) = (A_0 + A_1 x)\lambda_0^x$ for $\alpha \le x \le \beta$ for some unique choice of constants $A_0, A_1 \in \mathbb{R}$.

In the next couple of exercises you are going to use first step analysis to show that a simple unbiased random walk on $\mathbb{Z}$ is null recurrent. We let $\{X_n\}_{n=0}^\infty$ be the Markov chain with values in $\mathbb{Z}$ with transition probabilities given by

$$P(X_{n+1} = x \pm 1 | X_n = x) = 1/2 \text{ for all } n \in \mathbb{N}_0 \text{ and } x \in \mathbb{Z}.$$

Further let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and

$$T_{a,b} := \min\{n : X_n \in \{a, b\}\} \text{ and } T_b := \inf\{n : X_n = b\}.$$

We know by Corollary[8] 19.55 that $\mathbb{E}_0[T_{a,b}] < \infty$ from which it follows that $P(T_{a,b} < \infty) = 1$ for all $a < 0 < b$.

**Exercise 19.16.** Let $w_x := P_x\left(X_{T_{a,b}} = b\right) := P\left(X_{T_{a,b}} = b | X_0 = x\right)$.

1. Use first step analysis to show for $a < x < b$ that
$$w_x = \frac{1}{2}(w_{x+1} + w_{x-1}) \tag{19.78}$$
provided we define $w_a = 0$ and $w_b = 1$.
2. Use the results of Exercise 19.15 to show
$$P_x\left(X_{T_{a,b}} = b\right) = w_x = \frac{1}{b-a}(x-a). \tag{19.79}$$
3. Let
$$T_b := \begin{cases} \min\{n : X_n = b\} & \text{if } \{X_n\} \text{ hits } b \\ \infty & \text{otherwise} \end{cases}$$
be the first time $\{X_n\}$ hits $b$. Explain why, $\left\{X_{T_{a,b}} = b\right\} \subset \{T_b < \infty\}$ and use this along with Eq. (19.79) to conclude[9] that $P_x(T_b < \infty) = 1$ for all $x < b$. (By symmetry this result holds true for all $x \in \mathbb{Z}$.)

[8] Apply this corollary to finite walk in $[a, b] \cap \mathbb{Z}$.
[9] The fact that $P_j(T_b < \infty) = 1$ is also follows from Example 12.58 above.

**Exercise 19.17.** The goal of this exercise is to give a second proof of the fact that $P_x(T_b < \infty) = 1$. Here is the outline:

1. Let $w_x := P_x(T_b < \infty)$. Again use first step analysis to show that $w_x$ satisfies Eq. (19.78) for all $x$ with $w_b = 1$.
2. Use Exercise 19.15 to show that there is a constant, $c$, such that
$$w_x = c(x-b) + 1 \text{ for all } x \in \mathbb{Z}.$$
3. Explain why $c$ must be zero to again show that $P_x(T_b < \infty) = 1$ for all $x \in \mathbb{Z}$.

**Exercise 19.18.** Let $T = T_{a,b}$ and $u_x := \mathbb{E}_x T := \mathbb{E}[T|X_0 = x]$.

1. Use first step analysis to show for $a < x < b$ that
$$u_x = \frac{1}{2}(u_{x+1} + u_{x-1}) + 1 \tag{19.80}$$
with the convention that $u_a = 0 = u_b$.
2. Show that
$$u_x = A_0 + A_1 x - x^2 \tag{19.81}$$
solves Eq. (19.80) for any choice of constants $A_0$ and $A_1$.
3. Choose $A_0$ and $A_1$ so that $u_x$ satisfies the boundary conditions, $u_a = 0 = u_b$. Use this to conclude that
$$\mathbb{E}_x T_{a,b} = -ab + (b+a)x - x^2 = -a(b-x) + bx - x^2. \tag{19.82}$$

*Remark 19.67.* Notice that $T_{a,b} \uparrow T_b = \inf\{n : X_n = b\}$ as $a \downarrow -\infty$, and so passing to the limit as $a \downarrow -\infty$ in Eq. (19.82) shows
$$\mathbb{E}_x T_b = \infty \text{ for all } x < b.$$

Combining the last couple of exercises together shows that $\{X_n\}$ is "null - recurrent."

**Exercise 19.19.** Let $T = T_b$. The goal of this exercise is to give a second proof of the fact and $u_x := \mathbb{E}_x T = \infty$ for all $x \neq b$. Here is the outline. Let $u_x := \mathbb{E}_x T \in [0,\infty] = [0,\infty) \cup \{\infty\}$.

1. Note that $u_b = 0$ and, by a first step analysis, that $u_x$ satisfies Eq. (19.80) for all $x \neq b$ – allowing for the possibility that some of the $u_x$ may be infinite.
2. Argue, using Eq. (19.80), that if $u_x < \infty$ for some $x < b$ then $u_y < \infty$ for all $y < b$. Similarly, if $u_x < \infty$ for some $x > b$ then $u_y < \infty$ for all $y > b$.
3. If $u_x < \infty$ for all $x > b$ then $u_x$ must be of the form in Eq. (19.81) for some $A_0$ and $A_1$ in $\mathbb{R}$ such that $u_b = 0$. However, this would imply, $u_x = \mathbb{E}_x T \to -\infty$ as $x \to \infty$ which is impossible since $\mathbb{E}_x T \geq 0$ for all $x$. Thus we must conclude that $\mathbb{E}_x T = u_x = \infty$ for all $x > b$. (A similar argument works if we assume that $u_x < \infty$ for all $x < b$.)

## 19.7 Appendix: Kolmogorov's extension theorem II

The Kolmogorov extension Theorem 11.58 generalizes to the case where $\mathbb{N}$ is replaced by an arbitrary index set, $T$. Let us set up the notation for this theorem. Let $T$ be an arbitrary index set, $\{(S_t, \mathcal{S}_t)\}_{t\in T}$ be a collection of standard Borel spaces, $S = \prod_{t\in T} S_t$, $\mathcal{S} := \otimes_{t\in T}\mathcal{S}_t$, and for $\Lambda \subset T$ let
$$\left(S_\Lambda := \prod_{t\in\Lambda} S_t, \mathcal{S}_\Lambda := \otimes_{t\in\Lambda}\mathcal{S}_t\right)$$
and $X_\Lambda : S \to S_\Lambda$ be the projection map, $X_\Lambda(x) := x|_\Lambda$. If $\Lambda \subset \Lambda' \subset T$, also let $X_{\Lambda,\Lambda'} : S_{\Lambda'} \to S_\Lambda$ be the projection map, $X_{\Lambda,\Lambda'}(x) := x|_\Lambda$ for all $x \in S_{\Lambda'}$.

**Theorem 19.68 (Kolmogorov).** *For each $\Lambda \subset_f T$ (i.e. $\Lambda \subset T$ and $\#(\Lambda) < \infty$), let $\mu_\Lambda$ be a probability measure on $(S_\Lambda, \mathcal{S}_\Lambda)$. We further suppose $\{\mu_\Lambda\}_{\Lambda\subset_f T}$ satisfy the following compatibility relations;*
$$\mu_{\Lambda'} \circ X_{\Lambda,\Lambda'}^{-1} = \mu_\Lambda \text{ for all } \Lambda \subset \Lambda' \subset_f T. \tag{19.83}$$
*Then there exists a unique probability measure, $P$, on $(S, \mathcal{S})$ such that $P \circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f T$.*

**Proof.** (For slight variation on the proof of this theorem given here, see Exercise 19.21.) Let
$$\mathcal{A} := \cup_{\Lambda\subset_f T} X_\Lambda^{-1}(\mathcal{S}_\Lambda)$$
and for $A = X_\Lambda^{-1}(A') \in \mathcal{A}$, let $P(A) := \mu_\Lambda(A')$. The compatibility conditions in Eq. (19.83) imply $P$ is a well defined finitely additive measure on the algebra, $\mathcal{A}$. We now complete the proof by showing $P$ is continuous on $\mathcal{A}$.

To this end, suppose $A_n := X_{\Lambda_n}^{-1}(A'_n) \in \mathcal{A}$ with $A_n \downarrow \emptyset$ as $n \to \infty$. Let $\Lambda := \cup_{n=1}^\infty \Lambda_n$ – a countable subset of $T$. Owing to Theorem 11.58, there is a unique probability measure, $P_\Lambda$, on $(S_\Lambda, \mathcal{S}_\Lambda)$ such that $P_\Lambda\left(X_\Gamma^{-1}(A)\right) = \mu_\Gamma(A)$ for all $\Gamma \subset_f \Lambda$ and $A \in \mathcal{S}_\Gamma$. Hence if we let $\tilde{A}_n := X_{\Lambda,\Lambda_n}^{-1}(A_n)$, we then have
$$P(A_n) = \mu_{\Lambda_n}(A'_n) = P_\Lambda\left(\tilde{A}_n\right)$$
with $\tilde{A}_n \downarrow \emptyset$ as $n \to \infty$. Since $P_\Lambda$ is a measure, we may conclude
$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P_\Lambda\left(\tilde{A}_n\right) = 0.$$
■

**Exercise 19.20.** Let us write $\Lambda \subset_c T$ to mean $\Lambda \subset T$ and $\Lambda$ is at most countable. Show
$$\mathcal{S} = \cup_{\Lambda\subset_c T} X_\Lambda^{-1}(\mathcal{S}_\Lambda). \tag{19.84}$$
**Hint:** Verify Eq. (19.84) by showing $\mathcal{S}_0 := \cup_{\Lambda\subset_c T} X_\Lambda^{-1}(\mathcal{S}_\Lambda)$ is a $\sigma$ – algebra.

**Exercise 19.21.** For each $\Lambda \subset T$, let $\mathcal{S}'_\Lambda := X_\Lambda^{-1}(\mathcal{S}_\Lambda) = \sigma(X_i : i \in \Lambda) \subset \mathcal{S}$. Show;

1. if $U, V \subset T$ then $\mathcal{S}'_U \cap \mathcal{S}'_V = \mathcal{S}'_{U\cap V}$.
2. By Theorem 11.58, if $U, V \subset_c T$, there exists unique probability measures, $P_U$ and $P_V$ on $\mathcal{S}'_U$ and $\mathcal{S}'_V$ respectively such that $P_U \circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f U$ and $P_V \circ X_\Lambda^{-1} = \mu_\Lambda$ for all $\Lambda \subset_f V$. Show $P_U = P_V$ on $\mathcal{S}'_U \cap \mathcal{S}'_V$. Hence for any $A \in \mathcal{S}$ we may define $P(A) := P_U(A)$ provided $A \in \mathcal{S}'_U$.
3. Show $P$ defined in the previous item is a countably additive measure on $\mathcal{S}$.

## 19.8 Removing the standard Borel restriction

**Theorem 19.69.** *Let $\{(S_n, \mathcal{S}_n)\}_{n\in\mathbb{N}_0}$ be a collection of measurable spaces, $S = \prod_{n=0}^\infty S_n$ and $\mathcal{S} := \otimes_{n=0}^\infty \mathcal{S}_n$. Moreover for each $n \in \mathbb{N}_0$ let $S^n := S_0 \times \cdots \times S_n$ and $\mathcal{S}^n := \mathcal{S}_0 \otimes \cdots \otimes \mathcal{S}_n$. We further suppose that $\mu_0$ is a given probability measure on $(S_0, \mathcal{S}_0)$ and $T_n : S^{n-1} \times \mathcal{S}_n \to [0,1]$ for $n = 1, 2, \dots$ are give probability kernels on $S^{n-1} \times S_n$. Finally let $\mu_n$ be the probability measure on $(S^n, \mathcal{S}^n)$ defined inductively by,*

$$\mu_n(dx_0, \dots, dx_n) = \mu_{n-1}(dx_0, \dots, dx_{n-1}) T_n(x_0, \dots, x_{n-1}, dx_n) \ \forall\, n \in \mathbb{N}. \tag{19.85}$$

*Then there exists a unique probability measure, $P$ on $(S, \mathcal{S})$ such that*

$$P(f) = \int_{S^n} F d\mu_n$$

*whenever $f(x) = F(x_0, \dots, x_n)$ for some $F \in (\mathcal{S}_n)_b$.*

*Remark 19.70 (Heuristic proof).* Before giving the formal proof of this theorem let me indicate the main ideas. Let $X_i : S \to S_i$ be the projection maps and $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$. If $P$ exists, then

$$\begin{aligned} P[F(X_0, \dots, X_{n+1}) | \mathcal{B}_n] &= T_n(X_0, \dots, X_n; F(X_0, \dots, X_n, \cdot)) \\ &= (TF(X_0, \dots, X_n, \cdot))(X_0, \dots, X_n). \end{aligned}$$

Indeed,

$$\begin{aligned} &\mathbb{E}[T_n(X_0, \dots, X_n; F(X_0, \dots, X_n, \cdot)) G(X_0, \dots, X_n)] \\ &= \int T_n(x_0, \dots, x_n, dx_{n+1}) F(x_0, \dots, x_n, x_{n+1}) G(x_0, \dots, x_n) d\mu_n(x_0, \dots, x_n) \\ &= \int F(x_0, \dots, x_n, x_{n+1}) G(x_0, \dots, x_n) d\mu_{n+1}(x_0, \dots, x_n, x_{n+1}) \\ &= \mathbb{E}[F(X_0, \dots, X_{n+1}) G(X_0, \dots, X_n)]. \end{aligned}$$

Now suppose that $f_n = F_n(X_0, \dots, X_n)$ is a decreasing sequence of functions such that $\lim_{n\to\infty} P(f_n) =: \varepsilon > 0$. Letting $f_\infty := \lim_{n\to\infty} f_n$ we would have $f_n \ge f_\infty$ for all $n$ and therefore $f_n \ge \mathbb{E}[f_\infty|\mathcal{B}_n] := \bar{f}_n$. We also use

$$\begin{aligned} &\bar{f}_n(X_0, X_1, \dots, X_n) \\ &\quad= \mathbb{E}[f_\infty|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[f_\infty|\mathcal{B}_{n+1}]|\mathcal{B}_n] \\ &\quad= \mathbb{E}[\bar{f}_{n+1}|\mathcal{B}_n] = \int \bar{f}_{n+1}(X_0, X_1, \dots, x_{n+1}) T_{n+1}(X_0, \dots, X_n, dx_{n+1}) \end{aligned}$$

and $P(\bar{f}_n) = P(f_\infty) = \lim_{m\to\infty} P(f_m) = \varepsilon > 0$ (we only use the case where $n = 0$ here). Since $P(\bar{f}_0(X_0)) = \varepsilon > 0$, there exists $x_0 \in S_0$ such that

$$\varepsilon \le \bar{f}_0(x_0) = \mathbb{E}[\bar{f}_1|\mathcal{B}_0] = \int \bar{f}_1(x_0, x_1) T_1(x_0, dx_1)$$

and so similarly there exists $x_1 \in S_1$ such that

$$\varepsilon \le \bar{f}_1(x_0, x_1) = \int \bar{f}_2(x_0, x_1, x_2) T_2(x_0, x_1, dx_2).$$

Again it follows that there must exists an $x_2 \in S_2$ such that $\varepsilon \le \bar{f}_2(x_0, x_1, x_2)$. We continue on this way to find and $x \in S$ such that

$$f_n(x) \ge \bar{f}_n(x_0, \dots, x_n) \ge \varepsilon \text{ for all } n.$$

Thus if $P(f_n) \downarrow \varepsilon > 0$ then $\lim_{n\to\infty} f_n(x) \ge \varepsilon \ne 0$ as desired.

**Proof.** Now onto the formal proof. Let $\mathbb{S}$ denote the space of finitely based bounded cylinder functions on $S$, i.e. functions of the form $f(x) = F(x_0, \dots, x_n)$ with $F \in \mathcal{S}_b^n$. For such an $f$ we define

$$I(f) := P_n(F).$$

It is easy to check that $I$ is a well defined positive linear functional on $\mathbb{S}$.

Now suppose that $0 \le f_n \in \mathbb{S}$ are forms a decreasing sequence of functions such that $\lim_{n\to\infty} I(f_n) = \varepsilon > 0$. We wish to show that $\lim_{n\to\infty} f_n(x) \ne 0$ for every $x \in S$. By assumption, $f_n(x) = F_n(x_0, \dots, x_{N_n})$ for some $N_n \in \mathbb{N}$ of which we may assume $N_0 < N_1 < N_2 < \dots$. Moreover if $N_0 = 2 < N_1 = 5 < N_2 = 7 < \dots$, we may replace $(f_0, f_1, \dots)$ by

$$(g_0, g_1, g_2, \dots) = (1, 1, f_0, f_0, f_0, f_1, f_1, f_2, \dots).$$

Noting that $\lim_{n\to\infty} g_n = \lim_{n\to\infty} f_n$, $\lim_{n\to\infty} I(g_n) = I(f_n)$, and $g_n(x) = G_n(x_0, \dots, x_n)$ for some $G_n \in \mathcal{S}_b^n$, we may now assume that $f_n(x) = F_n(x_0, \dots, x_n)$ with $F_n \in \mathcal{S}_b^n$.

For any $k \leq n$ let

$$F_n^k(x_0,\dots,x_k) := \int \cdots \int F_n(x_0,\dots,x_n) \prod_{l=k}^{n-1} T_l(x_0,\dots,x_l,dx_{l+1})$$

which is an explicit version of $P_n[F_n(x_0,\dots,x_n)|x_0,\dots,x_k]$"="$\mathbb{E}[f_n|\mathcal{B}_k](x)$. By construction of the measures $P_n$ it follows that

$$P_k F_n^k = P_n F_n = I(f_n) \text{ for all } k \leq n. \tag{19.86}$$

Since

$$F_n(x_0,\dots,x_n) = f_n(x) \leq f_{n+1}(x) = F_{n+1}(x_0,\dots,x_n,x_{n+1}),$$

it follows that

$$\begin{aligned} F_n^k(x_0,\dots,x_k) &= \int F_n(x_0,\dots,x_n) \prod_{l=k}^{n} T_l(x_0,\dots,x_l,dx_{l+1}) \\ &\leq \int F_{n+1}(x_0,\dots,x_n,x_{n+1}) \prod_{l=k}^{n} T_l(x_0,\dots,x_l,dx_{l+1}) \\ &= F_{n+1}^k(x_0,\dots,x_k). \end{aligned}$$

Thus we may define $F^k(x_0,\dots,x_k) :=\downarrow \lim_{n\to\infty} F_n^k(x_0,\dots,x_k)$ which is formally equal to $\mathbb{E}[f|\mathcal{B}_k](x)$. Hence we expect that

$$F^k(x_0,\dots,x_k) = \int F^{k+1}(x_0,\dots,x_k,x_{k+1}) T_k(x_0,\dots,x_k,dx_{k+1}) \tag{19.87}$$

by the tower property for conditional expectations. This is indeed that case since,

$$\begin{aligned} &\int F^{k+1}(x_0,\dots,x_k,x_{k+1}) T_k(x_0,\dots,x_k,dx_{k+1}) \\ &\quad = \lim_{n\to\infty} \int F_n^{k+1}(x_0,\dots,x_k,x_{k+1}) T_k(x_0,\dots,x_k,dx_{k+1}) \end{aligned}$$

while

$$\begin{aligned} &\int F_n^{k+1}(x_0,\dots,x_k,x_{k+1}) T_k(x_0,\dots,x_k,dx_{k+1}) \\ &= \int \left[\int \cdots \int F_n(x_0,\dots,x_n) \prod_{l=k+1}^{n} T_l(x_0,\dots,x_l,dx_{l+1})\right] T_k(x_0,\dots,x_k,dx_{k+1}) \\ &= \int \cdots \int F_n(x_0,\dots,x_n) \prod_{l=k}^{n} T_l(x_0,\dots,x_l,dx_{l+1}) \\ &= F_n^k(x_0,\dots,x_k). \end{aligned}$$

We may now pass to the limit as $n \to \infty$ in Eq. (19.86) to find

$$P_k\left(F^k\right) = \varepsilon > 0 \text{ for all } k.$$

For $k = 0$ it follows that $F^0(x_0) \geq \varepsilon > 0$ for some $x_0 \in S_0$ for otherwise $P_0(F_0) < \varepsilon$. But

$$\varepsilon \leq F^0(x_0) = \int F^1(x_0,x_1) T_1(x_0,dx_1)$$

and so there exists $x_1$ such that

$$\varepsilon \leq F^1(x_0,x_1) = \int F^2(x_0,x_1,x_2) T_2(x_0,x_1,dx_2)$$

and hence there exists $x_2$ such that $\varepsilon \leq F^2(x_0,x_1,x_2)$, etc. etc. Thus in the end we find an $x = (x_0,x_1,\dots) \in S$ such that $F^k(x_0,\dots,x_n) \geq \varepsilon$ for all $k$. Finally recall that

$$F_n^k(x_0,\dots,x_k) \geq F^k(x_0,\dots,x_k) \geq \varepsilon \text{ for all } k \leq n.$$

Taking $k = n$ then implies,

$$f_n(x) = F_n^n(x_0,\dots,x_n) \geq F^n(x_0,\dots,x_n) \geq \varepsilon \text{ for all } n.$$

Therefore we have constructed a $x \in S$ such that $f(x) = \lim_{n\to\infty} f_n(x) \geq \varepsilon > 0$.

We may now use the Caratheodory extension theorem to show that $P$ extends to a countably additive measure on $(S,\mathcal{S})$. Indeed suppose $A_n \in \mathcal{A}(X_i : i \in \mathbb{N}_0)$. If $A_n \downarrow \emptyset$ then $1_{A_n} \downarrow 0$ and by what we have just proved,

$$P(A_n) = P(1_{A_n}) \downarrow 0 \text{ as } n \to \infty.$$

■

**Corollary 19.71 (Infinite Product Measures).** *Let $\{(S_n,\mathcal{S}_n,\mu_n)\}_{n\in\mathbb{N}_0}$ be a collection of measurable spaces, then there exists $P$ on $(S,\mathcal{S})$ such that*

$$P(f) = P(f) = \int_{S^n} F(x_0,\dots,x_n)\, d\nu_0(x_0)\dots d\nu_n(x_n)$$

*whenever $f(x) = F(x_0,\dots,x_n)$ for some $F \in (\mathcal{S}_n)_b$.*

**Proof.** Let $\mu_0 = \nu_0$ and

$$T_n(x_0,\dots,x_{n-1},dx_{n+1}) = v_n(dx_n).$$

Then in this case we will have

$$\mu_n(dx_0,\dots,dx_n) = d\nu_0(x_0)\, d\nu_1(dx_1)\dots\nu_n(dx_n)$$

as desired. ■

## 19.9 *Appendix: More Probability Kernel Constructions

**Lemma 19.72.** *Suppose that $(X,\mathcal{M})$, $(Y,\mathcal{F})$, and $(Z,\mathcal{B})$ are measurable spaces and $Q : X\times\mathcal{F}\to[0,1]$ and $R : Y\times\mathcal{B}\to[0,1]$ are probability kernels. Then for every bounded measurable function, $F : (Y\times Z,\mathcal{F}\otimes\mathcal{B})\to(\mathbb{R},\mathcal{B}_{\mathbb{R}})$, the map*

$$y\to\int_Z R(y,dz)\,F(y,z)$$

*is measurable. Moreover, if we define $P(x;A)$ for $A\in\mathcal{F}\otimes\mathcal{B}$ and $x\in X$ by*

$$P(x,A)=\int_Y Q(x,dy)\int_Z R(y,dz)\,1_A(y,z),$$

*then $P : X\times\mathcal{F}\otimes\mathcal{B}\to[0,1]$ is a probability kernel such that*

$$P(x,F)=\int_Y Q(x,dy)\int_Z R(y,dz)\,F(y,z)$$

*for all bounded measurable functions, $F : (Y\times Z,\mathcal{F}\otimes\mathcal{B})\to(\mathbb{R},\mathcal{B}_{\mathbb{R}})$. We will denote the kernel $P$ by $Q\otimes R$ and write*

$$(Q\otimes R)(x,dy,dz)=Q(x,dy)\,R(y,dz).$$

*Moreover if $S(z,dw)$ is another probability kernel, then $((Q\otimes R)\otimes S)=(Q\otimes(R\otimes S))$.*

**Proof.** A routine exercise in using the multiplicative systems theorem. To verify the last assertion it suffices to consider the kernels on sets of the form $A\times B\times C$ in which case,

$$\begin{aligned}
&(Q\otimes(R\otimes S))(x,A\times B\times C)\\
&\quad=\int_Y Q(x,dy)\int_{Z\times W} RS(y,dz,dw)\,1_{A\times B\times C}(y,z,w)\\
&\quad=\int_Y Q(x,dy)\,1_A(y)\int_{Z\times W} RS(y;B\times C)\\
&\quad=\int_Y Q(x,dy)\,1_A(y)\int_{Z\times W} R(y,dz)\,S(z,dw)\,1_{B\times C}(z,w)\\
&\quad=\int_Y Q(x,dy)\,1_A(y)\int_Z R(y,dz)\,S(z,C)\,1_B(z)
\end{aligned}$$

while

$$\begin{aligned}
&((Q\otimes R)\otimes S)(x,A\times B\times C)\\
&\quad=\int_{Y\times Z} QR(x,dy,dz)\int_{Z\times W} S(z,dw)\,1_{A\times B\times C}(y,z,w)\\
&\quad=\int_{Y\times Z} QR(x,dy,dz)\,1_{A\times B}(y,z)\,S(z,C)\\
&\quad=\int_Y Q(x,dy)\int_Z R(y,dz)\,1_{A\times B}(y,z)\,S(z,C)\\
&\quad=\int_Y Q(x,dy)\,1_A(y)\int_Z R(y,dz)\,S(z,C)\,1_B(z).
\end{aligned}$$

■

**Corollary 19.73.** *Keeping the notation in Lemma 19.72, let $QR$ be the probability kernel given by $QR(x,dz)=\int_Y Q(x,dy)\,R(y,dz)$ so that*

$$QR(x;B)=Q\otimes R(x;Y\times B).$$

*Then we have $Q(RS)=(QR)S$.*

**Proof.** Let $C\in\mathcal{B}_W$, then

$$\begin{aligned}
Q(RS)(x;C)&=Q\otimes(RS)(x;Y\times C)=\int_Y Q(x,dy)\,(RS)(y;C)\\
&=\int_Y Q(x,dy)\,(R\otimes S)(y;Z\times C)=[Q\otimes(R\otimes S)](Y\times Z\times C).
\end{aligned}$$

Similarly one shows that

$$(QR)S(x;C)=[(Q\otimes R)\otimes S](Y\times Z\times C)$$

and then the result follows from Lemma 19.72. ■

# (Sub and Super) Martingales

Let us start with a reminder of a few key notions that were already introduced in Chapter 19. As usual we will let $(S, \mathcal{S})$ denote a measurable space called **state space.** (Often in this chapter we will take $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.) As in Chapter 19, we will fix a **filtered probability space,** $\left(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n \in \mathbb{N}_0}, P\right)$, i.e. $\mathcal{B}_n \subset \mathcal{B}_{n+1} \subset \mathcal{B}$ for all $n = 0, 1, 2 \dots$. We further define

$$\mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma\left(\cup_{n=0}^\infty \mathcal{B}_n\right) \subset \mathcal{B}. \tag{20.1}$$

Also recall that a sequence of random functions, $Y_n : \Omega \to S$ for $n \in \mathbb{N}_0$, are said to be **adapted** to the filtration if $Y_n$ is $\mathcal{B}_n/\mathcal{S}$ – measurable for all $n$.

**Definition 20.1.** *Let $X := \{X_n\}_{n=0}^\infty$ is a be an adapted sequence of integrable random variables. Then;*

1. *$X$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **martingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ a.s. for all $n \in \mathbb{N}_0$.*
2. *$X$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **submartingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \geq X_n$ a.s. for all $n \in \mathbb{N}_0$.*
3. *$X$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – **supermartingale** if $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \leq X_n$ a.s. for all $n \in \mathbb{N}_0$.*

It is often fruitful to view $X_n$ as your earnings at time $n$ while playing some game of chance. In this interpretation, your expected earnings at time $n+1$ given the history of the game up to time $n$ is the same, greater than, less than your earnings at time $n$ if $X = \{X_n\}_{n=0}^\infty$ is a martingale, submartingale or supermartingale respectively. In this interpretation, martingales are fair games, submartingales are games which are favorable to the gambler (unfavorable to the casino), and supermartingales are games which are unfavorable to the gambler (favorable to the casino), see Example 20.4.

By induction one shows that $X$ is a supermartingale, martingale, or submartingale iff

$$\mathbb{E}[X_n|\mathcal{B}_m] \overset{\leq}{\underset{\geq}{=}} X_m \text{ a.s for all } n \geq m, \tag{20.2}$$

to be read from top to bottom respectively. This last equation may also be expressed as

$$\mathbb{E}[X_n|\mathcal{B}_m] \overset{\leq}{\underset{\geq}{=}} X_{n \wedge m} \text{ a.s for all } m, n \in \mathbb{N}_0. \tag{20.3}$$

The reader should also note that $\mathbb{E}[X_n]$ is decreasing, constant, or increasing respectively. The next lemma shows that we may shrink the filtration, $\{\mathcal{B}_n\}_{n=0}^\infty$, within limits and still have $X$ retain the property of being a supermartingale, martingale, or submartingale.

**Lemma 20.2 (Shrinking the filtration).** *Suppose that $X$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively and $\{\mathcal{B}'_n\}_{n=0}^\infty$ is another filtration such that $\sigma(X_0, \dots, X_n) \subset \mathcal{B}'_n \subset \mathcal{B}_n$ for all $n$. Then $X$ is a $\{\mathcal{B}'_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively.*

**Proof.** Since $\{X_n\}_{n=0}^\infty$ is adapted to $\{\mathcal{B}_n\}_{n=0}^\infty$ and $\sigma(X_0, \dots, X_n) \subset \mathcal{B}'_n \subset \mathcal{B}_n$, for all $n$,

$$\mathbb{E}_{\mathcal{B}'_n} X_{n+1} = \mathbb{E}_{\mathcal{B}'_n} \mathbb{E}_{\mathcal{B}_n} X_{n+1} \overset{\leq}{\underset{\geq}{=}} \mathbb{E}_{\mathcal{B}'_n} X_n = X_n,$$

when $X$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively – read from top to bottom. ■

Enlarging the filtration is another matter all together. In what follows we will simply say $X$ is a supermartingale, martingale, submartingale if it is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale.

## 20.1 (Sub and Super) Martingale Examples

*Example 20.3.* Suppose that $\{Z_n\}_{n=0}^\infty$ are independent integrable random variables such that $\mathbb{E}Z_n = 0$ for all $n \geq 1$. Then $S_n := \sum_{k=0}^n Z_k$ is a martingale relative to the filtration, $\mathcal{B}_n^Z := \sigma(Z_0, \dots, Z_n)$. Indeed,

$$\mathbb{E}[S_{n+1} - S_n|\mathcal{B}_n] = \mathbb{E}[Z_{n+1}|\mathcal{B}_n] = \mathbb{E}Z_{n+1} = 0.$$

This same computation also shows that $\{S_n\}_{n\geq 0}$ is a submartingale if $\mathbb{E}Z_n \geq 0$ and supermartingale if $\mathbb{E}Z_n \leq 0$ for all $n$.

**Exercise 20.1.** Construct an example of a martingale, $\{M_n\}_{n=0}^\infty$ such that $\mathbb{E}|M_n| \to \infty$ as $n \to \infty$. [In particular, $\{M_n\}_{n=1}^\infty$ will be a martingale which is not of the form $M_n = \mathbb{E}_{\mathcal{B}_n} X$ for some $X \in L^1(P)$.] **Hint:** try taking $M_n = \sum_{k=0}^n Z_k$ for a judicious choice of $\{Z_k\}_{k=0}^\infty$ which you should take to be independent, mean zero, and having $\mathbb{E}|Z_n|$ growing rather rapidly.

*Example 20.4 (Setting the odds).* Let $S$ be a finite set (think of the outcomes of a spinner, or dice, or a roulette wheel) and $p : S \to (0,1)$ be a probability function[1]. Let $\{Z_n\}_{n=1}^{\infty}$ be random functions with values in $S$ such that $p(s) := P(Z_n = s)$ for all $s \in S$. ($Z_n$ represents the outcome of the $n^{\text{th}}$ – game.) Also let $\alpha : S \to [0,\infty)$ be the house's payoff function, i.e. for each dollar you (the gambler) bets on $s \in S$, the house will pay $\alpha(s)$ dollars back if $s$ is rolled. Further let $W : \Omega \to \mathbb{W}$ be measurable function into some other measure space, $(\mathbb{W}, \mathcal{F})$ which is to represent your random (or not so random) "whims.". We now assume that $Z_n$ is independent of $(W, Z_1, \dots, Z_{n-1})$ for each $n$, i.e. the dice are not influenced by the previous plays or your whims. If we let $\mathcal{B}_n := \sigma(W, Z_1, \dots, Z_n)$ with $\mathcal{B}_0 = \sigma(W)$, then we are assuming the $Z_n$ is independent of $\mathcal{B}_{n-1}$ for each $n \in \mathbb{N}$.

As a gambler, you are allowed to choose **before** the $n^{\text{th}}$ – game is played, the amounts $\left(\{C_n(s)\}_{s\in S}\right)$ that you want to bet on each of the possible outcomes of the $n^{\text{th}}$ – game. Assuming that you are not clairvoyant (i.e. can not see the future), these amounts may be random but must be $\mathcal{B}_{n-1}$ – measurable, that is $C_n(s) = C_n(W, Z_1, \dots, Z_{n-1}, s)$, i.e. $\{C_n(s)\}_{n=1}^{\infty}$ is "previsible" process (see Definition 20.5 below). Thus if $X_0$ denotes your initial wealth (assumed to be a non-random quantity) and $X_n$ denotes your wealth just after the $n^{\text{th}}$ – game is played, then

$$X_n - X_{n-1} = -\sum_{s\in S} C_n(s) + C_n(Z_n)\,\alpha(Z_n)$$

where $-\sum_{s\in S} C_n(s)$ is your total bet on the $n^{\text{th}}$ – game and $C_n(Z_n)\,\alpha(Z_n)$ represents the house's payoff to you for the $n^{\text{th}}$ – game. Therefore it follows that

$$X_n = X_0 + \sum_{k=1}^{n}\left[-\sum_{s\in S} C_k(s) + C_n(Z_k)\,\alpha(Z_k)\right],$$

$X_n$ is $\mathcal{B}_n$ – measurable for each $n$, and

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_{n-1}}[X_n - X_{n-1}] &= -\sum_{s\in S} C_n(s) + \mathbb{E}_{\mathcal{B}_{n-1}}[C_n(Z_n)\,\alpha(Z_n)] \\ &= -\sum_{s\in S} C_n(s) + \sum_{s\in S} C_n(s)\,\alpha(s)\,p(s) \\ &= \sum_{s\in S} C_n(s)\,(\alpha(s)\,p(s) - 1).\end{aligned}$$

[1] To be concrete, take $S = \{2, \dots, 12\}$ representing the possible values for the sums of the upward pointing faces of two dice. Assuming the dice are independent and fair then determines $p : S \to (0,1)$. For example $p(2) = p(12) = 1/36$, $p(3) = p(11) = 1/18$, $p(7) = 1/6$, etc.

Thus it follows, that no matter the choice of the betting "strategy," $\{C_n(s) : s \in S\}_{n=1}^{\infty}$, we will have

$$\mathbb{E}_{\mathcal{B}_{n-1}}[X_n - X_{n-1}] = \begin{cases} \geq 0 \text{ if } \alpha(\cdot)\,p(\cdot) \geq 1 \\ = 0 \text{ if } \alpha(\cdot)\,p(\cdot) = 1 \\ \leq 0 \text{ if } \alpha(\cdot)\,p(\cdot) \leq 1 \end{cases},$$

that is $\{X_n\}_{n\geq 0}$ is a sub-martingale, martingale, or supermartingale depending on whether $\alpha \cdot p \geq 1$, $\alpha \cdot p = 1$, or $\alpha \cdot p \leq 1$.

**Moral:** If the Casino wants to be guaranteed to make money on average, it had better choose $\alpha : S \to [0,\infty)$ such that $\alpha(s) < 1/p(s)$ for all $s \in S$. In this case the expected earnings of the gambler will be decreasing which means the expected earnings of the Casino will be increasing.

**Definition 20.5.** *We say $\{C_n : \Omega \to S\}_{n=1}^{\infty}$ is* ***predictable or previsible*** *if each $C_n$ is $\mathcal{B}_{n-1}/\mathcal{S}$ – measurable for all $n \in \mathbb{N}$.*

A typical example is when $\{X_n : \Omega \to S\}_{n=0}^{\infty}$ is a sequence of measurable functions on a probability space $(\Omega, \mathcal{B}, P)$ and $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$. An application of Lemma 16.1 shows that a sequence of random variables, $\{Y_n\}_{n=0}^{\infty}$, is adapted to the filtration iff there are $\mathcal{S}^{\otimes(n+1)}/\mathcal{B}_{\mathbb{R}}$ – measurable functions, $f_n : S^{n+1} \to \mathbb{R}$, such that $Y_n = f_n(X_0, \dots, X_n)$ for all $n \in \mathbb{N}_0$ and a sequence of random variables, $\{Z_n\}_{n=1}^{\infty}$, is predictable iff there exists, there are measurable functions, $f_n : \mathbb{R}^n \to \mathbb{R}$ such that $Z_n = f_n(X_0, \dots, X_{n-1})$ for all $n \in \mathbb{N}$.

*Example 20.6 (Regular martingales).* Suppose that $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ is a filtered probability space and $X \in L^1(\Omega, \mathcal{B}, P)$. Then $X_n := \mathbb{E}[X|\mathcal{B}_n]$ is a martingale. Indeed, by the tower property of conditional expectations,

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_{n+1}]\,|\mathcal{B}_n] = \mathbb{E}[X|\mathcal{B}_n] = X_n \text{ a.s.}$$

When $X_n := \mathbb{E}[X|\mathcal{B}_n]$ for some $X \in L^1(P)$ we say that $\{X_n\}_{n=1}^{\infty}$ is a **regular martingale.**

*Example 20.7.* Suppose that $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$, and $P = m$ – Lebesgue measure. Let $\mathcal{P}_n = \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right\}_{k=0}^{2^n-1}$ and $\mathcal{B}_n := \sigma(\mathcal{P}_n)$ for each $n \in \mathbb{N}$. Then $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}$ is a martingale (Exercise 20.2) such that $\mathbb{E}|M_n| = 1$ for all $n$. However, there is no $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{B}_n]$. To verify this last assertion, suppose such an $X$ existed. We would then have for $2^n > k > 0$ and any $m > n$, that

$$\begin{aligned}\mathbb{E}\left[X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_m} X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] \\ &= \mathbb{E}\left[M_m : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = 0.\end{aligned}$$

Using $\mathbb{E}[X : A] = 0$ for all $A$ in the $\pi$ – system, $\mathcal{Q} := \cup_{n=1}^{\infty}\left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] : 0 \le k < 2^n\right\}$, an application of the $\pi - \lambda$ theorem shows $\mathbb{E}[X : A] = 0$ for all $A \in \sigma(\mathcal{Q}) = \mathcal{B}$. Therefore $X = 0$ a.s. by Proposition 9.22. But this is impossible since $1 = \mathbb{E}M_n = \mathbb{E}X$.

**Moral:** not all $L^1$ – bounded martingales are regular, i.e. as in Example 20.6. Proposition 20.8 shows what is missing from this martingale in order for it to be of the form in Example 20.6. See the comments after Example 20.11 for another $L^1$ – bounded martingale which is not of the form in example 20.6.

**Exercise 20.2.** Show that $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}$ as defined in Example 20.7 is a martingale.

**Proposition 20.8.** *Suppose $1 \le p < \infty$ and $X \in L^p(\Omega, \mathcal{B}, P)$. Then the collection of random variables, $\Gamma := \{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subset \mathcal{B}\}$ is a bounded subset of $L^p(\Omega, \mathcal{B}, P)$ which is also uniformly integrable.*

**Proof.** Since $\mathbb{E}_{\mathcal{G}}$ is a contraction on all $L^p$ – spaces it follows that $\Gamma$ is bounded in $L^p$ with

$$\sup_{\mathcal{G} \subset \mathcal{B}} \|\mathbb{E}[X|\mathcal{G}]\|_p \le \|X\|_p .$$

For the $p > 1$ the uniform integrability of $\Gamma$ follows directly from Lemma 14.54.

We now concentrate on the $p = 1$ case. Recall that $|\mathbb{E}_{\mathcal{G}}X| \le \mathbb{E}_{\mathcal{G}}|X|$ a.s. and therefore,

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}}X| : |\mathbb{E}_{\mathcal{G}}X| \ge a] \le \mathbb{E}[|X| : |\mathbb{E}_{\mathcal{G}}X| \ge a] \text{ for all } a > 0.$$

But by Chebyshev's inequality,

$$P(|\mathbb{E}_{\mathcal{G}}X| \ge a) \le \frac{1}{a}\mathbb{E}|\mathbb{E}_{\mathcal{G}}X| \le \frac{1}{a}\mathbb{E}|X| .$$

Since $\{|X|\}$ is uniformly integrable, it follows from Proposition 14.48 that, by choosing $a$ sufficiently large, $\mathbb{E}[|X| : |\mathbb{E}_{\mathcal{G}}X| \ge a]$ is as small as we please uniformly in $\mathcal{G} \subset \mathcal{B}$ and therefore,

$$\lim_{a \to \infty} \sup_{\mathcal{G} \subset \mathcal{B}} \mathbb{E}[|\mathbb{E}_{\mathcal{G}}X| : |\mathbb{E}_{\mathcal{G}}X| \ge a] = 0.$$

■

*Example 20.9.* This example generalizes Example 20.7. Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ is a filtered probability space and $Q$ is another probability measure on $(\Omega, \mathcal{B})$. Let us assume that $Q|_{\mathcal{B}_n} \ll P|_{\mathcal{B}_n}$ for all $n$, which by the Raydon-Nikodym Theorem 17.8, implies there exists $0 \le X_n \in L^1(\Omega, \mathcal{B}_n, P)$ with $\mathbb{E}X_n = 1$ such that $dQ|_{\mathcal{B}_n} = X_n dP|_{\mathcal{B}_n}$, or equivalently put, for any $B \in \mathcal{B}_n$ we have



$$Q(B) = \int_B X_n dP = \mathbb{E}[X_n : B].$$

Since $B \in \mathcal{B}_n \subset \mathcal{B}_{n+1}$, we also have $\mathbb{E}[X_{n+1} : B] = Q(B) = \mathbb{E}[X_n : B]$ for all $B \in \mathcal{B}_n$ and hence $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ a.s., i.e. $X = \{X_n\}_{n=0}^{\infty}$ is a positive martingale.

Example 20.7 is of this form with $Q = \delta_0$. Notice that $\delta_0|_{\mathcal{B}_n} \ll m|_{\mathcal{B}_n}$ for all $n < \infty$ while $\delta_0 \perp m$ on $\mathcal{B}_{[0,1]} = \mathcal{B}_\infty$. See Section 21.4 for more in the direction of this example.

**Lemma 20.10.** *Let $X := \{X_n\}_{n=0}^{\infty}$ be an adapted process of integrable random variables on a filtered probability space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ and let $d_n := X_n - X_{n-1}$ with $X_{-1} := \mathbb{E}X_0$. Then $X$ is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \ge 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \le 0$ respectively) for all $n \in \mathbb{N}_0$.*

*Conversely if $\{d_n\}_{n=1}^{\infty}$ is an adapted sequence of integrable random variables and $X_0$ is a $\mathcal{B}_0$ -measurable integrable random variable. Then $X_n = X_0 + \sum_{j=1}^{n} d_j$ is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \ge 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \le 0$ respectively) for all $n \in \mathbb{N}$.*

**Proof.** We prove the assertions for martingales only, the other all being similar. Clearly $X$ is a martingale iff

$$0 = \mathbb{E}[X_{n+1}|\mathcal{B}_n] - X_n = \mathbb{E}[X_{n+1} - X_n|\mathcal{B}_n] = \mathbb{E}[d_{n+1}|\mathcal{B}_n].$$

The second assertion is an easy consequence of the first assertion. ■

*Example 20.11.* Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a sequence of independent integrable random variables, $X_n = Z_0 \dots Z_n$, and $\mathcal{B}_n := \sigma(Z_0, \cdots, Z_n)$. (Observe that $\mathbb{E}|X_n| = \prod_{k=0}^{n} \mathbb{E}|Z_k| < \infty$.) Since

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[X_n Z_{n+1}|\mathcal{B}_n] = X_n \mathbb{E}[Z_{n+1}|\mathcal{B}_n] = X_n \cdot \mathbb{E}[Z_{n+1}] \text{ a.s.,}$$

it follows that $\{X_n\}_{n=0}^{\infty}$ is a martingale if $\mathbb{E}Z_n = 1$. If we further assume, for all $n$, that $Z_n \ge 0$ so that $X_n \ge 0$, then $\{X_n\}_{n=0}^{\infty}$ is a supermartingale (submartingale) provided $\mathbb{E}Z_n \le 1$ ($\mathbb{E}Z_n \ge 1$) for all $n$.

Let us specialize the above example even more by taking $Z_n \overset{d}{=} p + U$ where $p \ge 0$ and $U$ is the uniform distribution on $[0, 1]$. In this case we have by the strong law of large numbers that

$$\frac{1}{n}\ln X_n = \frac{1}{n}\sum_{k=0}^{n} \ln Z_k \to \mathbb{E}[\ln(p + U)] \text{ a.s.} \tag{20.4}$$

An elementary computation shows

$$\mathbb{E}[\ln(p+U)] = \int_0^1 \ln(p+x)\,dx = \int_p^{p+1} \ln(p+x)\,dx$$
$$= (x\ln x - x)_{x=p}^{x=p+1} = (p+1)\ln(p+1) - p\ln p - 1$$

The function $f(p) := \mathbb{E}[\ln(p+U)]$ satisfies has a zero at $p = p_c \cong 0.542\,21$

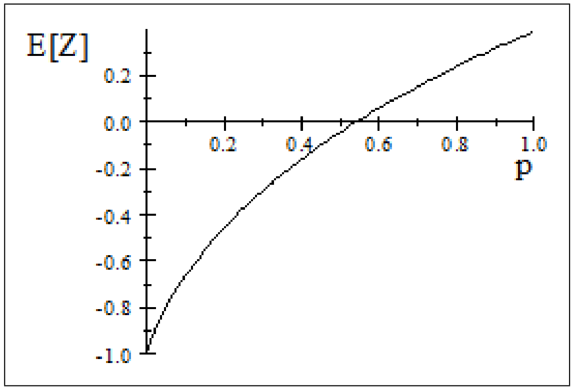



**Fig. 20.1.** The graph of $\mathbb{E}[\ln(p+U)]$ as a function of $p$. This function has a zero at $p = p_c \cong 0.542\,21$.

and $f(p) < 0$ for $p < p_c$ while $f(p) > 0$ for $p > p_c$, see Figure 20.1. Combining these observations with Eq. (20.4) implies,

$$X_n \to \lim_{n\to\infty} \exp(n\mathbb{E}[\ln(p+U)]) = \begin{cases} 0 & \text{if } p < p_c \\ ? & \text{if } p = p_c \\ \infty & \text{if } p > p_c \end{cases} \text{ a.s.}$$

Notice that $\mathbb{E}Z_n = p + 1/2$ and therefore $X_n$ is a martingale precisely when $p = 1/2$ and is a sub-martingale for $p > 1/2$. So for $1/2 < p < p_c$, $\{X_n\}_{n=1}^\infty$ is a positive sub-martingale, $\mathbb{E}X_n = (p+1/2)^{n+1} \to \infty$ yet $\lim_{n\to\infty} X_n = 0$ a.s. Have a look at the excel file (Product_positive-(sub)martingales.xls) in order to construct sample paths for the $\{X_n\}_{n=0}^\infty$.

**Proposition 20.12.** *Suppose that $X = \{X_n\}_{n=0}^\infty$ is a martingale and $\varphi$ is a convex function such that $\varphi(X_n) \in L^1$ for all $n$. Then $\varphi(X) = \{\varphi(X_n)\}_{n=0}^\infty$ is a submartingale. If $\varphi$ is also assumed to be increasing, it suffices to assume that $X$ is a submartingale in order to conclude that $\varphi(X)$ is a submartingale. (For example if $X$ is a positive submartingale, $p \in (1,\infty)$, and $\mathbb{E}X_n^p < \infty$ for all $n$, then $X^p := \{X_n^p\}_{n=0}^\infty$ is another positive submartingale.*

**Proof.** When $X$ is a martingale, by the conditional Jensen's inequality 16.31,

$$\varphi(X_n) = \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \le \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

which shows $\varphi(X)$ is a submartingale. Similarly, if $X$ is a submartingale and $\varphi$ is convex and increasing, then $\varphi$ preserves the inequality, $X_n \le \mathbb{E}_{\mathcal{B}_n} X_{n+1}$, and hence

$$\varphi(X_n) \le \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \le \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

so again $\varphi(X)$ is a submartingale. ∎

**Proposition 20.13 (Markov Chains and Martingales).** *Suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}\}_{n\in\mathbb{N}_0}, \{X_n : \Omega \to S\}_{n\ge 0}, Q, P\right)$ is a time homogeneous Markov chain and $f : \mathbb{N}_0 \times S \to \mathbb{R}$ be measurable function which is either non-negative or satisfies $\mathbb{E}[|f(n,X_n)|] < \infty$ for all $n$ and let $Z_n := f(n,X_n)$. Then $\{Z_n\}_{n=0}^\infty$ is a (sub-martingale) martingale if $(Qf(n+1,\cdot) \le f(n\cdot))$ $Qf(n+1,\cdot) = f(n,\cdot)$ for all $n \ge 0$. In particular if $f : S \to \mathbb{R}$ is a function such that $(Qf \le f)$ $Qf = f$ then $Z_n = f(X_n)$ is a (sub-martingale) martingale. (Also see Exercise 20.5 below.)*

**Proof.** Using the Markov property and the definition of $Q$, we have

$$\mathbb{E}[Z_{n+1}|\mathcal{B}_n] = \mathbb{E}[f(n+1,X_{n+1})|\mathcal{B}_n] = [Qf(n+1,\cdot)](X_n).$$

The latter expression is (less than or equal) equal to $Z_n$ if $(Qf(n+1,\cdot) \le f(n\cdot))$ $Qf(n+1,\cdot) = f(n,\cdot)$ for all $n \ge 0$. ∎

One way to find solutions to the equation $Qf(n+1,\cdot) = f(n,\cdot)$ at least for a finite number of $n$ is to let $g : S \to \mathbb{R}$ be an arbitrary function and $T \in \mathbb{N}$ be given and then define

$$f(n,y) := \left(Q^{T-n}g\right)(y) \text{ for } 0 \le n \le T.$$

Then $Qf(n+1,\cdot) = Q\left(Q^{T-n-1}g\right) = Q^{T-n}g = f(n,\cdot)$ and we will have that

$$Z_n = f(n,X_n) = \left(Q^{T-n}g\right)(X_n)$$

is a Martingale for $0 \le n \le T$. If $f(n,\cdot)$ satisfies $Qf(n+1,\cdot) = f(n,\cdot)$ for all $n$ then we must have, with $f_0 := f(0,\cdot)$,

$$f(n,\cdot) = Q^{-n} f_0$$

where $Q^{-1}g$ denotes a function $h$ solving $Qh = g$. In general $Q$ is not invertible and hence there may be no solution to $Qh = g$ or there might be many solutions.

*Example 20.14.* In special cases one can often make sense of these expressions (see Exercise 20.5). Let $S = \mathbb{Z}$, $S_n = X_0 + X_1 + \cdots + X_n$, where $\{X_i\}_{i=1}^\infty$ are i.i.d. with $P(X_i = 1) = p \in (0,1)$ and $P(X_i = -1) = q := 1 - p$, and $X_0$ is $S-$valued random variable independent of $\{X_i\}_{i=1}^\infty$ as in Exercise 19.48. Recall that

$\{S_n\}_{n=0}^\infty$ is a time homogeneous Markov chain with transition kernel determined by $Qf(x) = pf(x+1) + qf(x-1)$. As we have seen if $f(x) = a + b(q/p)^x$, then $Qf = f$ and therefore

$$M_n = a + b(q/p)^{S_n}$$

is a martingale for all $a, b \in \mathbb{R}$. This is easily verified directly as well;

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n}\left(\frac{q}{p}\right)^{S_{n+1}} &= \mathbb{E}_{\mathcal{B}_n}\left(\frac{q}{p}\right)^{S_n+X_{n+1}} = \left(\frac{q}{p}\right)^{S_n}\mathbb{E}_{\mathcal{B}_n}\left(\frac{q}{p}\right)^{X_{n+1}} \\ &= \left(\frac{q}{p}\right)^{S_n}\mathbb{E}\left(\frac{q}{p}\right)^{X_{n+1}} = \left(\frac{q}{p}\right)^{S_n}\cdot\left[\left(\frac{q}{p}\right)^1 p + \left(\frac{q}{p}\right)^{-1} q\right] \\ &= \left(\frac{q}{p}\right)^{S_n}\cdot[q+p] = \left(\frac{q}{p}\right)^{S_n}.\end{aligned}$$

Now suppose that $\lambda \neq 0$ and observe that $Q\lambda^x = \left(p\lambda + q\lambda^{-1}\right)\lambda^x$. Thus it follows that we may set $Q^{-1}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-1}\lambda^x$ and therefore conclude that

$$f(n,x) := Q^{-n}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^x$$

satisfies $Qf(n+1,\cdot) = f(n,\cdot)$. So if we suppose that $X_0$ is a bounded so that $S_n$ is bounded for all $n$, we will have $\left\{M_n = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^{S_n}\right\}_{n\geq 0}$ is a martingale for all $\lambda \neq 0$.

**Exercise 20.3.** For $\theta \in \mathbb{R}$ let

$$f_\theta(n,x) := Q^{-n}e^{\theta x} = \left(pe^\theta + qe^{-\theta}\right)^{-n}e^{\theta x}$$

so that $Qf_\theta(n+1,\cdot) = f_\theta(n,\cdot)$ for all $\theta \in \mathbb{R}$. Compute;

1. $f_\theta^{(k)}(n,x) := \left(\frac{d}{d\theta}\right)^k f_\theta(n,x)$ for $k = 1, 2$.
2. Use your results to show,

$$M_n^{(1)} := S_n - n(p-q)$$

and

$$M_n^{(2)} := (S_n - n(p-q))^2 - 4npq$$

are martingales.

(If you are ambitious you might also find $M_n^{(3)}$.)

*Remark 20.15.* If $\{M_n(\theta)\}_{n=0}^\infty$ is a martingale depending differentiability on a parameter $\theta \in \mathbb{R}$. Then for all $A \in \mathcal{B}_n$,

$$\mathbb{E}\left[\frac{d}{d\theta}M_{n+1}(\theta) : A\right] = \frac{d}{d\theta}\mathbb{E}\left[M_{n+1}(\theta) : A\right] = \frac{d}{d\theta}\mathbb{E}\left[M_n(\theta) : A\right] = \mathbb{E}\left[\frac{d}{d\theta}M_n(\theta) : A\right]$$

provided it is permissible to interchange $\frac{d}{d\theta}$ with the expectations in this equation. Thus under "suitable" hypothesis, we will have $\left\{\frac{d}{d\theta}M_n(\theta)\right\}_{n\geq 0}$ is another martingale.

## 20.2 Decompositions

**Notation 20.16** *Given a sequence* $\{Z_k\}_{k=0}^\infty$, *let* $\Delta_k Z := Z_k - Z_{k-1}$ *for* $k = 1, 2, \ldots$.

**Lemma 20.17 (Doob Decomposition).** *Each adapted sequence,* $\{Z_n\}_{n=0}^\infty$, *of integrable random variables has a unique decomposition,*

$$Z_n = M_n + A_n \tag{20.5}$$

*where* $\{M_n\}_{n=0}^\infty$ *is a martingale and* $A_n$ *is a predictable process such that* $A_0 = 0$. *Moreover this decomposition is given by* $A_0 = 0$,

$$A_n := \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}}[\Delta_k Z] \text{ for } n \geq 1 \tag{20.6}$$

*and*

$$M_n = Z_n - A_n = Z_n - \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}}[\Delta_k Z] \tag{20.7}$$

$$= Z_0 + \sum_{k=1}^n \left(Z_k - \mathbb{E}_{\mathcal{B}_{k-1}}Z_k\right). \tag{20.8}$$

*In particular,* $\{Z_n\}_{n=0}^\infty$ *is a submartingale (supermartingale) iff* $A_n$ *is increasing (decreasing) almost surely.*

**Proof.** Assuming $Z_n$ has a decomposition as in Eq. (20.5), then

$$\mathbb{E}_{\mathcal{B}_n}[\Delta_{n+1}Z] = \mathbb{E}_{\mathcal{B}_n}[\Delta_{n+1}M + \Delta_{n+1}A] = \Delta_{n+1}A \tag{20.9}$$

wherein we have used $M$ is a martingale and $A$ is predictable so that $\mathbb{E}_{\mathcal{B}_n}[\Delta_{n+1}M] = 0$ and $\mathbb{E}_{\mathcal{B}_n}[\Delta_{n+1}A] = \Delta_{n+1}A$. Hence we must define, for $m \geq 1$,

$$A_n := \sum_{k=1}^n \Delta_k A = \sum_{k=1}^n \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z]$$

which is a predictable process. This proves the uniqueness of the decomposition and the validity of Eq. (20.6).

For existence, from Eq. (20.6) it follows that

$$\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} Z] = \Delta_{n+1} A = \mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} A] .$$

Hence, if we define $M_n := Z_n - A_n$, then

$$\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} M] = \mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} Z - \Delta_{n+1} A] = 0$$

and hence $\{M_n\}_{n=0}^\infty$ is a martingale. Moreover, Eq. (20.8) follows from Eq. (20.7) since,

$$M_n = Z_0 + \sum_{k=1}^n \left(\Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z]\right)$$

and

$$\begin{aligned}\Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z] &= Z_k - Z_{k-1} - \mathbb{E}_{\mathcal{B}_{k-1}} [Z_k - Z_{k-1}] \\ &= Z_k - Z_{k-1} - \left(\mathbb{E}_{\mathcal{B}_{k-1}} Z_k - Z_{k-1}\right) = Z_k - \mathbb{E}_{\mathcal{B}_{k-1}} Z_k .\end{aligned}$$

■

*Remark 20.18.* Suppose that $X = \{X_n\}_{n=0}^\infty$ is a submartingale and $X_n = M_n + A_n$ is it Doob decomposition. Then $A_\infty =\uparrow \lim_{n\to\infty} A_n$ exists a.s.,

$$\mathbb{E}A_n = \mathbb{E}[X_n - M_n] = \mathbb{E}X_n - \mathbb{E}M_0 = \mathbb{E}[X_n - X_0] \tag{20.10}$$

and hence by MCT,

$$\mathbb{E}A_\infty =\uparrow \lim_{n\to\infty} \mathbb{E}[X_n - X_0] . \tag{20.11}$$

Hence if $\lim_{n\to\infty} \mathbb{E}[X_n - X_0] = \sup_n \mathbb{E}[X_n - X_0] < \infty$, then $\mathbb{E}A_\infty < \infty$ and so by DCT, $A_n \to A_\infty$ in $L^1(\Omega, \mathcal{B}, P)$. In particular if $\sup_n \mathbb{E}|X_n| < \infty$, we may conclude that $\{X_n\}_{n=0}^\infty$ is $L^1(\Omega, \mathcal{B}, P)$ convergent iff $\{M_n\}_{n=0}^\infty$ is $L^1(\Omega, \mathcal{B}, P)$ convergent. (We will see below in Corollary 20.56 that $X_\infty := \lim_{n\to\infty} X_n$ and $M_\infty := \lim_{n\to\infty} M_n$ exist almost surely under the assumption that $\sup_n \mathbb{E}|X_n| < \infty$.)

*Example 20.19.* Suppose that $N = \{N_n\}_{n=0}^\infty$ is a square integrable martingale, i.e. $\mathbb{E}N_n^2 < \infty$ for all $n$. Then from Proposition 20.12, $X := \left\{X_n = N_n^2\right\}_{n=0}^\infty$ is a positive submartingale. In this case

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X &= \mathbb{E}_{\mathcal{B}_{k-1}} \left(N_k^2 - N_{k-1}^2\right) = \mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1})(N_k + N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1})(N_k - N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}} (N_k - N_{k-1})^2\end{aligned}$$

wherein the second to last equality we have used

$$\mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1}) N_{k-1}] = N_{k-1} \mathbb{E}_{\mathcal{B}_{k-1}} (N_k - N_{k-1}) = 0 \text{ a.s.}$$

in order to change $(N_k + N_{k-1})$ to $(N_k - N_{k-1})$. Hence the increasing predictable process, $A_n$, in the Doob decomposition may be written as

$$A_n = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} (\Delta_k N)^2 . \tag{20.12}$$

**Exercise 20.4 (Very similar to above example?).** Suppose $\{M_n\}_{n=0}^\infty$ is a square integrable martingale. Show;

1. $\mathbb{E}\left[M_{n+1}^2 - M_n^2 | \mathcal{B}_n\right] = \mathbb{E}\left[(M_{n+1} - M_n)^2 | \mathcal{B}_n\right]$. Conclude from this that the Doob decomposition of $M_n^2$ is of the form,
$$M_n^2 = N_n + A_n$$
where
$$A_n := \sum_{1\le k\le n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{B}_{k-1}\right] .$$
2. If we further assume that $M_k - M_{k-1}$ is independent of $\mathcal{B}_{k-1}$ for all $k = 1, 2, \ldots,$ explain why,
$$A_n = \sum_{1\le k\le n} \mathbb{E}(M_k - M_{k-1})^2 .$$

The next exercise shows how to characterize Markov processes via martingales.

**Exercise 20.5 (Martingale problem I).** Suppose that $\{X_n\}_{n=0}^\infty$ is an $(S, \mathcal{S})$ – valued adapted process on some filtered probability space $\left(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n\in\mathbb{N}_0}, P\right)$ and $Q$ is a probability kernel on $S$. To each $f : S \to \mathbb{R}$ which is bounded and measurable, let

$$M_n^f := f(X_n) - \sum_{k<n} (Qf(X_k) - f(X_k)) = f(X_n) - \sum_{k<n} ((Q - I) f)(X_k) .$$

Show;

1. If $\{X_n\}_{n\geq 0}$ is a time homogeneous Markov chain with transition kernel, $Q$, then $\left\{M_n^f\right\}_{n\geq 0}$ is a martingale for each $f \in \mathcal{S}_b$.
2. Conversely if $\left\{M_n^f\right\}_{n\geq 0}$ is a martingale for each $f \in \mathcal{S}_b$, then $\{X_n\}_{n\geq 0}$ is a time homogeneous Markov chain with transition kernel, $Q$.

*Remark 20.20.* If $X$ is a real valued random variable, then $X = X^+ - X^-$, $|X| = X^+ + X^-$, $X^+ \leq |X| = 2X^+ - X$, so that

$$\mathbb{E}X^+ \leq \mathbb{E}|X| = 2\mathbb{E}X^+ - \mathbb{E}X.$$

Hence if $\{X_n\}_{n=0}^{\infty}$ is a submartingale then

$$\mathbb{E}X_n^+ \leq \mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_n^+ - \mathbb{E}X_0$$

from which it follows that

$$\sup_n \mathbb{E}X_n^+ \leq \sup_n \mathbb{E}|X_n| \leq 2\sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0. \tag{20.13}$$

In particular, an integrable submartingale $\{X_n\}_{n=0}^{\infty}$ is $L^1(P)$ bounded iff $\{X_n^+\}_{n=0}^{\infty}$ is $L^1(P)$ bounded.

**Theorem 20.21 (Krickeberg Decomposition).** *Suppose that $X$ is an integrable submartingale such that $C := \sup_n \mathbb{E}[X_n^+] < \infty$ or equivalently $\sup_n \mathbb{E}|X_n| < \infty$, see Eq. (20.13). Then*

$$M_n :=\uparrow \lim_{p\to\infty} \mathbb{E}\left[X_p^+|\mathcal{B}_n\right] \text{ exists a.s.},$$

*$M = \{M_n\}_{n=0}^{\infty}$ is a positive martingale, $Y = \{Y_n\}_{n=0}^{\infty}$ with $Y_n := X_n - M_n$ is a positive supermartingale, and hence $X_n = M_n - Y_n$. So $X$ can be decomposed into the difference of a positive martingale and a positive supermartingale.*

**Proof.** From Proposition 20.12 we know that $X^+ = \{X_n^+\}$ is a still a positive submartingale. Therefore for each $n \in \mathbb{N}$, and $p \geq n$,

$$\mathbb{E}_{\mathcal{B}_n}\left[X_{p+1}^+\right] = \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_p}\left[X_{p+1}^+\right] \geq \mathbb{E}_{\mathcal{B}_n} X_p^+ \text{ a.s.}$$

Therefore $\mathbb{E}_{\mathcal{B}_n} X_p^+$ is increasing in $p$ for $p \geq n$ and therefore, $M_n := \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]$ exists in $[0,\infty]$. By Fatou's lemma, we know that

$$\mathbb{E}M_n \leq \liminf_{p\to\infty} \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]\right] \leq \liminf_{p\to\infty} \mathbb{E}\left[X_p^+\right] = C < \infty$$

which shows $M$ is integrable. By cMCT and the tower property of conditional expectation,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n} M_{n+1} &= \mathbb{E}_{\mathcal{B}_n} \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] = \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] \\ &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] = M_n \text{ a.s.},\end{aligned}$$

which shows $M = \{M_n\}$ is a martingale.

We now define $Y_n := M_n - X_n$. Using the submartingale property of $X^+$ implies,

$$\begin{aligned}Y_n = M_n - X_n &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n = \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n^+ + X_n^- \\ &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+ - X_n^+\right] + X_n^- \geq 0 \text{ a.s.}.\end{aligned}$$

Moreover,

$$\mathbb{E}\left[Y_{n+1}|\mathcal{B}_n\right] = \mathbb{E}\left[M_{n+1} - X_{n+1}|\mathcal{B}_n\right] = M_n - \mathbb{E}\left[X_{n+1}|\mathcal{B}_n\right] \geq M_n - X_n = Y_n$$

wherein we have use $M$ is a martingale in the second equality and $X$ is submartingale the last inequality. ■

## 20.3 Stopping Times

**Definition 20.22.** *Again let $\{\mathcal{B}_n\}_{n=0}^{\infty}$ be a filtration on $(\Omega, \mathcal{B})$ and assume that $\mathcal{B} = \mathcal{B}_\infty := \vee_{n=0}^{\infty}\mathcal{B}_n := \sigma\left(\cup_{n=0}^{\infty}\mathcal{B}_n\right)$. A function, $\tau : \Omega \to \bar{\mathbb{N}} := \mathbb{N} \cup \{0, \infty\}$ is said to be a stopping time if $\{\tau \leq n\} \in \mathcal{B}_n$ for all $n \in \bar{\mathbb{N}}$. Equivalently put, $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff the process, $n \to 1_{\tau\leq n}$ is adapted.*

**Lemma 20.23.** *Let $\{\mathcal{B}_n\}_{n=0}^{\infty}$ be a filtration on $(\Omega, \mathcal{B})$ and $\tau : \Omega \to \bar{\mathbb{N}}$ be a function. Then the following are equivalent;*

1. *$\tau$ is a stopping time.*
2. *$\{\tau \leq n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*
3. *$\{\tau > n\} = \{\tau \geq n+1\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*
4. *$\{\tau = n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.*

*Moreover if any of these conditions hold for $n \in \mathbb{N}_0$ then they also hold for $n = \infty$.*

**Proof.** (1. $\iff$ 2.) Observe that if $\{\tau \leq n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$, then $\{\tau < \infty\} = \cup_{n=1}^{\infty}\{\tau \leq n\} \in \mathcal{B}_\infty$ and therefore $\{\tau = \infty\} = \{\tau < \infty\}^c \in \mathcal{B}_\infty$ and hence $\{\tau \leq \infty\} = \{\tau < \infty\} \cup \{\tau = \infty\} \in \mathcal{B}_\infty$. Hence in order to check that $\tau$ is a stopping time, it suffices to show $\{\tau \leq n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.

The equivalence of 2., 3., and 4. follows from the identities

$$\{\tau > n\}^c = \{\tau \le n\},$$
$$\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}, \text{ and}$$
$$\{\tau \le n\} = \cup_{k=0}^n \{\tau = k\}$$

from which we conclude that 2. $\implies$ 3. $\implies$ 4. $\implies$ 1. ■

Clearly any constant function, $\tau : \Omega \to \bar{\mathbb{N}}$, is a stopping time. The reader should also observe that if $\mathcal{B}_n = \sigma(X_0, \dots, X_n)$, then $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff, for each $n \in \mathbb{N}_0$ there exists a measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $1_{\{\tau=n\}} = f_n(X_0, \dots, X_n)$. In other words, if $\tau(\omega) = n$ and $\omega'$ is any other point in $\Omega$ such that $X_k(\omega) = X_k(\omega')$ for $k \le n$ then $\tau(\omega') = n$. Here is another common example of a stopping time.

*Example 20.24 (Hitting times).* Let $(S, \mathcal{S})$ be a state space, $X := \{X_n : \Omega \to S\}_{n=0}^\infty$ be an adapted process on the filtered space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ and $A \in \mathcal{S}$. Then the **first hitting time of** $A$,

$$\tau := \inf\{n \in \mathbb{N}_0 : X_n \in A\},$$

(with convention that $\inf \emptyset = \infty$) is a stopping time. To see this, observe that

$$\{\tau = n\} = \{X_0 \in A^c, \dots, X_{n-1} \in A^c, X_n \in A\} \in \sigma(X_0, \dots, X_n) \subset \mathcal{B}_n.$$

More generally if $\sigma$ is a stopping time, then the **first hitting time after** $\sigma$,

$$\tau := \inf\{k \ge \sigma : X_k \in A\},$$

is also a stopping time. Indeed,

$$\begin{aligned}\{\tau = n\} &= \{\sigma \le n\} \cap \{X_\sigma \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \\ &= \cup_{0 \le k \le n} \{\sigma = k\} \cap \{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\}\end{aligned}$$

which is in $\mathcal{B}_n$ for all $n$. Here we use the convention that

$$\{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\} = \{X_n \in A\} \text{ if } k = n.$$

On the other hand the last hitting time, $\tau = \sup\{n \in \mathbb{N}_0 : X_n \in A\}$, of a set $A$ is typically not a stopping time. Indeed, in this case

$$\{\tau = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \dots\} \in \sigma(X_n, X_{n+1}, \dots)$$

which typically will not be in $\mathcal{B}_n$.

**Proposition 20.25 (New Stopping Times from Old).** *Let $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ be a filtered measure space and suppose $\sigma$, $\tau$, and $\{\tau_n\}_{n=1}^\infty$ are all stopping times. Then*

1. *$\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are all stopping times.*
2. *If $\tau_k \uparrow \tau_\infty$ or $\tau_k \downarrow \tau_\infty$, then $\tau_\infty$ is a stopping time.*
3. *In general, $\sup_k \tau_k = \lim_{k\to\infty} \max\{\tau_1, \dots, \tau_k\}$ and $\inf_k \tau_k = \lim_{k\to\infty} \min\{\tau_1, \dots, \tau_k\}$ are also stopping times.*

**Proof.**

1. Since $\{\tau \wedge \sigma > n\} = \{\tau > n\} \cap \{\sigma > n\} \in \mathcal{B}_n$, $\{\tau \vee \sigma \le n\} = \{\tau \le n\} \cap \{\sigma \le n\} \in \mathcal{B}_n$ for all $n$, and
$$\{\tau + \sigma = n\} = \cup_{k=0}^n \{\tau = k, \sigma = n-k\} \in \mathcal{B}_n$$
for all $n$, $\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are all stopping times.
2. If $\tau_k \uparrow \tau_\infty$, then $\{\tau_\infty \le n\} = \cap_k \{\tau_k \le n\} \in \mathcal{B}_n$ and so $\tau_\infty$ is a stopping time. Similarly, if $\tau_k \downarrow \tau_\infty$, then $\{\tau_\infty > n\} = \cap_k \{\tau_k > n\} \in \mathcal{B}_n$ and so $\tau_\infty$ is a stopping time. (Recall that $\{\tau_\infty > n\} = \{\tau_\infty \ge n+1\}$.)
3. This follows from items 1. and 2.

■

**Lemma 20.26.** *If $\tau$ is a stopping time, then the processes, $f_n := 1_{\{\tau \le n\}}$, and $f_n := 1_{\{\tau = n\}}$ are adapted and $f_n := 1_{\{\tau < n\}}$ is predictable. Moreover, if $\sigma$ and $\tau$ are two stopping times, then $f_n := 1_{\sigma < n \le \tau}$ is predictable.*

**Proof.** These are all trivial to prove. For example, if $f_n := 1_{\sigma < n \le \tau}$, then $f_n$ is $\mathcal{B}_{n-1}$ measurable since,

$$\{\sigma < n \le \tau\} = \{\sigma < n\} \cap \{n \le \tau\} = \{\sigma < n\} \cap \{\tau < n\}^c \in \mathcal{B}_{n-1}.$$

■

**Notation 20.27 (Stochastic intervals)** *If $\sigma, \tau : \Omega \to \bar{\mathbb{N}}$, let*

$$(\sigma, \tau] := \left\{(\omega, n) \in \Omega \times \bar{\mathbb{N}} : \sigma(\omega) < n \le \tau(\omega)\right\}$$

*and we will write $1_{(\sigma,\tau]}$ for the process, $1_{\sigma < n \le \tau}$.*

Our next goal is to define the "stopped" $\sigma$ – algebra, $\mathcal{B}_\tau$. To motivate the upcoming definition, suppose $X_n : \Omega \to \mathbb{R}$ are given functions for all $n \in \mathbb{N}_0$, $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$, and $\tau : \Omega \to \mathbb{N}_0$ is a $\mathcal{B}_\cdot$ – stopping time. Recalling that a function $Y : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ measurable iff $Y(\omega) = f_n(X_0(\omega), \dots X_n(\omega))$ for some measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$, it is reasonable to suggest that $Y$ is $\mathcal{B}_\tau$ measurable iff $Y(\omega) = f_{\tau(\omega)}\left(X_0(\omega), \dots X_{\tau(\omega)}(\omega)\right)$, where $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ are measurable random variables. If this is the case, then we would have $1_{\tau=n} Y = f_n(X_0, \dots, X_n)$ is $\mathcal{B}_n$ – measurable for all $n$. Hence we should define $A \subset \Omega$ to be in $\mathcal{B}_\tau$ iff $1_A$ is $\mathcal{B}_\tau$ measurable iff $1_{\tau=n} 1_A$ is $\mathcal{B}_n$ measurable for all $n$ which happens iff $\{\tau = n\} \cap A \in \mathcal{B}_n$ for all $n$.

**Definition 20.28 (Stopped $\sigma-$ algebra).** *Given a stopping time $\tau$ on a filtered measure space $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty)$ with $\mathcal{B}_\infty:=\vee_{n=0}^\infty\mathcal{B}_n:=\sigma(\cup_{n=0}^\infty\mathcal{B}_n)$, let*

$$\mathcal{B}_\tau:=\{A\subset\Omega:\{\tau=n\}\cap A\in\mathcal{B}_n \text{ for all } n\le\infty\}. \tag{20.14}$$

**Lemma 20.29.** *Suppose $\sigma$ and $\tau$ are stopping times.*

1. *A set, $A\subset\Omega$ is in $\mathcal{B}_\tau$ iff $A\cap\{\tau\le n\}\in\mathcal{B}_n$ for all $n\le\infty$.*
2. *$\mathcal{B}_\tau$ is a sub-$\sigma$-algebra of $\mathcal{B}_\infty$.*
3. *$\tau$ is $\mathcal{B}_\tau$ – measurable.*
4. *For all $n\in\bar{\mathbb{N}}_0$, $\{\tau=n\}\in\mathcal{B}_\tau\cap\mathcal{B}_n$ and $\mathcal{B}_\tau=\mathcal{B}_n$ on $\{\tau=n\}$.*
5. *If $\sigma\le\tau$, then $\mathcal{B}_\sigma\subset\mathcal{B}_\tau$.*

**Proof.** We take each item in turn.

1. Since
$$A\cap\{\tau\le n\}=\cup_{k\le n}[A\cap\{\tau\le k\}] \text{ and}$$
$$A\cap\{\tau=n\}=[A\cap\{\tau\le n\}]\setminus[A\cap\{\tau\le n-1\}],$$
it easily follows that $A\subset\Omega$ is in $\mathcal{B}_\tau$ iff $A\cap\{\tau\le n\}\in\mathcal{B}_n$ for all $n\le\infty$.
2. 2. Since $\Omega\cap\{\tau\le n\}=\{\tau\le n\}\in\mathcal{B}_n$ for all $n$, it follows that $\Omega\in\mathcal{B}_\tau$. If $A\in\mathcal{B}_\tau$, then, for all $n\in\mathbb{N}_0$,
$$A^c\cap\{\tau\le n\}=\{\tau\le n\}\setminus A=\{\tau\le n\}\setminus[A\cap\{\tau\le n\}]\in\mathcal{B}_n.$$
This shows $A^c\in\mathcal{B}_\tau$. Similarly if $\{A_k\}_{k=1}^\infty\subset\mathcal{B}_\tau$, then
$$\{\tau\le n\}\cap(\cap_{k=1}^\infty A_k)=\cap_{k=1}^\infty(\{\tau\le n\}\cap A_k)\in\mathcal{B}_n$$
and hence $\cap_{k=1}^\infty A_k\in\mathcal{B}_\tau$. This completes the proof the $\mathcal{B}_\tau$ is a $\sigma-$ algebra. Since $A=A\cap\{\tau\le\infty\}$, it also follows that $\mathcal{B}_\tau\subset\mathcal{B}_\infty$.
3. For $n,k\in\bar{\mathbb{N}}_0$ we have
$$\{\tau=n\}\cap\{\tau=k\}=\begin{cases}\emptyset & \text{if } k\ne n\\ \{\tau=k\} & \text{if } k=n\end{cases}$$
and so in any case $\{\tau=n\}\cap\{\tau=k\}\in\mathcal{B}_k$. This show $\{\tau=n\}\in\mathcal{B}_\tau$ for all $n\in\bar{\mathbb{N}}_0$ and therefore $\tau$ is $\mathcal{B}_\tau$ – measurable.
4. For all $n\in\bar{\mathbb{N}}_0$, $\{\tau=n\}\in\mathcal{B}_n$ as $\tau$ is a stopping time and $\{\tau=n\}\in\mathcal{B}_\tau$ as $\tau$ is $\mathcal{B}_\tau$ – measurable and so $\{\tau=n\}\in\mathcal{B}_\tau\cap\mathcal{B}_n$. For the second assertion, if $A\in(\mathcal{B}_\tau)_{\{\tau=n\}}$, then $A\in\mathcal{B}_\tau$ and $A\subset\{\tau=n\}$ which implies $A=A\cap\{\tau=n\}\in\mathcal{B}_n$ and so $A\in(\mathcal{B}_n)_{\{\tau=n\}}$. Conversely, if $A\in(\mathcal{B}_n)_{\{\tau=n\}}$, then $A\in\mathcal{B}_n$ and $A\subset\{\tau=n\}$. Thus if $k\in\bar{\mathbb{N}}_0$ we have
$$A\cap\{\tau=k\}=\begin{cases}\emptyset \text{ if } k\ne n\\ A \text{ if } k=n\end{cases}$$
from which it follows that $A\cap\{\tau=k\}\in\mathcal{B}_k$ for all $k\in\bar{\mathbb{N}}_0$, i.e. $A\in[\mathcal{B}_\tau]_{\{\tau=n\}}$.
5. Now suppose $\sigma\le\tau$ and $A\in\mathcal{B}_\sigma$. Since $A\cap\{\sigma\le n\}$ and $\{\tau\le n\}$ are in $\mathcal{B}_n$ for all $n\le\infty$, we find
$$A\cap\{\tau\le n\}=[A\cap\{\sigma\le n\}]\cap\{\tau\le n\}\in\mathcal{B}_n\ \forall\ n\le\infty$$
which shows $A\in\mathcal{B}_\tau$.

■

**Proposition 20.30 ($\mathcal{B}_\tau$ – measurable random variables).** *Let $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty)$ be a filtered measure space. Let $\tau$ be a stopping time and $Z:\Omega\to\mathbb{R}$ be a function. Then the following are equivalent;*

1. *$Z$ is $\mathcal{B}_\tau$ – measurable,*
2. *$1_{\{\tau\le n\}}Z$ is $\mathcal{B}_n$ – measurable for all $n\le\infty$,*
3. *$1_{\{\tau=n\}}Z$ is $\mathcal{B}_n$ – measurable for all $n\le\infty$.*
4. *There exists, $Y_n:\Omega\to\mathbb{R}$ which are $\mathcal{B}_n$ – measurable for all $n\le\infty$ such that*
$$Z=Y_\tau=\sum_{n\in\bar{\mathbb{N}}}1_{\{\tau=n\}}Y_n.$$

**Proof.** 1. $\Longrightarrow$ 2. By definition, if $A\in\mathcal{B}_\tau$, then $1_{\{\tau\le n\}}1_A=1_{\{\tau\le n\}\cap A}$ is $\mathcal{B}_n$ – measurable for all $n\le\infty$. Consequently any simple $\mathcal{B}_\tau$ – measurable function, $Z$, satisfies $1_{\{\tau\le n\}}Z$ is $\mathcal{B}_n$ – measurable for all $n$. So by the usual limiting argument (Theorem 8.39), it follows that $1_{\{\tau\le n\}}Z$ is $\mathcal{B}_n$ – measurable for all $n$ for any $\mathcal{B}_\tau$ – measurable function, $Z$.

2. $\Longrightarrow$ 3. This property follows from the identity,

$$1_{\{\tau=n\}}Z=1_{\{\tau\le n\}}Z-1_{\{\tau<n\}}Z.$$

3. $\Longrightarrow$ 4. Simply take $Y_n=1_{\{\tau=n\}}Z$.

4. $\Longrightarrow$ 1. Since $Z=\sum_{n\in\bar{\mathbb{N}}}1_{\{\tau=n\}}Y_n$, it suffices to show $1_{\{\tau=n\}}Y_n$ is $\mathcal{B}_\tau$ – measurable if $Y_n$ is $\mathcal{B}_n$ – measurable. Further, by the usual limiting arguments using Theorem 8.39, it suffices to assume that $Y_n=1_A$ for some $A\in\mathcal{B}_n$. In this case $1_{\{\tau=n\}}Y_n=1_{A\cap\{\tau=n\}}$. Hence we must show $A\cap\{\tau=n\}\in\mathcal{B}_\tau$ which indeed is true because

$$A\cap\{\tau=n\}\cap\{\tau=k\}=\begin{cases}\emptyset\in\mathcal{B}_k & \text{if } k\ne n\\ A\cap\{\tau=n\}\in\mathcal{B}_k & \text{if } k=n\end{cases}.$$

**Alternatively proof for** 1. $\Longrightarrow$ 2. If $Z$ is $\mathcal{B}_\tau$ measurable, then $\{Z\in B\}\cap\{\tau\le n\}\in\mathcal{B}_n$ for all $n\le\infty$ and $B\in\mathcal{B}_\mathbb{R}$. Hence if $B\in\mathcal{B}_\mathbb{R}$ with $0\notin B$, then

$$\left\{1_{\{\tau\le n\}}Z\in B\right\}=\{Z\in B\}\cap\{\tau\le n\}\in\mathcal{B}_n \text{ for all } n$$

and similarly,

$$\left\{1_{\{\tau\le n\}}Z=0\right\}^c=\left\{1_{\{\tau\le n\}}Z\neq 0\right\}=\{Z\neq 0\}\cap\{\tau\le n\}\in\mathcal{B}_n \text{ for all } n.$$

From these two observations, it follows that $\left\{1_{\{\tau\le n\}}Z\in B\right\}\in\mathcal{B}_n$ for all $B\in\mathcal{B}_\mathbb{R}$ and therefore, $1_{\{\tau\le n\}}Z$ is $\mathcal{B}_n$ – measurable. ■

**Exercise 20.6.** Suppose $\tau$ is a stopping time, $(S,\mathcal{S})$ is a measurable space, and $Z:\Omega\to S$ is a function. Show that $Z$ is $\mathcal{B}_\tau/\mathcal{S}$ measurable iff $Z|_{\{\tau=n\}}$ is $(\mathcal{B}_n)_{\{\tau=n\}}/\mathcal{S}$ – measurable for all $n\in\bar{\mathbb{N}}_0$.

**Lemma 20.31 ($\mathcal{B}_\sigma$ – conditioning).** *Suppose $\sigma$ is a stopping time and $Z\in L^1(\Omega,\mathcal{B},P)$ or $Z\ge 0$, then*

$$\mathbb{E}[Z|\mathcal{B}_\sigma]=\sum_{n\le\infty}1_{\sigma=n}\mathbb{E}[Z|\mathcal{B}_n]=Y_\sigma \tag{20.15}$$

*where*

$$Y_n:=\mathbb{E}[Z|\mathcal{B}_n] \text{ for all } n\in\bar{\mathbb{N}}. \tag{20.16}$$

**Proof.** From Lemma 20.29 we know $\mathcal{B}_\sigma=\mathcal{B}_n$ on $\{\sigma=n\}$ and therefore by the localization Lemma 16.29,

$$1_{\{\sigma=n\}}\mathbb{E}_{\mathcal{B}_\sigma}Z=1_{\{\sigma=n\}}\mathbb{E}_{\mathcal{B}_n}Z \text{ a.s.}$$

Summing this equation on $n$ shows

$$\mathbb{E}_{\mathcal{B}_\sigma}Z=\sum_{n\le\infty}1_{\{\sigma=n\}}\mathbb{E}_{\mathcal{B}_n}Z=Y_\sigma.$$

**Alternative direct proof.** By Proposition 20.30, $Y_\sigma$ is $\mathcal{B}_\sigma$ – measurable. Moreover if $Z$ is integrable, then

$$\begin{aligned}\sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}|Y_n|\right]&=\sum_{n\le\infty}\mathbb{E}1_{\{\sigma=n\}}|\mathbb{E}[Z|\mathcal{B}_n]|\\&\le\sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}\mathbb{E}[|Z||\mathcal{B}_n]\right]\\&=\sum_{n\le\infty}\mathbb{E}\left[\mathbb{E}\left[1_{\{\sigma=n\}}|Z||\mathcal{B}_n\right]\right]\\&=\sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}|Z|\right]=\mathbb{E}|Z|<\infty \end{aligned}\tag{20.17}$$

and therefore

$$\begin{aligned}\mathbb{E}|Y_\sigma|&=\mathbb{E}\left|\sum_{n\le\infty}\left[1_{\{\sigma=n\}}Y_n\right]\right|\\&\le\sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}|Y_n|\right]\le\mathbb{E}|Z|<\infty.\end{aligned}$$

Furthermore if $A\in\mathcal{B}_\sigma$, then

$$\begin{aligned}\mathbb{E}[Z:A]&=\sum_{n\le\infty}\mathbb{E}[Z:A\cap\{\sigma=n\}]=\sum_{n\le\infty}\mathbb{E}[Y_n:A\cap\{\sigma=n\}]\\&=\sum_{n\le\infty}\mathbb{E}\left[1_{\{\sigma=n\}}Y_n:A\right]=\mathbb{E}\left[\sum_{n\le\infty}1_{\{\sigma=n\}}Y_n:A\right]\\&=\mathbb{E}[Y_\sigma:A],\end{aligned}$$

wherein the interchange of the sum and the expectation in the second to last equality is justified by the estimate in 20.17 or by the fact that everything in sight is positive when $Z\ge 0$. ■

Theorem 20.32 extends the tower property of conditional expectations to conditioning relative to stopped $\sigma$ – algebras. Some of the results of the next exercise are useful in the proof of this theorem.

**Exercise 20.7.** Suppose $\sigma$ and $\tau$ are two stopping times. Show;

1. $\{\sigma<\tau\},\{\sigma=\tau\}$, and $\{\sigma\le\tau\}^*$ are all in $\mathcal{B}_\sigma\cap\mathcal{B}_\tau$,
2. $\mathcal{B}_{\sigma\wedge\tau}=\mathcal{B}_\sigma\cap\mathcal{B}_\tau$,
3. $\mathcal{B}_{\sigma\vee\tau}=\mathcal{B}_\sigma\vee\mathcal{B}_\tau:=\sigma(\mathcal{B}_\sigma\cup\mathcal{B}_\tau)$,[2] and
4. $\mathcal{B}_\sigma=\mathcal{B}_{\sigma\wedge\tau}$ on $C$ where $C$ is any one of the following three sets; $\{\sigma\le\tau\}$, $\{\sigma<\tau\}$, or $\{\sigma=\tau\}$.

*As an example, since

$$\{\sigma\le\tau\}\cap\{\sigma\wedge\tau=n\}=\{\sigma\le\tau\}\cap\{\sigma=n\}=\{n\le\tau\}\cap\{\sigma=n\}\in\mathcal{B}_n$$

for all $n\in\mathbb{N}_0$, it follows that $\{\sigma\le\tau\}\in\mathcal{B}_\sigma\cap\mathcal{B}_\tau$.

**Theorem 20.32 (Tower Property II).** *Let $X\in L^1(\Omega,\mathcal{B},P)$ or $X:\Omega\to[0,\infty]$ be a $\mathcal{B}$ – measurable function. If $\sigma$ and $\tau$ are* ***any*** *two stopping times, then*

$$1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}},\ 1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}=1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}},\ \text{and} \tag{20.18}$$

$$\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X=\mathbb{E}_{\mathcal{B}_\tau}\mathbb{E}_{\mathcal{B}_\sigma}X=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X. \tag{20.19}$$

**Proof.** As usual it suffices to consider the case where $X\ge 0$ and this case there will be now convergence issues to worry about. Equation 20.18 follows from the localization Lemma 16.29 as explained in the second proof. Nevertheless let us first give a self-contained proof.

**First Proof.** Notice that

[2] In fact, you will likely show in your proof that every set in $\mathcal{B}_\sigma\vee\mathcal{B}_\tau$ may be written as a disjoint union of a set from $\mathcal{B}_\sigma$ with a set from $\mathcal{B}_\tau$.

$$1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}=\sum_{n\le\infty}1_{\tau\le\sigma}1_{\tau=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau\le\sigma}\sum_{n\le\infty}1_{\tau\wedge\sigma=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}$$

and similarly,

$$1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}=1_{\tau>\sigma}\sum_{n\le\infty}1_{\sigma=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau>\sigma}\sum_{n\le\infty}1_{\tau\wedge\sigma=n}\mathbb{E}_{\mathcal{B}_n}=1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.$$

Using these remarks and the fact that $\{\tau\le\sigma\}$ and $\{\tau>\sigma\}$ are both in $\mathcal{B}_{\sigma\wedge\tau}=\mathcal{B}_\sigma\cap\mathcal{B}_\tau$ we find;

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}&=\mathbb{E}_{\mathcal{B}_\sigma}\left(1_{\tau\le\sigma}+1_{\tau>\sigma}\right)\mathbb{E}_{\mathcal{B}_\tau}=\mathbb{E}_{\mathcal{B}_\sigma}1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}\\&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}\mathbb{E}_{\mathcal{B}_\tau}\\&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.\end{aligned}$$

**Second Proof.** In this proof we are going to make use of the localization Lemma 16.29. Since $\mathcal{B}_{\sigma\wedge\tau}\subset\mathcal{B}_\sigma$, it follows by item 4. of Exercise 20.7 that $\mathcal{B}_\sigma=\mathcal{B}_{\sigma\wedge\tau}$ on $\{\sigma\le\tau\}$ and on $\{\sigma<\tau\}$. We will actually use the first statement in the form, $\mathcal{B}_\tau=\mathcal{B}_{\sigma\wedge\tau}$ on $\{\tau\le\sigma\}$. From Lemma 20.31, we have

$$\begin{aligned}1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}\text{ and}\\1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}&=1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}.\end{aligned}$$

Using these relations and the basic properties of conditional expectation we arrive at,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X&=\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}\left[1_{\tau\le\sigma}X+1_{\tau>\sigma}X\right]\\&=\mathbb{E}_{\mathcal{B}_\sigma}\left[1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\tau}X\right]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X\\&=\mathbb{E}_{\mathcal{B}_\sigma}\left[1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X\right]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}\mathbb{E}_{\mathcal{B}_\tau}X\\&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_\sigma}\left[\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X\right]+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X\\&=1_{\tau\le\sigma}\mathbb{E}_{\mathcal{B}_{\tau\wedge\sigma}}X+1_{\tau>\sigma}\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X=\mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X\text{ a.s.}\end{aligned}$$

■

**Exercise 20.8.** Show, by example, that it is not necessarily true that

$$\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2}=\mathbb{E}_{\mathcal{G}_1\wedge\mathcal{G}_2}$$

for arbitrary $\mathcal{G}_1$ and $\mathcal{G}_2$ – sub-sigma algebras of $\mathcal{B}$.

**Hint:** it suffices to take $(\Omega,\mathcal{B},P)$ with $\Omega=\{1,2,3\}$, $\mathcal{B}=2^\Omega$, and $P(\{j\})=\frac{1}{3}$ for $j=1,2,3$.

**Exercise 20.9 (Geometry of commuting projections).** Suppose that $H$ is a Hilbert space and $H_i\subset H$ for $i=1,2$ are two closed subspaces. Let $P_i=P_{H_i}$ denote orthogonal projection onto $H_i$ and $P=P_M$ be orthogonal projection onto $M:=H_1\cap H_2$. Show;

1. Suppose there exists $M_0\subset H_1\cap H_2$ such that $M_1\perp M_2$ where $M_i=\{h\in H_i:h\perp M_0\}$ so that $H_1=M_0\overset{\perp}{\oplus}M_1$ and $H_2=M_0\overset{\perp}{\oplus}M_2$. Then $M_0=H_1\cap H_2$ and $P_1P_2=P=P_2P_1$.
2. If $P_1P_2=P_2P_1$, then $P_1P_2=P=P_2P_1$. Moreover if we let $M_0=H_1\cap H_2$ and $M_i$ be as above, then $M_1\perp M_2$.

**Exercise 20.10.** Let $\sigma$ and $\tau$ be stopping times and apply the results of Exercise 20.9 with $M_0:=L^2(\Omega,\mathcal{B}_{\sigma\wedge\tau},P)$, $H_1:=L^2(\Omega,\mathcal{B}_\sigma,P)$, and $H_2=L^2(\Omega,\mathcal{B}_\tau,P)$ to give another proof of Theorem 20.32.

### 20.3.1 Summary of some of the more notable Chapter 20 convergence results

As a guide to the reader, let us pause to summarize some of the key convergence results which are going to be proved in the remainder of this chapter. Suppose that $\{M_n\}_{n=1}^\infty$ is a martingale. Recall that $\{M_n\}$ is a **regular martingale** if $M_n=\mathbb{E}[X|\mathcal{B}_n]$ for some $X\in L^1(P)$. Here is a list of some of the key convergence results to come.

1. If $M:=\{M_n\}_{n=0}^\infty$ is an $L^1$ – bounded martingale (i.e. $C:=\sup_n\mathbb{E}|M_n|<\infty$), then $M_\infty:=\lim_{n\to\infty}M_n$ exists a.s. and satisfies, $\mathbb{E}|M_\infty|<\infty$, see Corollary 20.56.
2. Suppose $X=\{X_n\}_{n=0}^\infty$ is a supermartingale, martingale, or submartingale with either $\mathbb{E}|X_n|<\infty$ for all $n$ or $X_n\ge0$ for all $n$. Then for every stopping time, $\tau$, $X^\tau$ is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, or submartingale respectively, see the optional stopping Theorem 20.39.
3. Suppose that $\sigma$ and $\tau$ are two stopping times and $\tau$ is bounded, i.e. there exists $N\in\mathbb{N}$ such that $\tau\le N<\infty$ a.s. If $X=\{X_n\}_{n=0}^\infty$ is a supermartingale, martingale, or submartingale, with either $\mathbb{E}|X_n|<\infty$ of $X_n\ge0$ for all $0\le n\le N$, then
$$\mathbb{E}[X_\tau|\mathcal{B}_\sigma]\begin{matrix}\le\\=\\\ge\end{matrix}X_{\sigma\wedge\tau}\text{ a.s.}$$
respectively from top to bottom, see the optional stopping Theorem 20.40.
4. Suppose that $M:=\{M_n\}_{n=0}^\infty$ is an $L^1$ – bounded martingale then (see Theorem 20.67) the following are equivalent;
   a) $M:=\{M_n\}_{n=0}^\infty$ is a regular martingale.
   b) $M_n=\mathbb{E}[M_\infty|\mathcal{B}_n]$ for $n\in\mathbb{N}$.
   c) $M_n\to M_\infty$ in $L^1(\Omega,\mathcal{B},P)$.
   d) $\{M_n\}_{n=0}^\infty$ is uniformly integrable.
5. If $1<p<\infty$ and $M:=\{M_n\}_{n=0}^\infty$ is an $L^p$ – bounded martingale. Then $M_n\to M_\infty$ almost surely and in $L^p$. In particular, $\{M_n\}$ is a regular martingale, see Theorem 20.69.

6. Suppose that $M = \{M_n\}_{n=0}^{\infty}$ is a regular martingale, $\sigma$ and $\tau$ are **arbitrary** stopping times, then (see the optional stopping Theorem 20.70)
$$M_\tau = \mathbb{E}[M_\infty|\mathcal{B}_\tau], \quad \mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty \text{ and } \mathbb{E}[M_\tau|\mathcal{B}_\sigma] = M_{\sigma\wedge\tau} \text{ a.s..}$$
7. Let $\{\mathcal{B}_n : n \le 0\}$ be a reverse filtration (still have $\mathcal{B}_m \subset \mathcal{B}_n$ if $m \le n$) and $\{X_n\}_{n\le 0}$ be a backwards submartingale, i.e. $\mathbb{E}[X_n|\mathcal{B}_m] \ge X_m$ if $m \le n$. The Backwards (or reverse) submartingale convergence theorem 20.79 asserts $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. in $\{-\infty\} \cup \mathbb{R}$ and $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$. If we further assume that
$$C := \lim_{n\to-\infty} \mathbb{E}X_n = \inf_{n\le 0} \mathbb{E}X_n > -\infty,$$
then $X_n = M_n + A_n$ where
a) $\{M_n\}_{-\infty<n\le 0}$ is a martingale, $\{A_n\}_{-\infty<n\le 0}$ is a predictable process such that $A_{-\infty} = \lim_{n\to-\infty} A_n = 0$,
b) $\{X_n\}_{n\le 0}$ is uniformly integrability,
c) $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$, and
d) $\lim_{n\to-\infty} \mathbb{E}|X_n - X_{-\infty}| = 0$.

## 20.4 Stochastic Integrals and Optional Stopping

**Notation 20.33** *Suppose that* $\{c_n\}_{n=1}^{\infty}$ *and* $\{x_n\}_{n=0}^{\infty}$ *are two sequences of numbers, let* $c \cdot \Delta x = \{(c\cdot\Delta x)_n\}_{n\in\mathbb{N}_0}$ *denote the sequence of numbers defined by* $(c\cdot\Delta x)_0 = 0$ *and*
$$(c\cdot\Delta x)_n = \sum_{j=1}^{n} c_j (x_j - x_{j-1}) = \sum_{j=1}^{n} c_j \Delta_j x \text{ for } n \ge 1.$$
*(For convenience of notation later we will interpret* $\sum_{j=1}^{0} c_j\Delta_j x = 0$.*)*

For a gambling interpretation of $(c\cdot\Delta x)_n$, let $x_j$ represent the price of a stock at time $j$. Suppose that you, the investor, buys $c_j$ shares at time $j-1$ and then sells these shares back at time $j$. With this interpretation, $c_j\Delta_j x$ represents your profit (or loss if negative) in the time interval from $j-1$ to $j$ and $(c\cdot\Delta x)_n$ represents your profit (or loss) from time 0 to time $n$. By the way, if you want to buy 5 shares of the stock at time $n = 3$ and then sell them all at time 9, you would take $c_k = 5\cdot 1_{3<k\le 9}$ so that
$$(c\cdot\Delta x)_9 = 5\cdot\sum_{3<k\le 9}\Delta_k x = 5\cdot(x_9 - x_3)$$
would represent your profit (loss) for this transaction. The next example formalizes this observation.

*Example 20.34.* Suppose that $0 \le \sigma \le \tau$ where $\sigma, \tau \in \bar{\mathbb{N}}_0$ and let $c_n := 1_{\sigma<n\le\tau}$. Then
$$(c\cdot\Delta x)_n = \sum_{j=1}^{n} 1_{\sigma<j\le\tau}(x_j - x_{j-1}) = \sum_{j=1}^{\infty} 1_{\sigma<j\le\tau\wedge n}(x_j - x_{j-1})$$
$$= \sum_{j=1}^{\infty} 1_{\sigma\wedge n<j\le\tau\wedge n}(x_j - x_{j-1}) = x_{\tau\wedge n} - x_{\sigma\wedge n}.$$
More generally if $\sigma, \tau \in \bar{\mathbb{N}}_0$ are arbitrary and $c_n := 1_{\sigma<n\le\tau}$ we will have $c_n := 1_{\sigma\wedge\tau<n\le\tau}$ and therefore
$$(c\cdot\Delta x)_n = x_{\tau\wedge n} - x_{\sigma\wedge\tau\wedge n}.$$

**Proposition 20.35 (The Discrete Stochastic Integral).** *Let* $X = \{X_n\}_{n=0}^{\infty}$ *be an adapted integrable process, i.e.* $\mathbb{E}|X_n| < \infty$ *for all* $n$. *If* $X$ *is a martingale and* $\{C_n\}_{n=1}^{\infty}$ *is a predictable sequence of bounded random variables, then* $\{(C\cdot\Delta X)_n\}_{n=1}^{\infty}$ *is still a martingale. If* $X := \{X_n\}_{n=0}^{\infty}$ *is a submartingale (supermartingale) (necessarily real valued) and* $C_n \ge 0$, *then* $\{(C\cdot\Delta X)_n\}_{n=1}^{\infty}$ *is a submartingale (supermartingale).*

*Conversely if* $X$ *is an adapted process of integrable functions such that* $\mathbb{E}[(C\cdot\Delta X)_n] = 0$ *for all bounded predictable processes,* $\{C_n\}_{n=1}^{\infty}$, *then* $X$ *is a martingale. Similarly if* $X$ *is real valued adapted process such that*
$$\mathbb{E}[(C\cdot\Delta X)_n] \overset{\le}{\underset{\ge}{=}} 0 \tag{20.20}$$
*for all* $n$ *and for all bounded, non-negative predictable processes,* $C$, *then* $X$ *is a supermartingale, martingale, or submartingale respectively. (In other words,* $X$ *is a sub-martingale if no matter what your (non-negative) betting strategy is you will make money on average.)*

**Proof.** For any adapted process $X$, we have
$$\begin{aligned}\mathbb{E}\left[(C\cdot\Delta X)_{n+1}|\mathcal{B}_n\right] &= \mathbb{E}\left[(C\cdot\Delta X)_n + C_{n+1}(X_{n+1}-X_n)|\mathcal{B}_n\right]\\ &= (C\cdot\Delta X)_n + C_{n+1}\mathbb{E}\left[(X_{n+1}-X_n)|\mathcal{B}_n\right].\end{aligned} \tag{20.21}$$
The first assertions easily follow from this identity.

Now suppose that $X$ is an adapted process of integrable functions such that $\mathbb{E}[(C\cdot\Delta X)_n] = 0$ for all bounded predictable processes, $\{C_n\}_{n=1}^{\infty}$. Taking expectations of Eq. (20.21) then allows us to conclude that
$$\mathbb{E}\left[C_{n+1}\mathbb{E}\left[(X_{n+1}-X_n)|\mathcal{B}_n\right]\right] = 0$$
for all bounded $\mathcal{B}_n$ – measurable random variables, $C_{n+1}$. Taking $C_{n+1} := \operatorname{sgn}(\mathbb{E}[(X_{n+1}-X_n)|\mathcal{B}_n])$ shows $|\mathbb{E}[(X_{n+1}-X_n)|\mathcal{B}_n]| = 0$ a.s. and hence $X$ is

a martingale. Similarly, if for all non-negative, predictable $C$, Eq. (20.20) holds for all $n \geq 1$, and $C_n \geq 0$, then taking $A \in \mathcal{B}_n$ and $C_k = \delta_{k,n+1} 1_A$ in Eq. (20.13) allows us to conclude that

$$\mathbb{E}\left[X_{n+1} - X_n : A\right] = \mathbb{E}\left[(C \cdot \Delta X)_{n+1}\right] \overset{\leq}{\underset{\geq}{=}} 0,$$

i.e. $X$ is a supermartingale, martingale, or submartingale respectively. ■

*Example 20.36.* Suppose that $\{X_n\}_{n=0}^{\infty}$ are mean zero independent integrable random variables and $f_k : \mathbb{R}^k \to \mathbb{R}$ are bounded measurable functions for $k \in \mathbb{N}$. Then $\{Y_n\}_{n=0}^{\infty}$, defined by $Y_0 = 0$ and

$$Y_n := \sum_{k=1}^{n} f_k\left(X_0, \dots, X_{k-1}\right)\left(X_k - X_{k-1}\right) \text{ for } n \in \mathbb{N}, \tag{20.22}$$

is a martingale sequence relative to $\left\{\mathcal{B}_n^X\right\}_{n \geq 0}$.

**Notation 20.37** *Given an adapted process, $X$, and a stopping time $\tau$, let $X_n^\tau := X_{\tau \wedge n}$. We call $X^\tau := \{X_n^\tau\}_{n=0}^{\infty}$ the* ***process $X$ stopped by*** *$\tau$.*

Observe that

$$|X_n^\tau| = |X_{\tau \wedge n}| = \left|\sum_{0 \leq k \leq n} 1_{\tau = k} X_k\right| \leq \sum_{0 \leq k \leq n} 1_{\tau = k} |X_k| \leq \sum_{0 \leq k \leq n} |X_k|,$$

so that $X_n^\tau \in L^1(P)$ for all $n$ provided $X_n \in L^1(P)$ for all $n$.

*Example 20.38.* Suppose that $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale, with $\mathbb{E}|X_n| < \infty$ and let $\sigma$ and $\tau$ be stopping times. Then for any $A \in \mathcal{B}_\sigma$, the process $C_n := 1_A \cdot 1_{\sigma < n \leq \tau}$ is predictable since for all $n \in \mathbb{N}$ we have

$$\begin{aligned} A \cap \{\sigma < n \leq \tau\} &= (A \cap \{\sigma < n\}) \cap \{n \leq \tau\} \\ &= (A \cap \{\sigma \leq n - 1\}) \cap \{\tau \leq n - 1\}^c \in \mathcal{B}_{n-1}. \end{aligned}$$

Therefore by Proposition 20.35, $\{(C \cdot \Delta X)_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale respectively where

$$\begin{aligned} (C \cdot \Delta X)_n &= \sum_{k=1}^{n} 1_A \cdot 1_{\sigma < k \leq \tau} \Delta_k X = 1_A \cdot \sum_{k=1}^{n} 1_{\sigma \wedge \tau < k \leq \tau} \Delta_k X \\ &= \sum_{k=1}^{\infty} 1_A \cdot 1_{\sigma \wedge \tau \wedge n < k \leq \tau \wedge n} \Delta_k X = 1_A \left(X_n^\tau - X_n^{\sigma \wedge \tau}\right). \end{aligned}$$

**Theorem 20.39 (Optional stopping theorem).** *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale with either $\mathbb{E}|X_n| < \infty$ for all $n$ or $X_n \geq 0$ for all $n$. Then for every stopping time, $\tau$, $X^\tau$ is a $\{\mathcal{B}_n\}_{n=0}^{\infty}$ – supermartingale, martingale, or submartingale respectively.*

**Proof.** When $\mathbb{E}|X_n| < \infty$ for all $n \geq 0$ we may take $\sigma = 0$ and $A = \Omega$ in Example 20.38 in order to learn that $\{X_n^\tau - X_0\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale respectively and therefore so is $\{X_n^\tau = X_0 + X_n^\tau - X_0\}_{n=0}^{\infty}$. When $X_n$ is only non-negative we have to give a different proof which does not involve any subtractions (which might be undefined).

For the second proof we simply observe that $1_{\tau \leq n} X_\tau = \sum_{k=0}^{n} 1_{\tau = k} X_k$ is $\mathcal{B}_n$ measurable, $\{\tau > n\} \in \mathcal{B}_n$, and

$$X_{\tau \wedge (n+1)} = 1_{\tau \leq n} X_\tau + 1_{\tau > n} X_{n+1}.$$

Therefore

$$\begin{aligned} \mathbb{E}_{\mathcal{B}_n}\left[X_{(n+1)}^\tau\right] = \mathbb{E}_{\mathcal{B}_n}\left[X_{\tau \wedge (n+1)}\right] &= 1_{\tau \leq n} X_\tau + 1_{\tau > n} \mathbb{E}_{\mathcal{B}_n} X_{n+1} \\ &\begin{Bmatrix} \leq \\ = \\ \geq \end{Bmatrix} 1_{\tau \leq n} X_\tau + 1_{\tau > n} X_n = X_{\tau \wedge n}, \end{aligned}$$

where the top, middle, bottom (in)equality holds depending on whether $X$ is a supermartingale, martingale, or submartingale respectively. (This second proof works for both cases at once. For another proof see Remark 20.41.) ■

**Theorem 20.40 (Optional sampling theorem I).** *Suppose that $\sigma$ and $\tau$ are two stopping times and $\tau$ is bounded, i.e. there exists $N \in \mathbb{N}$ such that $\tau \leq N < \infty$ a.s. If $X = \{X_n\}_{n=0}^{\infty}$ is a supermartingale, martingale, or submartingale, with either $\mathbb{E}|X_n| < \infty$ of $X_n \geq 0$ for all $0 \leq n \leq N$, then*

$$\mathbb{E}\left[X_\tau | \mathcal{B}_\sigma\right] \overset{\leq}{\underset{\geq}{=}} X_{\sigma \wedge \tau} \text{ a.s.} \tag{20.23}$$

*respectively*[3] *from top to bottom.*

**Proof.** First suppose that $\mathbb{E}|X_n| < \infty$ for $0 \leq n \leq N$ and let $A \in \mathcal{B}_\sigma$. From Example 20.38 we know that $1_A\left(X_n^\tau - X_n^{\sigma \wedge \tau}\right)$ is a supermartingale, martingale, or submartingale respectively and in particular for all $n \in \mathbb{N}_0$ we have

$$\mathbb{E}\left[1_A\left(X_n^\tau - X_n^{\sigma \wedge \tau}\right)\right] \overset{\leq}{\underset{\geq}{=}} 0 \text{ respectively.}$$

[3] This is the natural generalization of Eq. (20.3) to the stopping time setting.

Taking $n = N$ in this equation using $\sigma\wedge\tau \le \tau \le N$ then implies, for all $A \in \mathcal{B}_\sigma$, that

$$\mathbb{E}\left[(X_\tau - X_{\sigma\wedge\tau}) : A\right] \overset{\le}{\underset{\ge}{=}} 0 \text{ respectively}$$

and this is equivalent to Eq. (20.23).

When we only assume that $X_n \ge 0$ for all $n$ we again have to give a different proof which avoids subtractions which may be undefined. One way to do this is to use Theorem 20.39 in order to conclude that $X^\tau$ is a supermartingale, martingale, or submartingale respectively and in particular that

$$\mathbb{E}\left[X_\tau|\mathcal{B}_n\right] = \mathbb{E}\left[X_N^\tau|\mathcal{B}_n\right] \overset{\le}{\underset{\ge}{=}} X_{n\wedge N}^\tau \text{ for all } n \le \infty.$$

Combining this result with Lemma 20.31 then implies

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] = \sum_{n\le\infty} 1_{\sigma=n}\mathbb{E}\left[X_\tau|\mathcal{B}_n\right] \overset{\le}{\underset{\ge}{=}} \sum_{n\le\infty} 1_{\sigma=n}X_{n\wedge N}^\tau = X_{\sigma\wedge N}^\tau = X_{\sigma\wedge\tau}. \quad (20.24)$$

(This second proof again covers both cases at once!) ■

**Exercise 20.11.** Give another proof of Theorem 20.40 when $\mathbb{E}|X_n| < \infty$ by using the tower property in Theorem 20.32 along with the Doob decomposition of Lemma 20.17.

**Exercise 20.12.** Give yet another (full) proof of Theorem 20.40 using the following outline;

1. Show by induction on $n$ starting with $n = N$ that

$$\mathbb{E}\left[X_\tau|\mathcal{B}_n\right] \overset{\le}{\underset{\ge}{=}} X_{\tau\wedge n} \text{ a.s. for all } 0 \le n \le N. \quad (20.25)$$

2. Observe the above inequality holds as an equality for $n > N$ as well.
3. Combine this result with Lemma 20.17 to complete the proof.

This argument makes it clear why we must at least initially assume that $\tau \le N$ for some $N \in \mathbb{N}$. To relax this restriction will require a limiting argument which will be the topic of Section 20.8 below.

*Remark 20.41.* Theorem 20.40 can be used to give a simple proof of the Optional stopping Theorem 20.39. For example, if $X = \{X_n\}_{n=0}^\infty$ is a submartingale and $\tau$ is a stopping time, then

$$\mathbb{E}_{\mathcal{B}_n} X_{\tau\wedge(n+1)} \ge X_{[\tau\wedge(n+1)]\wedge n} = X_{\tau\wedge n},$$

i.e. $X^\tau$ is a submartingale.

## 20.5 Submartingale Maximal Inequalities

**Notation 20.42 (Running Maximum)** *If $X = \{X_n\}_{n=0}^\infty$ is a sequence of (extended) real numbers, we let*

$$X_N^* := \max\{X_0, \dots, X_N\}. \quad (20.26)$$

**Proposition 20.43 (Maximal Inequalities of Bernstein and Lévy).** *Let $\{X_n\}$ be a submartingale on a filtered probability space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$. Then[4] for any $a \ge 0$ and $N \in \mathbb{N}$,*

$$aP(X_N^* \ge a) \le \mathbb{E}\left[X_N : X_N^* \ge a\right] \le \mathbb{E}\left[X_N^+\right], \quad (20.27)$$

$$aP\left(\min_{n\le N} X_n \le -a\right) \le \mathbb{E}\left[X_N : \min_{k\le N} X_k > -a\right] - \mathbb{E}\left[X_0\right] \quad (20.28)$$
$$\le \mathbb{E}\left[X_N^+\right] - \mathbb{E}\left[X_0\right], \quad (20.29)$$

*and*

$$aP\left(|X|_N^* \ge a\right) \le 2\mathbb{E}\left[X_N^+\right] - \mathbb{E}\left[X_0\right]. \quad (20.30)$$

**Proof.** Let $\tau := \inf\{n : X_n \ge a\}$ and observe that

$$X_N^* \ge X_\tau \ge a \text{ on } \{\tau \le N\} = \{X_N^* \ge a\} \quad (20.31)$$

and (by the optional sampling Theorem 20.40) $\mathbb{E}\left[X_N|\mathcal{B}_\tau\right] \ge X_{N\wedge\tau}$. Since $\{\tau \le N\} \in \mathcal{B}_{\tau\wedge N} \subset \mathcal{B}_\tau$, we learn

$$\mathbb{E}\left[X_\tau : \tau \le N\right] = \mathbb{E}\left[X_{\tau\wedge N} : \tau \le N\right] \le \mathbb{E}\left[\mathbb{E}\left[X_N|\mathcal{B}_\tau\right] : \tau \le N\right] = \mathbb{E}\left[X_N : \tau \le N\right]$$

which combined with Eq. (20.31) implies,

$$\begin{aligned} a \cdot P(X_N^* \ge a) &= \mathbb{E}\left[a : X_N^* \ge a\right] = \mathbb{E}\left[a : \tau \le N\right] \\ &\le \mathbb{E}\left[X_\tau : \tau \le N\right] \le \mathbb{E}\left[X_N : \tau \le N\right] = \mathbb{E}\left[X_N : X_N^* \ge a\right] \\ &\le \mathbb{E}\left[X_N^+ : X_N^* \ge a\right] \le \mathbb{E}\left[X_N^+\right], \end{aligned}$$

i.e. Eq. (20.27) holds.

**More generally** if $X$ is **any** integrable process and $\tau$ is the random time defined by, $\tau := \inf\{n : X_n \ge a\}$ we still have Eq. (20.31) and

$$\begin{aligned} aP(X_N^* \ge a) &= \mathbb{E}\left[a : \tau \le N\right] \\ &\le \mathbb{E}\left[X_\tau : \tau \le N\right] \quad (20.32) \\ &= \mathbb{E}\left[X_N : \tau \le N\right] - \mathbb{E}\left[X_N - X_\tau : \tau \le N\right] \\ &= \mathbb{E}\left[X_N : \tau \le N\right] - \mathbb{E}\left[X_N - X_{\tau\wedge N}\right]. \quad (20.33) \end{aligned}$$

[4] The first inequality is the most important.

Let me emphasize again that in deriving Eq. (20.33), we have **not** used any special properties (not even adaptedness) of $X$. If $X$ is now assumed to be a submartingale, by the optional sampling Theorem 20.40, $\mathbb{E}_{\mathcal{B}_{\tau\wedge N}} X_N \geq X_{\tau\wedge N}$ and in particular $\mathbb{E}[X_N - X_{\tau\wedge N}] \geq 0$. Combining this observation with Eq. (20.33) and Eq. (20.31) again gives Eq. (20.27).

Secondly we may apply Eq. (20.33) with $X_n$ replaced by $-X_n$ to find

$$aP\left(\min_{n\leq N} X_n \leq -a\right) = aP\left(-\min_{n\leq N} X_n \geq a\right) = aP\left(\max_{n\leq N}(-X_n) \geq a\right) \\ \leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] \tag{20.34}$$

where now,

$$\tau := \inf\{n : -X_n \geq a\} = \inf\{n : X_n \leq -a\}.$$

By the optional sampling Theorem 20.40, $\mathbb{E}[X_{\tau\wedge N} - X_0] \geq 0$ and adding this to right side of Eq. (20.34) gives the estimate

$$\begin{aligned} aP\left(\min_{n\leq N} X_n \leq -a\right) &\leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] + \mathbb{E}[X_{\tau\wedge N} - X_0] \\ &\leq \mathbb{E}[X_N - X_0] - \mathbb{E}[X_N : \tau \leq N] \\ &= \mathbb{E}[X_N : \tau > N] - \mathbb{E}[X_0] \\ &= \mathbb{E}\left[X_N : \min_{k\leq N} X_k > -a\right] - \mathbb{E}[X_0] \end{aligned}$$

which proves Eq. (20.28) and hence Eq. (20.29). Adding Eqs. (20.27) and (20.29) gives the estimate in Eq. (20.30) since

$$\left\{|X|_N^* \geq a\right\} = \{X_N^* \geq a\} \cup \left\{\min_{n\leq N} X_n \leq -a\right\}.$$

■

*Remark 20.44.* It is of course possible to give a direct proof of Proposition 20.43. For example,

$$\begin{aligned} \mathbb{E}\left[X_N : \max_{n\leq N} X_n \geq a\right] &= \sum_{k=1}^N \mathbb{E}[X_N : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[X_k : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[a : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &= aP\left(\max_{n\leq N} X_n \geq a\right) \end{aligned}$$

which proves Eq. (20.27).

**Corollary 20.45.** *Suppose that $\{Y_n\}_{n=1}^\infty$ is a non-negative supermartingale, $a > 0$ and $N \in \mathbb{N}$, then*

$$aP\left(\max_{n\leq N} Y_n \geq a\right) \leq \mathbb{E}[Y_0 \wedge a] - \mathbb{E}\left[Y_N : \max_{n\leq N} Y_n < a\right] \leq \mathbb{E}[Y_0 \wedge a]. \tag{20.35}$$

**Proof.** Let $X_n := -Y_n$ in Eq. (20.28) to learn

$$aP\left(\min_{n\leq N}(-Y_n) \leq -a\right) \leq \mathbb{E}\left[-Y_N : \min_{n\leq N}(-Y_n) > -a\right] + \mathbb{E}[Y_0]$$

or equivalently that

$$aP\left(\max_{n\leq N} Y_n \geq a\right) \leq \mathbb{E}[Y_0] - \mathbb{E}\left[Y_N : \max_{n\leq N} Y_n < a\right] \leq \mathbb{E}[Y_0]. \tag{20.36}$$

Since $\varphi_a(x) := a \wedge x$ is concave and nondecreasing, it follows by Jensen's inequality that

$$\mathbb{E}[\varphi_a(Y_n)|\mathcal{B}_m] \leq \varphi_a(\mathbb{E}[Y_n|\mathcal{B}_m]) \leq \varphi_a(Y_n) \text{ for all } n \geq m.$$

In this way we see that $\varphi_a(Y_n) = Y_n \wedge a$ is a supermartingale as well. Applying Eq. (20.36) with $Y_n$ replaced by $Y_n \wedge a$ proves Eq. (20.35). ■

**Lemma 20.46.** *Suppose that $X$ and $Y$ are two non-negative random variables such that $P(Y \geq y) \leq \frac{1}{y}\mathbb{E}[X : Y \geq y]$ for all $y > 0$. Then for all $p \in (1,\infty)$,*

$$\mathbb{E}Y^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p. \tag{20.37}$$

**Proof.** We will begin by proving Eq. (20.37) under the additional assumption that $Y \in L^p(\Omega, \mathcal{B}, P)$. Since

$$\begin{aligned} \mathbb{E}Y^p &= p\mathbb{E}\int_0^\infty 1_{y\leq Y} \cdot y^{p-1} dy = p\int_0^\infty \mathbb{E}[1_{y\leq Y}] \cdot y^{p-1} dy \\ &= p\int_0^\infty P(Y \geq y) \cdot y^{p-1} dy \leq p\int_0^\infty \frac{1}{y}\mathbb{E}[X : Y \geq y] \cdot y^{p-1} dy \\ &= p\mathbb{E}\int_0^\infty X 1_{y\leq Y} \cdot y^{p-2} dy = \frac{p}{p-1}\mathbb{E}\left[XY^{p-1}\right]. \end{aligned}$$

Now apply Hölder's inequality, with $q = p(p-1)^{-1}$, to find

$$\mathbb{E}\left[XY^{p-1}\right] \leq \|X\|_p \cdot \left\|Y^{p-1}\right\|_q = \|X\|_p \cdot [\mathbb{E}|Y|^p]^{1/q}.$$

Combining thew two inequalities shows and solving for $\|Y\|_p$ shows $\|Y\|_p \le \frac{p}{p-1}\|X\|_p$ which proves Eq. (20.37) under the additional restriction of $Y$ being in $L^p(\Omega,\mathcal{B},P)$.

To remove the integrability restriction on $Y$, for $M>0$ let $Z:=Y\wedge M$ and observe that

$$P(Z\ge y)=P(Y\ge y)\le\frac{1}{y}\mathbb{E}[X:Y\ge y]=\frac{1}{y}\mathbb{E}[X:Z\ge y]\text{ if }y\le M$$

while

$$P(Z\ge y)=0=\frac{1}{y}\mathbb{E}[X:Z\ge y]\text{ if }y>M.$$

Since $Z$ is bounded, the special case just proved shows

$$\mathbb{E}[(Y\wedge M)^p]=\mathbb{E}Z^p\le\left(\frac{p}{p-1}\right)^p\mathbb{E}X^p.$$

We may now use the MCT to pass to the limit, $M\uparrow\infty$, and hence conclude that Eq. (20.37) holds in general. ∎

**Corollary 20.47 (Doob's Inequality).** *If $X=\{X_n\}_{n=0}^\infty$ be a non-negative submartingale and $1<p<\infty$, then*

$$\mathbb{E}X_N^{*p}\le\left(\frac{p}{p-1}\right)^p\mathbb{E}X_N^p. \tag{20.38}$$

**Proof.** Equation 20.38 follows by applying Lemma 20.46 with the aid of Proposition 20.43. ∎

**Corollary 20.48 (Doob's Inequality).** *If $\{M_n\}_{n=0}^\infty$ is a martingale and $1<p<\infty$, then for all $a>0$,*

$$P\left(|M|_N^*\ge a\right)\le\frac{1}{a}\mathbb{E}\left[|M|_N:M_N^*\ge a\right]\le\frac{1}{a}\mathbb{E}\left[|M_N|\right] \tag{20.39}$$

*and*

$$\mathbb{E}|M|_N^{*p}\le\left(\frac{p}{p-1}\right)^p\mathbb{E}|M_N|^p. \tag{20.40}$$

**Proof.** By the conditional Jensen's inequality, it follows that $X_n:=|M_n|$ is a submartingale. Hence Eq. (20.39) follows from Eq. (20.27) and Eq. (20.40) follows from Eq. (20.38). ∎

*Example 20.49.* Let $\{X_n\}$ be a sequence of independent integrable random variables with mean zero, $S_0=0$, $S_n:=X_1+\dots+X_n$ for $n\in\mathbb{N}$, and $|S|_n^*=\max_{j\le n}|S_j|$. Since $\{S_n\}_{n=0}^\infty$ is a martingale, by cJensen's inequality, $\{|S_n|^p\}_{n=1}^\infty$ is a (possibly extended) submartingale for any $p\in[1,\infty)$. Therefore an application of Eq. (20.27) of Proposition 20.43 show

$$P\left(|S|_N^*\ge\alpha\right)=P\left(|S|_N^{*p}\ge\alpha^p\right)\le\frac{1}{\alpha^p}\mathbb{E}\left[|S_N|^p:S_N^*\ge\alpha\right].$$

(When $p=2$, this is Kolmogorov's inequality in Theorem 22.46 below.) From Corollary 20.48 we also know that

$$\mathbb{E}|S|_N^{*p}\le\left(\frac{p}{p-1}\right)^p\mathbb{E}|S_N|^p.$$

In particular when $p=2$, this inequality becomes,

$$\mathbb{E}|S|_N^{*2}\le 4\cdot\mathbb{E}|S_N|^2=4\cdot\sum_{n=1}^N\mathbb{E}|X_n|^2.$$

## 20.6 Submartingale Upcrossing Inequality and Convergence Theorems

The main results of this section are consequences of the following example and lemma which say that the optimal strategy for betting on a sub-martingale is to go "all in." Any other strategy, including buy low and sell high, will not fare better (on average) than going *all in.*

*Example 20.50.* Suppose that $\{X_n\}_{n=0}^\infty$ represents the value of a stock which is known to be a sub-martingale. At time $n-1$ you are allowed buy $C_n\in[0,1]$ shares of the stock which you will then sell at time $n$. Your net gain (loss) in this transaction is $C_nX_n-C_nX_{n-1}=C_n\Delta_nX$ and your wealth at time $n$ will be

$$W_n=W_0+\sum_{k=1}^n C_k\Delta_kX.$$

The next lemma asserts that the way to maximize your expected gain is to choose $C_k=1$ for all $k$, i.e. buy the maximum amount of stock you can at each stage. We will refer to this as the **all in** strategy..

**Lemma 20.51 ("All In").** *If $\{X_n\}_{n=0}^\infty$ is a sub-martingale and $\{C_k\}_{k=1}^\infty$ is a previsible process with values in $[0,1]$, then*

$$\mathbb{E}\left(\sum_{k=1}^n C_k\Delta_kX\right)\le\mathbb{E}[X_n-X_0]$$

*with equality when $C_k=1$ for all $k$, i.e. the optimal strategy is to go all in.*

**Proof.** Notice that $\{1 - C_k\}_{k=1}^{\infty}$ is a previsible non-negative process and therefore by Proposition 20.35,

$$\mathbb{E}\left(\sum_{k=1}^{n}(1 - C_k)\,\Delta_k X\right) \geq 0.$$

Since

$$X_n - X_0 = \sum_{k=1}^{n} \Delta_k X = \sum_{k=1}^{n} C_k \Delta_k X + \sum_{k=1}^{n}(1 - C_k)\,\Delta_k X,$$

it follows that

$$\mathbb{E}\left[X_n - X_0\right] = \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right) + \mathbb{E}\left(\sum_{k=1}^{n}(1 - C_k)\,\Delta_k X\right) \geq \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right).$$

■

We are now going to apply Lemma 20.51 to the time honored gambling strategy of buying low and selling high in order to prove the important "upcrossing" inequality of Doob, see Theorem 20.53. To be more precise, suppose that $\{X_n\}_{n=0}^{\infty}$ is a sub-martingale representing a stock price and $-\infty < a < b < \infty$ are given numbers. The (sub-optimal) strategy we wish to employ is to buy the stock when it first drops below $a$ and then sell the first time it rises above $b$ and then repeat this strategy over and over again.

Given a function, $\mathbb{N}_0 \ni n \to X_n \in \mathbb{R}$ and $-\infty < a < b < \infty$, let

$$\tau_0 = \inf\{n \geq 0 : X_n \leq a\},\ \tau_1 = \inf\{n \geq \tau_0 : X_n \geq b\}$$
$$\tau_2 = \inf\{n \geq \tau_1 : X_n \leq a\},\ \tau_3 := \inf\{n \geq \tau_2 : X_n \geq b\}$$
$$\vdots$$
$$\tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \leq a\},\ \tau_{2k+1} := \inf\{n \geq \tau_{2k} : X_n \geq b\} \qquad (20.41)$$
$$\vdots$$

with the usual convention that $\inf \emptyset = \infty$ in the definitions above, see Figures 20.2 and 20.3.

In terms of these stopping time our betting strategy may be describe as,

$$C_n = \sum_{k=0}^{\infty} 1_{\tau_{2k} < n \leq \tau_{2k+1}} \text{ for } n \in \mathbb{N}, \qquad (20.42)$$

see Figure 20.3 for a more intuitive description of $\{C_n\}_{n=1}^{\infty}$.

Observe that $\tau_0 \geq 0$ and $\tau_{n+1} \geq \tau_n + 1$ for all $n \geq 1$ and hence $\tau_n \geq n$ for all $n \geq 0$. Further, for each $N \in \bar{\mathbb{N}}$ let



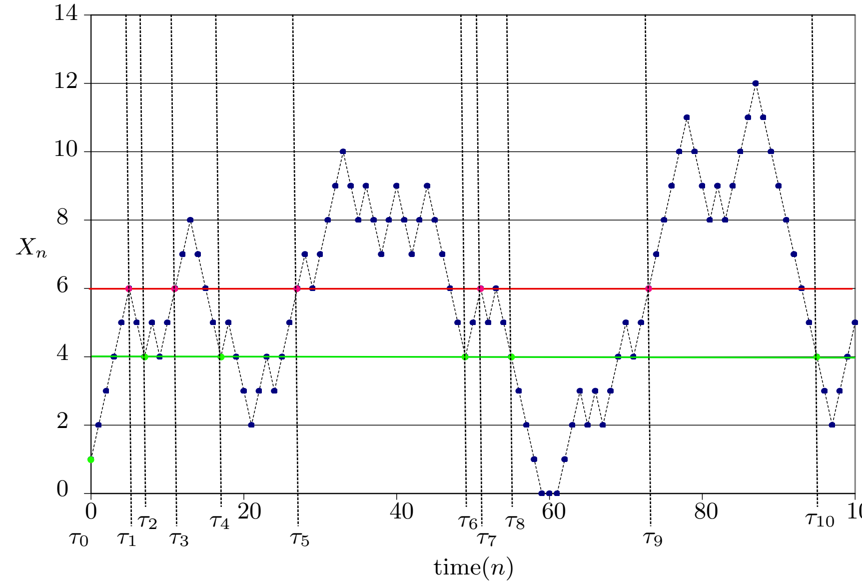



**Fig. 20.2.** A sample path or the positive part of a random walk with level crossing of $a = 1$ and $b = 2$ being marked off.

$$U_N^X(a, b) = \max\{k \geq 1 : \tau_{2k-1} \leq N\} \qquad (20.43)$$

be the **number of upcrossings of** $X$ **across** $[a, b]$ in the time interval, $[0, N]$.

In Figure 20.3 you will notice that there are two upcrossings and at the end we are holding a stock for a loss of no more than $(a - X_N)_+$. In this example $X_0 = 0.90$ and we do not purchase a stock until time 1, i.e. $C_n = 1$ for the first time at $n = 2$. On the other hand if $X_0 < a$, then on the **first** upcrossing we would be guaranteed to make at least

$$b - X_0 = b - a + a - X_0 = b - a + (a - X_0)_+.$$

With these observations in mind, if there is at least one upcrossing, then

$$W_N := \sum_{k=1}^{N} C_k \Delta_k X \geq (b - a)\,U_N^X(a, b) + (a - X_0)_+ - (a - X_N)_+ \qquad (20.44)$$
$$= (b - a)\,U_N^X(a, b) + (X_0 - a)_- - (X_N - a)_-. \qquad (20.45)$$

In words the inequality in Eq. (20.45) states that our net gain in buying at or below $a$ and selling at or above $b$ is at least equal to $(b - a)$ times the number

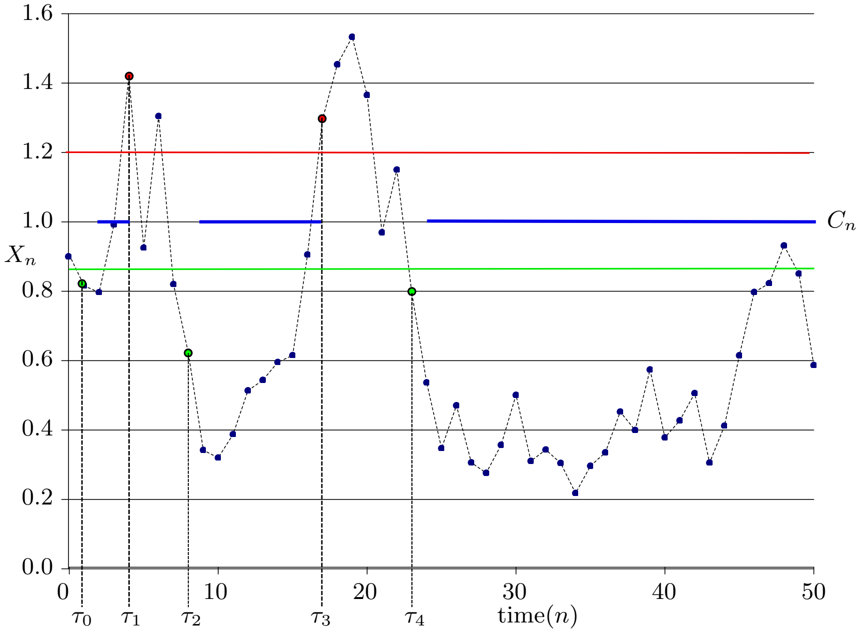



**Fig. 20.3.** In this figure we are taking $a = 0.85$ and $b = 1.20$. There are two upcrossings and we imaging buying below 0.85 and selling above 1.20. The graph of $C_n$ is given in blue in the above figure.

of times we buy low and sell high plus a possible bonus for buying below $a$ at time 0 and a penalty for holding the stock below $a$ at the end of the day. The key inequality in Eq. (20.45) may also be verified when no upcrossings occur. Here are the three case to consider.

1. If $X_n > a$ for all $0 \le n \le N$, then $C_n = 0$ for all $n$ so $W_N = 0$ while $(X_0 - a)_- - (X_N - a)_- = 0 - 0 = 0$ as well.
2. If $X_0 \le a$ and $X_n < b$ for all $0 \le n \le N$, then $C_n = 1$ for all $n$ so that
$$\begin{aligned} W_N &= X_N - X_0 = (X_N - a) - (X_0 - a) \\ &= (X_N - a) + (X_0 - a)_- \ge -(X_N - a)_- + (X_0 - a)_- . \end{aligned}$$
3. If $X_0 > a$, but $\tau_1 \le N$ and $X_n < b$ for all $0 \le n \le N$, then
$$W_N = X_N - X_{\tau_1} \ge X_N - a \ge -(X_N - a)_- = -(X_N - a)_- + (X_0 - a)_- .$$

**Lemma 20.52.** *If $\{X_n\}_{n=0}^\infty$ is a submartingale, then $\{X_n^+ = \max(X_n, 0)\}_{n=0}^\infty$ is a submartingale.*

**Proof.** This follows by an application of the conditional Jensen's inequality applied with $\varphi(x) = x_+$ which is convex and increasing. We may also, however, easily give a direct proof. Indeed, $X_{n+1} \le X_{n+1}^+$ and $0 \le X_{n+1}^+$ and therefore,

$$X_n \le \mathbb{E}_{\mathcal{B}_n} X_{n+1} \le \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+ \text{ and } 0 \le \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+$$

and therefore $X_n^+ = \max\{0, X_n\} \le \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+$. ■

**Theorem 20.53 (Doob's Upcrossing Inequality).** *If $\{X_n\}_{n=0}^\infty$ is a submartingale and $-\infty < a < b < \infty$, then for all $N \in \mathbb{N}$,*

$$\mathbb{E}\left[U_N^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+\right].$$

**First Proof.** Let $\{C_k\}_{k=1}^\infty$ be the buy low sell high strategy defined in Eq. (20.42). Taking expectations of the inequality in Eq. (20.45) making use of Lemma 20.51 implies,

$$\begin{aligned} \mathbb{E}[X_N - a - (X_0 - a)] &= \mathbb{E}[X_N - X_0] \ge \mathbb{E}[(C \cdot \Delta X)_N] \\ &\ge (b-a)\,\mathbb{E}U_N^X(a,b) + \mathbb{E}(X_0 - a)_- - \mathbb{E}(X_N - a)_- . \end{aligned}$$

The result follows from this inequality and the fact that $(X_n - a) = (X_n - a)_+ - (X_n - a)_-$.

**Second Proof.** It is easily verified that $\{X_n - a\}_{n=0}^\infty$ is still a submartingale and then by Lemma 20.52 it follows that $\left\{(X_n - a)_+\right\}_{n=0}^\infty$ is still a sub-martingale.[5] We also note

$$U_N^X(a,b) = U_N^{(X-a)_+}(0, b-a)$$

and if $\{W_n\}_{n=0}^\infty$ are the winnings of the buy at 0 and sell above $b-a$ strategy for $\left\{(X_n - a)_+\right\}_{n=0}^\infty$ then it is easily[6] seen that

$$W_N - W_0 \ge (b-a)\,U_N^{(X-a)_+}(0, b-a) = (b-a)\,U_N^X(a,b).$$

Therefore it follows from Lemma 20.51 that

$$(b-a)\,\mathbb{E}\left[U_N^X(a,b)\right] \le \mathbb{E}[W_N - W_0] \le \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+ .$$

■

[5] Alternatively use Jensen's inequality with $\varphi(x) = (x-a)_+$ which is convex and increasing.
[6] This is where this proof is conceptually a bit simpler than the first proof.

*Remark 20.54 (*Third Proof).* Here is a variant on the above proof which may safely be skipped. We first suppose that $X_n \geq 0$, $a = 0$ and $b > 0$. Let

$$\tau_0 = \inf\{n \geq 0 : X_n = 0\}, \; \tau_1 = \inf\{n \geq \tau_0 : X_n \geq b\}$$
$$\tau_2 = \inf\{n \geq \tau_1 : X_n = 0\}, \; \tau_3 := \inf\{n \geq \tau_2 : X_n \geq b\}$$
$$\vdots$$
$$\tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n = 0\}, \; \tau_{2k+1} := \inf\{n \geq \tau_{2k} : X_n \geq b\}$$
$$\vdots$$

a sequence of stopping times. Suppose that $N$ is given and we choose $k$ such that $2k \geq N$ in which case $\tau_{2k} \geq 2k \geq N$. Thus if $\tau_n' := \tau_n \wedge N$, then $\tau_n' = N$ for all $n \geq 2k$. Therefore,

$$\begin{aligned} X_N - X_0 &= \sum_{n=1}^{2k} \left( X_{\tau_n'} - X_{\tau_{n-1}'} \right) \\ &= \sum_{n=1}^{k} \left( X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} \right) + \sum_{n=1}^{k} \left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) \\ &\geq bU_N^X(0,b) + \sum_{n=1}^{k} \left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right), \end{aligned} \tag{20.46}$$

wherein we have used $X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} \geq b$ if there were an upcrossing in the interval $\left[\tau_{2n-2}', \tau_{2n-1}'\right]$ and $X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} \geq 0$ otherwise, see Figure 20.4. Taking expectations of Eq. (20.46) implies

$$\mathbb{E}X_N - \mathbb{E}X_0 \geq b\mathbb{E}U_N^X(0,b) + \sum_{n=1}^{k} \mathbb{E}\left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) \geq b\mathbb{E}U_N^X(0,b)$$

wherein we have used the optional sampling theorem to guarantee,

$$\mathbb{E}\left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) \geq 0.$$

Following the second proof of Theorem 20.53, $\left\{(X_n - a)_+\right\}_{n=0}^{\infty}$ is still a sub-martingal and

$$U_N^X(a,b) = U_N^{(X-a)_+}(0, b-a)$$

and therefore

$$\begin{aligned} (b-a)\,\mathbb{E}\left[U_N^X(a,b)\right] &= (b-a)\,\mathbb{E}\left[U_N^{(X-a)_+}(0,b-a)\right] \\ &\leq \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+ . \end{aligned}$$

The third proof is now complete, nevertheless it is worth contemplating a bit how is that $\mathbb{E}\left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) \geq 0$ given that are strategy being employed here is to buy high and sell low. On $\{\tau_{2n} \leq N\}$, $X_{\tau_{2n}} - X_{\tau_{2n-1}} = 0 - X_{\tau_{2n-1}} \leq -b$ and therefore,

$$\begin{aligned} 0 \leq& \mathbb{E}\left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) \\ &= \mathbb{E}\left( X_{\tau_{2n}} - X_{\tau_{2n-1}} : \tau_{2n} \leq N \right) + \mathbb{E}\left( X_{\tau_{2n}'} - X_{\tau_{2n-1}'} : \tau_{2n} > N \right) \\ \leq& -bP(\tau_{2n} \leq N) + \mathbb{E}\left( X_N - X_{\tau_{2n-1}'} : \tau_{2n} > N \right). \end{aligned}$$

Therefore we must have

$$\mathbb{E}\left( X_N - X_{\tau_{2n-1} \wedge N} : \tau_{2n} > N \right) \geq bP(\tau_{2n} \leq N)$$

so that $X_N$ must be sufficiently large sufficiently often on the set where $\tau_{2n} > N$.

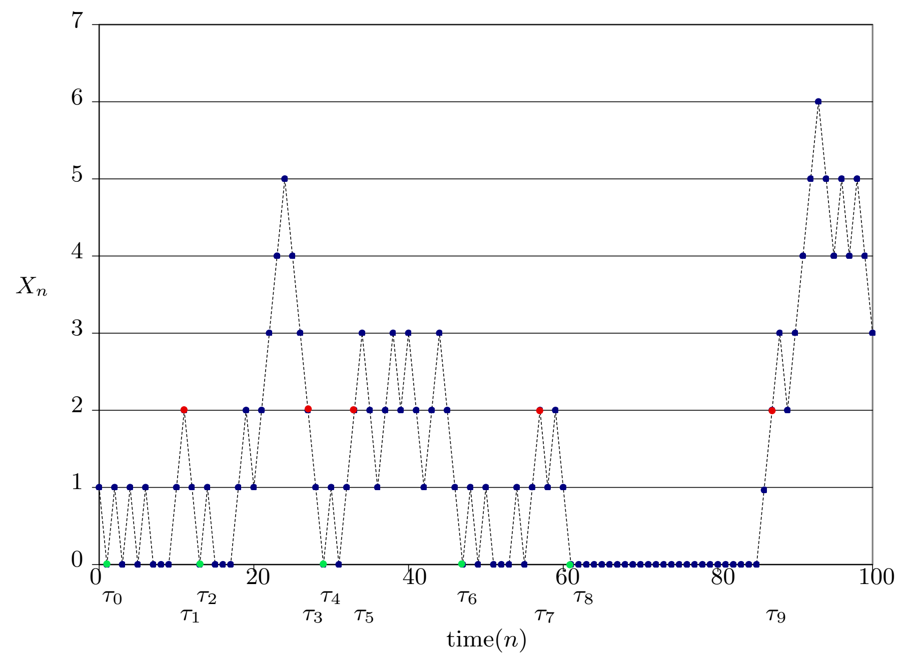



**Fig. 20.4.** A sample path of a positive submartingale along with stopping times $\tau_{2j}$ and $\tau_{2j+1}$ which are the successive hitting times of 0 and 2 respectively. If we take $N = 70$ in this case, then observe that $X_{\tau_7 \wedge 70} - X_{\tau_6 \wedge 70} \geq 2$ while $X_{\tau_9 \wedge 70} - X_{\tau_8 \wedge 70} = 0$.

**Lemma 20.55.** *Suppose* $X = \{X_n\}_{n=0}^{\infty}$ *is a sequence of extended real numbers such that* $U_\infty^X(a,b) < \infty$ *for all* $a, b \in \mathbb{Q}$ *with* $a < b$. *Then* $X_\infty := \lim_{n\to\infty} X_n$ *exists in* $\bar{\mathbb{R}}$.

**Proof.** If $\lim_{n\to\infty} X_n$ does not exists in $\bar{\mathbb{R}}$, then there would exists $a, b \in \mathbb{Q}$ such that

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n$$

and for this choice of $a$ and $b$, we must have $X_n < a$ and $X_n > b$ infinitely often. Therefore, $U_\infty^X(a,b) = \infty$. ■

**Corollary 20.56.** *Suppose* $\{X_n\}_{n=0}^{\infty}$ *is an integrable submartingale such that* $\sup_n \mathbb{E}X_n^+ < \infty$ *(or equivalently* $C := \sup_n \mathbb{E}|X_n| < \infty$, *see Remark 20.20), then* $X_\infty := \lim_{n\to\infty} X_n$ *exists in* $\mathbb{R}$ *a.s. and* $X_\infty \in L^1(\Omega, \mathcal{B}, P)$. *Moreover* $\{X_n\}_{n\in\bar{\mathbb{N}}_0}$ *is a submartingale (that is we also have* $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ *a.s. for all* $n$*), iff* $\{X_n^+\}_{n=1}^{\infty}$ *is uniformly integrable.*

**Proof.** For any $-\infty < a < b < \infty$, by Doob's upcrossing inequality (Theorem 20.53) and the MCT,

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \le \frac{1}{b-a}\left[\sup_N \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+\right] < \infty$$

where

$$U_\infty^X(a,b) := \lim_{N\to\infty} U_N^X(a,b)$$

is the total number of upcrossings of $X$ across $[a,b]$.[7] In particular it follows that

$$\Omega_0 := \cap\left\{U_\infty^X(a,b) < \infty : a, b \in \mathbb{Q} \text{ with } a < b\right\}$$

has probability one. Hence by Lemma 20.55, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists in $\bar{\mathbb{R}}$. By Fatou's lemma we know that

$$\mathbb{E}[|X_\infty|] = \mathbb{E}\left[\liminf_{n\to\infty} |X_n|\right] \le \liminf_{n\to\infty} \mathbb{E}[|X_n|] \le C < \infty \tag{20.47}$$

and therefore that $X_\infty \in \mathbb{R}$ a.s.

Since (as we have already shown) $X_n^+ \to X_\infty^+$ a.s., if $\{X_n^+\}_{n=1}^{\infty}$ is uniformly integrable, then $X_n^+ \to X_\infty^+$ in $L^1(P)$ by Vitalli's convergence Theorem 14.51. Thus if $A \in \mathcal{B}_n$,then using $\mathbb{E}_{\mathcal{B}_n} X_m \ge X_n$ a.s. for $m \ge n$ it follows that

$$\begin{aligned}
\mathbb{E}[X_n 1_A] &\le \limsup_{m\to\infty} \mathbb{E}[X_m 1_A] = \limsup_{m\to\infty}\left(\mathbb{E}\left[X_m^+ 1_A\right] - \mathbb{E}\left[X_m^- 1_A\right]\right) \\
&= \mathbb{E}\left[X_\infty^+ 1_A\right] - \liminf_{m\to\infty} \mathbb{E}\left[X_m^- 1_A\right] \\
&\le \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[\liminf_{m\to\infty} X_m^- 1_A\right] \\
&= \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[X_\infty^- 1_A\right] = \mathbb{E}[X_\infty 1_A],
\end{aligned}$$

wherein Fatou's lemma for the second inequality above. As $A \in \mathcal{B}_n$ was arbitrary, it now follows that $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ a.s. for $n$.

Conversely if we suppose that $X_n \le \mathbb{E}[X_\infty|\mathcal{B}_n]$ a.s. for $n$, then by Lemma 20.52 (or cJensen's inequality with $\varphi(x) = x \vee 0$ being an increasing convex function),

$$X_n^+ \le (\mathbb{E}[X_\infty|\mathcal{B}_n])^+ \le \mathbb{E}\left[X_\infty^+|\mathcal{B}_n\right] \text{ a.s. for all } n$$

and therefore $\{X_n^+\}_{n=1}^{\infty}$ is uniformly integrable by Proposition 20.8 and Exercise 14.5.

**Second Proof.** We may also give another proof of the first assertion based on the Krickeberg decomposition Theorem 20.21 and the supermartingale convergence Corollary 20.65 below. Indeed, by the Krickeberg decomposition Theorem 20.21, $X_n = M_n - Y_n$ where $M$ is a positive martingale and $Y$ is a positive supermartingale. Hence by two applications of Corollary 20.65 we may conclude that

$$X_\infty = \lim_{n\to\infty} X_n = \lim_{n\to\infty} M_n - \lim_{n\to\infty} Y_n$$

exists in $\mathbb{R}$ almost surely. ■

*Remark 20.57.* If $\{X_n\}_{n=0}^{\infty}$ is a submartingale such that $\{X_n^+\}_{n=0}^{\infty}$ is uniformly integrable, it does not necessarily follows that $\{X_n\}_{n=0}^{\infty}$ is uniformly integrable. Indeed, let $X_n = -M_n$ where $M_n$ is the non-uniformly integrable martingale in Example 20.7. Then $X_n$ is a negative (sub)martingale and hence $X_n^+ \equiv 0$ is uniformly integrable but $\{X_n\}_{n=0}^{\infty}$ is **not** uniformly integrable. This also shows that assuming the positive part of a martingale is uniformly integrable is not sufficient to show the martingale itself is uniformly integrable. Keep in mind in this example that $\lim_{n\to\infty} X_n = 0$ a.s. while $\mathbb{E}X_n = 1$ for all $n$ and so clearly $\lim_{n\to\infty} \mathbb{E}X_n = 1 \ne 0 = \mathbb{E}[\lim_{n\to\infty} X_n]$ in this case.

**Notation 20.58** *Given a probability space,* $(\Omega, \mathcal{B}, P)$ *and* $A, B \in \mathcal{B}$, *we say* $A = B$ *a.s. iff* $P(A \triangle B) = 0$ *or equivalently iff* $1_A = 1_B$ *a.s.*

**Corollary 20.59 (Localizing Corollary 20.56).** *Suppose* $M = \{M_n\}_{n=0}^{\infty}$ *is a martingale and* $c < \infty$ *such that* $\Delta_n M := M_n - M_{n-1} \le c$ *a.s. for all* $n$. *Then*

$$\left\{\lim_{n\to\infty} M_n \text{ exists in } \mathbb{R}\right\} = \left\{\sup_n M_n < \infty\right\} \text{ a.s.}$$

[7] Notice that $(X_N - a)_+ \le |X_N - a| \le |X_N| + a$ so that $\sup_N \mathbb{E}(X_N - a)_+ \le C + a < \infty$.

**Proof.** For any $a \in \mathbb{N}$, let $\tau_a := \inf\{n : M_n \geq a\}$. Then by the optional stopping theorem, $n \to M_n^{\tau_a}$ is still a martingale. Since $M_n^{\tau_a} \leq a + c$,[8] it follows that $\mathbb{E}(M_n^{\tau_a})_+ \leq a + c < \infty$ for all $n$. Hence we may apply Corollary 20.56 to conclude, $\lim_{n\to\infty} M_n^{\tau_a} = M_\infty^{\tau_a}$ exists in $\mathbb{R}$ almost surely. Therefore $n \to M_n$ is convergent in $\mathbb{R}$ almost surely on the set

$$\cup_a \{M^{\tau_a} = M\} = \left\{\sup_n M_n < \infty\right\}.$$

Conversely if $n \to M_n$ is convergent in $\mathbb{R}$, then $\sup_n M_n < \infty$. ∎

**Corollary 20.60.** *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale, and $c < \infty$ such that $|\Delta_n M| \leq c$ a.s. for all $n$. Let*

$$C := \left\{\lim_{n\to\infty} M_n \text{ exists in } \mathbb{R}\right\} \text{ and}$$
$$D := \left\{\limsup_{n\to\infty} M_n = \infty \text{ and } \liminf_{n\to\infty} M_n = -\infty\right\}.$$

*Then, $P(C \cup D) = 1$. (In words, either $\lim_{n\to\infty} M_n$ exists in $\mathbb{R}$ or $\{M_n\}_{n=1}^\infty$ is "wildly" oscillating as $n \to \infty$.)*

**Proof.** Since both $M$ and $-M$ satisfy the hypothesis of Corollary 20.59, we may conclude that (almost surely),

$$C = \left\{\sup_n M_n < \infty\right\} = \left\{\inf_n M_n > -\infty\right\} \text{ a.s.}$$

and hence almost surely,

$$C^c = \left\{\sup_n M_n = \infty\right\} = \left\{\inf_n M_n = -\infty\right\}$$
$$= \left\{\sup_n M_n = \infty\right\} \cap \left\{\inf_n M_n = -\infty\right\} = D.$$

∎

**Corollary 20.61.** *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $A_n \in \mathcal{B}_n$ for all $n$. Then*

$$\left\{\sum_n 1_{A_n} = \infty\right\} = \{A_n \text{ i.o.}\} = \left\{\sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty\right\} \text{ a.s.} \quad (20.48)$$

[8] If $n < \tau_a$ then $M_n < a$ and if $n \geq \tau_a$ then $M_n^{\tau_a} = M_{\tau_a} \leq M_{\tau_a - 1} + c < a + c$.

**Proof.** Let $\Delta_n M := 1_{A_n} - \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}]$ so that $\mathbb{E}[\Delta_n M|\mathcal{B}_{n-1}] = 0$ for all $n$. Thus if

$$M_n := \sum_{k\leq n} \Delta_n M = \sum_{k\leq n} (1_{A_n} - \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}]),$$

then $M$ is a martingale with $|\Delta_n M| \leq 1$ for all $n$. Let $C$ and $C$ be as in Corollary 20.60. Since $\{A_n \text{ i.o.}\} = \{\sum_n 1_{A_n} = \infty\}$, it follows that

$$\{A_n \text{ i.o.}\} = \left\{\sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty\right\} \text{ a.s. on } C.$$

Moreover, on $\{\sup_n M_n = \infty\}$ we must have $\sum_n 1_{A_n} = \infty$ and on $\{\inf_n M_n = -\infty\}$ that $\sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty$ and so

$$\sum_n 1_{A_n} = \infty \text{ and } \sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty \text{ a.s. on } D.$$

Thus it follows that Eq. (20.48) holds on $C \cup D$ a.s. which completes the proof since $\Omega = C \cup D$ a.s.. ∎

See Durrett [12, Chapter 4.3] for more in this direction.

## 20.7 *Supermartingale inequalities

As the optional sampling theorem was our basic tool for deriving submartingale inequalities, the following optional switching lemma will be our basic tool for deriving positive supermartingale inequalities.

**Lemma 20.62 (Optional switching lemma).** *Suppose that $X$ and $Y$ are two supermartingales and $\tau$ is a stopping time such that $X_\tau \geq Y_\tau$ on $\{\tau < \infty\}$. Then*

$$Z_n = 1_{n<\tau} X_n + 1_{n\geq\tau} Y_n = \begin{cases} X_n & \text{if } n < \tau \\ Y_n & \text{if } n \geq \tau \end{cases}$$

*is again a supermartingale. (In short we can switch from $X$ to $Y$ at time, $\tau$, provided $Y \leq X$ at the switching time, $\tau$.) This lemma is valid if $X_n, Y_n \in L^1(\Omega, \mathcal{B}_n, P)$ for all $n$ or if both $X_n, Y_n \geq 0$ for all $n$. In the latter case, we should be using the extended notion of conditional expectations.*

**Proof.** We begin by observing,

$$\begin{aligned} Z_{n+1} &= 1_{n+1<\tau} X_{n+1} + 1_{n+1\geq\tau} Y_{n+1} \\ &= 1_{n+1<\tau} X_{n+1} + 1_{n\geq\tau} Y_{n+1} + 1_{\tau=n+1} Y_{n+1} \\ &\leq 1_{n+1<\tau} X_{n+1} + 1_{n\geq\tau} Y_{n+1} + 1_{\tau=n+1} X_{n+1} \\ &= 1_{n<\tau} X_{n+1} + 1_{n\geq\tau} Y_{n+1}. \end{aligned}$$

Since $\{n < \tau\}$ and $\{n \geq \tau\}$ are $\mathcal{B}_n$ – measurable, it now follows from the supermartingale property of $X$ and $Y$ that

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n} Z_{n+1} &\leq \mathbb{E}_{\mathcal{B}_n}\left[1_{n<\tau} X_{n+1} + 1_{n\geq\tau} Y_{n+1}\right] \\ &= 1_{n<\tau}\mathbb{E}_{\mathcal{B}_n}\left[X_{n+1}\right] + 1_{n\geq\tau}\mathbb{E}_{\mathcal{B}_n}\left[Y_{n+1}\right] \\ &\leq 1_{n<\tau} X_n + 1_{n\geq\tau} Y_n = Z_n.\end{aligned}$$

■

### 20.7.1 Maximal Inequalities

**Theorem 20.63 (Supermartingale maximal inequality).** *Let $X$ be a positive supermartingale (in the extended sense) and $a \in \mathcal{B}_0$ with $a \geq 0$, then*

$$aP\left[\sup_n X_n \geq a|\mathcal{B}_0\right] \leq a \wedge X_0 \tag{20.49}$$

*and moreover*

$$P\left[\sup_n X_n = \infty|\mathcal{B}_0\right] = 0 \text{ on } \{X_0 < \infty\}. \tag{20.50}$$

*In particular if $X_0 < \infty$ a.s. then $\sup_n X_n < \infty$ a.s.*

**Proof.** Simply apply Corollary 20.45 with $Y_n = ((2a) \wedge X_n) \cdot 1_A$ where $A \in \mathcal{B}_0$ to find

$$a\mathbb{E}\left(P\left[\sup_n X_n \geq a|\mathcal{B}_0\right] : A\right) = aP\left[\sup_n X_n \geq a : A\right] \leq \mathbb{E}\left[a \wedge X_0 : A\right].$$

Since this holds for all $A \in \mathcal{B}_0$, Eq. (20.49) follows.

**Second Proof.** Let $\tau := \inf\{n : X_n \geq a\}$ which is a stopping time since,

$$\{\tau \leq n\} = \{X_n \geq a\} \in \mathcal{B}_n \text{ for all } n.$$

Since $X_\tau \geq a$ on $\{\tau < \infty\}$ and $Y_n := a$ is a supermartingale, it follows by the switching Lemma 20.62 that

$$Z_n := 1_{n<\tau} X_n + a 1_{n\geq\tau}$$

is a supermartingale (in the extended sense). In particular it follows

$$aP\left(\tau \leq n|\mathcal{B}_0\right) = \mathbb{E}_{\mathcal{B}_0}\left[a 1_{n\geq\tau}\right] \leq \mathbb{E}_{\mathcal{B}_0} Z_n \leq Z_0,$$

and

$$Z_0 = 1_{0<\tau}\, X_0 + a 1_{\tau=0} = 1_{X_0<a} X_0 + 1_{X_0\geq a} a = a \wedge X_0.$$

Therefore, using the cMCT,

$$\begin{aligned}aP\left[\sup_n X_n \geq a|\mathcal{B}_0\right] &= aP\left[\tau < \infty|\mathcal{B}_0\right] = \lim_{n\to\infty} aP\left(\tau \leq n|\mathcal{B}_0\right) \\ &\leq Z_0 = a \wedge X_0\end{aligned}$$

which proves Eq. (20.49).

For the last assertion, take $a > 0$ to be constant in Eq. (20.49) and then use the cDCT to let $a \uparrow \infty$ to conclude

$$P\left[\sup_n X_n = \infty|\mathcal{B}_0\right] = \lim_{a\uparrow\infty} P\left[\sup_n X_n \geq a|\mathcal{B}_0\right] \leq \lim_{a\uparrow\infty} 1 \wedge \frac{X_0}{a} = 1_{X_0=\infty}.$$

Multiplying this equation by $1_{X_0<\infty}$ and then taking expectations implies

$$\mathbb{E}\left[1_{\sup_n X_n=\infty} 1_{X_0<\infty}\right] = \mathbb{E}\left[1_{X_0=\infty} 1_{X_0<\infty}\right] = 0$$

which implies $1_{\sup_n X_n=\infty} 1_{X_0<\infty} = 0$ a.s., i.e. $\sup_n X_n < \infty$ a.s. on $\{X_0 < \infty\}$.

■

### 20.7.2 The upcrossing inequality and convergence result

**Theorem 20.64 (Dubin's Upcrossing Inequality).** *Suppose $X = \{X_n\}_{n=0}^\infty$ is a positive supermartingale and $0 < a < b < \infty$. Then*

$$P\left(U_\infty^X(a,b) \geq k|\mathcal{B}_0\right) \leq \left(\frac{a}{b}\right)^k \left(1 \wedge \frac{X_0}{a}\right), \text{ for } k \geq 1 \tag{20.51}$$

*and $U_\infty(a,b) < \infty$ a.s. and in fact*

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \leq \frac{1}{b/a - 1} = \frac{a}{b-a} < \infty.$$

**Proof.** Since

$$U_N^X(a,b) = U_N^{X/a}(1, b/a),$$

it suffices to consider the case where $a = 1$ and $b > 1$. Let $\tau_n$ be the stopping times defined in Eq. (20.41) with $a = 1$ and $b > 1$, i.e.

$$\begin{aligned}\tau_0 = 0,\ & \tau_1 = \inf\{n \geq \tau_0 : X_n \leq 1\} \\ \tau_2 = \inf\{n \geq \tau_1 : X_n \geq b\},\ & \tau_3 := \inf\{n \geq \tau_2 : X_n \leq 1\} \\ & \vdots \\ \tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \geq b\},\ & \tau_{2k+1} := \inf\{n \geq \tau_{2k} : X_n \leq 1\}, \\ & \vdots\end{aligned}$$

see Figure 20.2.

Let $k \geq 1$ and use the switching Lemma 20.62 repeatedly to define a new positive supermatingale $Y_n = Y_n^{(k)}$ (see Exercise 20.13 below) as follows,

$$\begin{aligned} Y_n^{(k)} &= 1_{n<\tau_1} + 1_{\tau_1 \leq n < \tau_2} X_n \\ &+ b 1_{\tau_2 \leq n < \tau_3} + b X_n 1_{\tau_3 \leq n < \tau_4} \\ &+ b^2 1_{\tau_4 \leq n < \tau_5} + b^2 X_n 1_{\tau_5 \leq n < \tau_6} \\ &\vdots \\ &+ b^{k-1} 1_{\tau_{2k-2} \leq n < \tau_{2k-1}} + b^{k-1} X_n 1_{\tau_{2k-1} \leq n < \tau_{2k}} \\ &+ b^k 1_{\tau_{2k} \leq n}. \end{aligned} \tag{20.52}$$

Since $\mathbb{E}[Y_n|\mathcal{B}_0] \leq Y_0$ a.s., $Y_n \geq b^k 1_{\tau_{2k} \leq n}$, and

$$Y_0 = 1_{0<\tau_1} + 1_{\tau_1 = 0} X_0 = 1_{X_0 > 1} + 1_{X_0 \leq 1} X_0 = 1 \wedge X_0,$$

we may infer that

$$b^k P(\tau_{2k} \leq n | \mathcal{B}_0) = \mathbb{E}\left[b^k 1_{\tau_{2k} \leq n} | \mathcal{B}_0\right] \leq \mathbb{E}[Y_n | \mathcal{B}_0] \leq 1 \wedge X_0 \text{ a.s.}$$

Using *cMCT*, we may now let $n \to \infty$ to conclude

$$P\left(U^X(1,b) \geq k | \mathcal{B}_0\right) \leq P(\tau_{2k} < \infty | \mathcal{B}_0) \leq \frac{1}{b^k}(1 \wedge X_0) \text{ a.s.}$$

which is Eq. (20.51). Using cDCT, we may let $k \uparrow \infty$ in this equation to discover $P\left(U_\infty^X(1,b) = \infty | \mathcal{B}_0\right) = 0$ a.s. and in particular, $U_\infty^X(1,b) < \infty$ a.s. In fact we have

$$\begin{aligned} \mathbb{E}\left[U_\infty^X(1,b)\right] &= \sum_{k=1}^\infty P\left(U_\infty^X(1,b) \geq k\right) \leq \sum_{k=1}^\infty \mathbb{E}\left[\frac{1}{b^k}(1 \wedge X_0)\right] \\ &= \frac{1}{b} \frac{1}{1 - 1/b} \mathbb{E}[(1 \wedge X_0)] \leq \frac{1}{b-1} < \infty. \end{aligned}$$

■

**Exercise 20.13.** In this exercise you are asked to fill in the details showing $Y_n$ in Eq. (20.52) is still a supermartingale. To do this, define $Y_n^{(k)}$ via Eq. (20.52) and then show (making use of the switching Lemma 20.62 twice) $Y_n^{(k+1)}$ is a supermartingale under the assumption that $Y_n^{(k)}$ is a supermartingale. Finish off the induction argument by observing that the constant process, $U_n := 1$ and $V_n = 0$ are supermartingales such that $U_{\tau_1} = 1 \geq 0 = V_{\tau_1}$ on $\{\tau_1 < \infty\}$, and therefore by the switching Lemma 20.62,

$$Y_n^{(1)} = 1_{0 \leq n < \tau_1} U_n + 1_{\tau_1 \leq n} V_n = 1_{0 \leq n < \tau_1}$$

is also a supermartingale.



**Corollary 20.65 (Positive Supermartingale convergence).** *Suppose $X = \{X_n\}_{n=0}^\infty$ is a positive supermartingale (possibly in the extended sense), then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and we have*

$$\mathbb{E}[X_\infty | \mathcal{B}_n] \leq X_n \text{ for all } n \in \bar{\mathbb{N}}. \tag{20.53}$$

*In particular,*

$$\mathbb{E}X_\infty \leq \mathbb{E}X_n \leq \mathbb{E}X_0 \text{ for all } n < \infty. \tag{20.54}$$

**Proof.** The set,

$$\Omega_0 := \cap \left\{U_\infty^X(a,b) < \infty : a, b \in \mathbb{Q} \text{ with } a < b\right\},$$

has full measure ($P(\Omega_0) = 1$) by Dubin's upcrossing inequality in Theorem 20.64. So by Lemma 20.55, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists[9] in $[0,\infty]$. For definiteness, let $X_\infty = 0$ on $\Omega_0^c$. Equation (20.53) is now a consequence of cFatou;

$$\mathbb{E}[X_\infty | \mathcal{B}_n] = \mathbb{E}\left[\lim_{m\to\infty} X_m | \mathcal{B}_n\right] \leq \liminf_{m\to\infty} \mathbb{E}[X_m | \mathcal{B}_n] \leq \liminf_{m\to\infty} X_n = X_n \text{ a.s.}$$

The supermartingale property guarantees that $\mathbb{E}X_n \leq \mathbb{E}X_0$ for all $n < \infty$ while taking expectations of Eq. (20.53) implies $\mathbb{E}X_\infty \leq \mathbb{E}X_n$. ■

**Theorem 20.66 (Optional sampling II – Positive supermartingales).** *Suppose that $X = \{X_n\}_{n=0}^\infty$ is a positive supermartingale, $X_\infty := \lim_{n\to\infty} X_n$ (which exists a.s. by Corollary 20.65), and $\sigma$ and $\tau$ are* ***arbitrary*** *stopping times. Then $X_n^\tau := X_{\tau\wedge n}$ is a positive $\{\mathcal{B}_n\}_{n=0}^\infty$ – super martingale, $X_\infty^\tau = \lim_{n\to\infty} X_{\tau\wedge n}^\tau$, and*

$$\mathbb{E}[X_\tau | \mathcal{B}_\sigma] \leq X_{\sigma\wedge\tau} \text{ a.s.} \tag{20.55}$$

*Moreover, if $\mathbb{E}X_0 < \infty$, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_\infty^\tau] < \infty$.*

**Proof.** We already know that $X^\tau$ is a positive supermatingale by optional stopping Theorem 20.39. Hence an application of Corollary 20.65 implies that $\lim_{n\to\infty} X_n^\tau = \lim_{n\to\infty} X_{\tau\wedge n}$ is convergent and

$$\mathbb{E}\left[\lim_{n\to\infty} X_n^\tau | \mathcal{B}_m\right] \leq X_m^\tau = X_{\tau\wedge m} \text{ for all } m < \infty. \tag{20.56}$$

On the set $\{\tau < \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ and on the set $\{\tau = \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = \lim_{n\to\infty} X_n = X_\infty = X_\tau$ a.s. Therefore it follows that $\lim_{n\to\infty} X_n^\tau = X_\tau$ and Eq. (20.56) may be expressed as

$$\mathbb{E}[X_\tau | \mathcal{B}_m] \leq X_{\tau\wedge m} \text{ for all } m < \infty. \tag{20.57}$$

An application of Lemma 20.31 now implies

$$\mathbb{E}[X_\tau | \mathcal{B}_\sigma] = \sum_{m\leq\infty} 1_{\sigma=m} \mathbb{E}[X_\tau | \mathcal{B}_m] \leq \sum_{m\leq\infty} 1_{\sigma=m} X_{\tau\wedge m} = X_{\tau\wedge\sigma} \text{ a.s.}$$

■

[9] If $\mathbb{E}X_0 < \infty$, this may also be deduced by applying Corollary 20.56 to $\{-X_n\}_{n=0}^\infty$.

## 20.8 Martingale Closure and Regularity Results

We are now going to give a couple of theorems which have already been alluded to in Exercises 15.6, 16.9, and 16.10.

**Theorem 20.67.** *Let $M := \{M_n\}_{n=0}^\infty$ be an $L^1$ – bounded martingale, i.e. $C := \sup_n \mathbb{E}|M_n| < \infty$ and let $M_\infty := \lim_{n\to\infty} M_n$ which exists a.s. and satisfies, $\mathbb{E}|M_\infty| < \infty$ by Corollary 20.56. Then the following are equivalent;*

1. *There exists $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{B}_n]$ for all $n$.*
2. *$\{M_n\}_{n=0}^\infty$ is uniformly integrable.*
3. *$M_n \to M_\infty$ in $L^1(\Omega, \mathcal{B}, P)$.*

*Moreover, if any of the above equivalent conditions hold we may take $X = M_\infty$, i.e. $M_n = \mathbb{E}[M_\infty|\mathcal{B}_n]$.*

**Proof.** 1. $\implies$ 2. This was already proved in Proposition 20.8.

2. $\implies$ 3. The knowledge that $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. along with the assumed uniform integrability implies $L^1$ – convergence by Vitali convergence Theorem 14.51.

3. $\implies$ 1. If $M_n \to M_\infty$ in $L^1(\Omega, \mathcal{B}, P)$, then by the martingale property and the $L^1(P)$ – continuity of conditional expectation we find,

$$M_n = \mathbb{E}[M_m|\mathcal{B}_n] \to \mathbb{E}[M_\infty|\mathcal{B}_n] \text{ as } m \to \infty,$$

and thus, $M_n = \mathbb{E}[M_\infty|\mathcal{B}_n]$ a.s. ■

**Definition 20.68.** *A martingale satisfying any and all of the equivalent statements in Theorem 20.67 is said to be* ***regular.***

**Exercise 20.14 (Rademacher's theorem).** Let $\Omega := (0,1]$, $\mathcal{B} := \mathcal{B}_{(0,1]}$, $P = m$ be Lebesgue measure, and $f \in L^1(P)$. To each partition $\Pi := \{0 = x_0 < x_1 < x_2 < \cdots < x_n = 1\}$ of $(0,1]$ we let $\mathcal{B}_\Pi := \sigma(J_i := (x_{i-1}, x_i] : 1 \le i \le n)$.

1. Show $\mathbb{E}[f|\mathcal{B}_\Pi](x) = \sum_{i=1}^n \frac{1}{x_i - x_{i-1}} \left[\int_{x_{i-1}}^{x_i} f(s)\,ds\right] \cdot 1_{(x_{i-1},x_i]}(x)$ for a.e. $x \in \Omega$.
2. For $f \in C([0,1],\mathbb{R})$, let
$$f_\Pi(x) := \sum_{i=1}^n \frac{\Delta_i f}{\Delta_i} 1_{J_i}(x) \tag{20.58}$$
where $\Delta_i f := f(x_i) - f(x_{i-1})$ and $\Delta_i := x_i - x_{i-1}$. Show if $\Pi'$ is another partition of $\Omega$ which refines $\Pi$, i.e. $\Pi \subset \Pi'$, then
$$f_\Pi = \mathbb{E}[f_{\Pi'}|\mathcal{B}_\Pi] \text{ a.s.}$$
3. Show for any $a, b \in \Pi$ with $a < b$ that
$$\frac{f(b) - f(a)}{b - a} = \frac{1}{b-a}\int_a^b f_\Pi(x)\,dx. \tag{20.59}$$
**Hint:** consider the partition $\Pi_0 := \{0 < a < b < 1\}$.
Now let $\mathcal{B}_n := \mathcal{B}_{\Pi_n}$ and where $\Pi_n := \left\{\frac{k}{2^n}\right\}_{k=0}^{2^n}$ an observe your have now shown $g_n := f_{\Pi_n}$ is a martingale.
4. Let us now further suppose that $|f(y) - f(x)| \le K|y - x|$ for all $x, y \in [0,1]$, i.e. $f$ is Lipschitz. From Eq. (20.58) it follows that $|g_n| := |f_{\Pi_n}| \le K$ so that $\{g_n\}_{n=1}^\infty$ is a bounded martingale. Use this along with Eq. (20.59) and Theorem 20.67 to conclude there exists $g \in L^\infty(P)$ such that
$$f(b) - f(a) = \int_a^b g(x)\,dx \text{ for all } 0 \le a < b \le 1.$$
[You may be interested to know that under these hypothesis, $f'(x)$ exists a.e. and $g(x) = f'(x)$ a.e.. Thus this a version of the fundamental theorem of calculus.]

**Theorem 20.69.** *Suppose $1 < p < \infty$ and $M := \{M_n\}_{n=0}^\infty$ is an $L^p$ – bounded martingale. Then $M_n \to M_\infty$ almost surely and in $L^p$. In particular, $\{M_n\}$ is a regular martingale.*

**Proof.** The almost sure convergence follows from Corollary 20.56. So, because of Theorem 14.51, to finish the proof it suffices to show $\{|M_n|^p\}_{n=0}^\infty$ is uniformly integrable. But by Doob's inequality, Corollary 20.48, and the MCT, we find

$$\mathbb{E}\left[\sup_k |M_k|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_k \mathbb{E}[|M_k|^p] < \infty.$$

As $|M_n|^p \le \sup_k |M_k|^p \in L^1(P)$ for all $n \in \mathbb{N}$, it follows by Example 14.39 and Exercise 14.5 that $\{|M_n|^p\}_{n=0}^\infty$ is uniformly integrable. ■

**Theorem 20.70 (Optional sampling III – regular martingales).** *Suppose that $M = \{M_n\}_{n=0}^\infty$ is a regular martingale, $\sigma$ and $\tau$ are* ***arbitrary*** *stopping times. Define $M_\infty := \lim_{n\to\infty} M_n$ which exists a.s.. Then $M_\infty \in L^1(P)$,*

$$M_\tau = \mathbb{E}[M_\infty|\mathcal{B}_\tau], \quad \mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty \tag{20.60}$$

*and*

$$\mathbb{E}[M_\tau|\mathcal{B}_\sigma] = M_{\sigma\wedge\tau} \text{ a.s.} \tag{20.61}$$

**Proof.** By Theorem 20.67, $M_\infty \in L^1(\Omega, \mathcal{B}, P)$ and $M_n := \mathbb{E}_{\mathcal{B}_n} M_\infty$ a.s. for all $n \le \infty$. By Lemma 20.31,

$$\mathbb{E}_{\mathcal{B}_\tau} M_\infty = \sum_{n\le\infty} 1_{\tau=n} \mathbb{E}_{\mathcal{B}_n} M_\infty = \sum_{n\le\infty} 1_{\tau=n} M_n = M_\tau.$$

Hence we have $|M_\tau| = |\mathbb{E}_{\mathcal{B}_\tau} M_\infty| \le \mathbb{E}_{\mathcal{B}_\tau} |M_\infty|$ a.s. and $\mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty$. An application of Theorem 20.32 now concludes the proof;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \mathbb{E}_{\mathcal{B}_\sigma} \mathbb{E}_{\mathcal{B}_\tau} M_\infty = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}} M_\infty = M_{\sigma\wedge\tau}.$$

■

**Definition 20.71.** *Let $M = \{M_n\}_{n=0}^\infty$ be a martingale. We say that $\tau$ is a* ***regular stopping time for*** *$M$ if $M^\tau$ is a regular martingale.*

*Example 20.72.* Every bounded martingale is regular. More generally if $\tau$ is a stopping time such that $M^\tau$ is bounded, then $\tau$ is a regular stopping time for $M$.

*Remark 20.73.* If $\tau$ is regular for $M$, then $\lim_{n\to\infty} M_n^\tau := M_\infty^\tau$ exists a.s. and in $L^1(P)$ and hence

$$M_n^\tau = \mathbb{E}[M_\infty^\tau|\mathcal{B}_n] \text{ for all } n \le \infty \tag{20.62}$$

and

$$\lim_{n\to\infty} M_n = M_\infty^\tau \text{ a.s. on } \{\tau = \infty\}. \tag{20.63}$$

**Definition 20.74.** *If $\tau$ is regular stopping time for $M$, then we define $M_\tau$ as,*

$$M_\tau := M_\infty^\tau = \lim_{n\to\infty} M_{n\wedge\tau} = \begin{cases} M_\tau & \text{if } \tau < \infty \\ M_\infty & \text{if } \tau = \infty \end{cases}. \tag{20.64}$$

From Eq. (20.62) and the definition of $M_\tau$ we have

$$M_{\tau\wedge n} = M_n^\tau = \mathbb{E}[M_\infty^\tau|\mathcal{B}_n] = \mathbb{E}[M_\tau|\mathcal{B}_n] \text{ for all } n \le \infty \tag{20.65}$$

and further note that

$$\mathbb{E}|M_\tau| = \lim_{n\to\infty} \mathbb{E}|M_n^\tau| \le \sup_n \mathbb{E}|M_n^\tau| < \infty.$$

**Theorem 20.75.** *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale and $\sigma, \tau$, are stopping times such that $\tau$ is a regular stopping time for $M$. Then*

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = M_{\tau\wedge\sigma}. \tag{20.66}$$

*If we further assume that $\sigma \le \tau$ a.s. then*

$$M_n^\sigma = \mathbb{E}_{\mathcal{B}_n}[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau] \tag{20.67}$$

*and $\sigma$ is also regular for $M$.*



**Proof.** By assumption, $M_\tau = \lim_{n\to\infty} M_{n\wedge\tau}$ exists almost surely and in $L^1(P)$ and $M_n^\tau = \mathbb{E}[M_\tau|\mathcal{B}_n]$ for $n \le \infty$.

1. Equation (20.66) is a consequence of;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \sum_{n\le\infty} 1_{\sigma=n} \mathbb{E}_{\mathcal{B}_n} M_\tau = \sum_{n\le\infty} 1_{\sigma=n} M_n^\tau = M_{\sigma\wedge\tau},$$

wherein Lemma 20.31 was used for the first equality.

2. Applying $\mathbb{E}_{\mathcal{B}_\sigma}$ to Eq. (20.65) using the optional sampling Theorem 20.40 and the tower property of conditional expectation (see Theorem 20.32) shows

$$M_n^\sigma = M_{\sigma\wedge n} = \mathbb{E}_{\mathcal{B}_\sigma} M_{\tau\wedge n} = \mathbb{E}_{\mathcal{B}_\sigma} \mathbb{E}_{\mathcal{B}_n} M_\tau = \mathbb{E}_{\mathcal{B}_n}[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau].$$

The regularity of $M^\sigma$ now follows by item 1. of Theorem 20.67.

■

**Proposition 20.76.** *Suppose that $M$ is a martingale and $\tau$ is a stopping time. Then $\tau$ is regular for $M$ iff;*

1. *$\mathbb{E}[|M_\tau| : \tau < \infty] < \infty$ and*
2. *$\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is a uniformly integrable sequence of random variables.*

*Moreover, condition 1. is automatically satisfied if $M$ is $L^1$ – bounded, i.e. if $C := \sup_n \mathbb{E}|M_n| < \infty$.*

**Proof.** ($\Longrightarrow$) If $\tau$ is regular for $M$, $M_\tau \in L^1(P)$ and $M_n^\tau = \mathbb{E}_{\mathcal{B}_n} M_\tau$ so that $M_n = \mathbb{E}_{\mathcal{B}_n} M_\tau$ a.s. on $\{n \le \tau\}$. In particular it follows that

$$\mathbb{E}[|M_\tau| : \tau < \infty] \le \mathbb{E}|M_\tau| < \infty$$

and

$$|M_n 1_{n<\tau}| = |\mathbb{E}_{\mathcal{B}_n} M_\tau 1_{n<\tau}| \le \mathbb{E}_{\mathcal{B}_n} |M_\tau| \text{ a.s.}$$

from which it follows that $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is uniformly integrable.

($\Longleftarrow$) Our goal is to show $\{M_n^\tau\}_{n=0}^\infty$ is uniformly integrable. We begin with the identity;

$$\begin{aligned}\mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a] =& \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] \\ &+ \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau].\end{aligned}$$

Since

$$\begin{aligned}\mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] &= \mathbb{E}[|M_\tau| : |M_\tau| \ge a,\ \tau \le n] \\ &\le \mathbb{E}[|M_\tau 1_{\tau<\infty}| : |M_\tau 1_{\tau<\infty}| \ge a],\end{aligned}$$

if follows (by assumption 1. that $\mathbb{E}[|M_\tau 1_{\tau<\infty}|] < \infty$) that

$$\lim_{a\to\infty}\sup_n \mathbb{E}\left[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n\right] = 0.$$

Moreover for any $a > 0$,

$$\sup_n \mathbb{E}\left[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau\right] = \sup_n \mathbb{E}\left[|M_n^\tau 1_{n<\tau}| : |M_n^\tau 1_{n<\tau}| \ge a\right]$$

and the latter term goes to zero as $a \to \infty$ by assumption 2. Hence we have shown,

$$\lim_{a\to\infty}\sup_n \mathbb{E}\left[|M_n^\tau| : |M_n^\tau| \ge a\right] = 0$$

as desired.

Now to prove the last assertion. If $C := \sup_n \mathbb{E}|M_n| < \infty$, the (by Corollary 20.56) $M_\infty := \lim_{n\to\infty} M_n$ a.s. and $\mathbb{E}|M_\infty| < \infty$. Therefore,

$$\mathbb{E}\left[|M_\tau| : \tau < \infty\right] \le \mathbb{E}|M_\tau| = \mathbb{E}\left[\lim_{n\to\infty}|M_{\tau\wedge n}|\right]$$
$$\le \liminf_{n\to\infty}\mathbb{E}|M_{\tau\wedge n}| \le \liminf_{n\to\infty}\mathbb{E}|M_n| < \infty$$

wherein we have used Fatou's lemma, the optional sampling theorem to conclude $M_{\tau\wedge n} = \mathbb{E}_{\mathcal{B}_{\tau\wedge n}} M_n$, cJensen to conclude $|M_{\tau\wedge n}| \le \mathbb{E}_{\mathcal{B}_{\tau\wedge n}}|M_n|$, and the tower property of conditional expectation to conclude $\mathbb{E}|M_{\tau\wedge n}| \le \mathbb{E}|M_n|$. ■

**Corollary 20.77.** *Suppose that $M$ is an $L^1$ – bounded martingale and $J \in \mathcal{B}_\mathbb{R}$ is a bounded set, then $\tau = \inf\{n : M_n \notin J\}$ is a regular stopping time for $M$.*

**Proof.** According to Proposition 20.76, it suffices to show $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is a uniformly integrable sequence of random variables. However, if we choose $A < \infty$ such that $J \subset [-A, A]$, since $M_n 1_{n<\tau} \in J$ we have $|M_n 1_{n<\tau}| \le A$ which is sufficient to complete the proof. ■

## 20.9 Backwards (Reverse) Submartingales

In this section we will consider submartingales indexed by $\mathbb{Z}_- := \{\dots, -n, -n+1, \dots, -2, -1, 0\}$. So again we assume that we have an increasing filtration, $\{\mathcal{B}_n : n \le 0\}$, i.e. $\dots \subset \mathcal{B}_{-2} \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \subset \mathcal{B}$. As usual, we say an adapted process $\{X_n\}_{n\le 0}$ is a submartingale (martingale) provided $\mathbb{E}[X_n - X_m|\mathcal{B}_n] \ge 0\ (= 0)$ for all $m \le n$. Observe that $\mathbb{E}X_n \ge \mathbb{E}X_m$ for $n \ge m$, so that $\mathbb{E}X_{-n}$ decreases as $n$ increases. Also observe that $\left(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0\right)$ is a "finite string" submartingale relative to the filtration, $\mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \dots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0$.

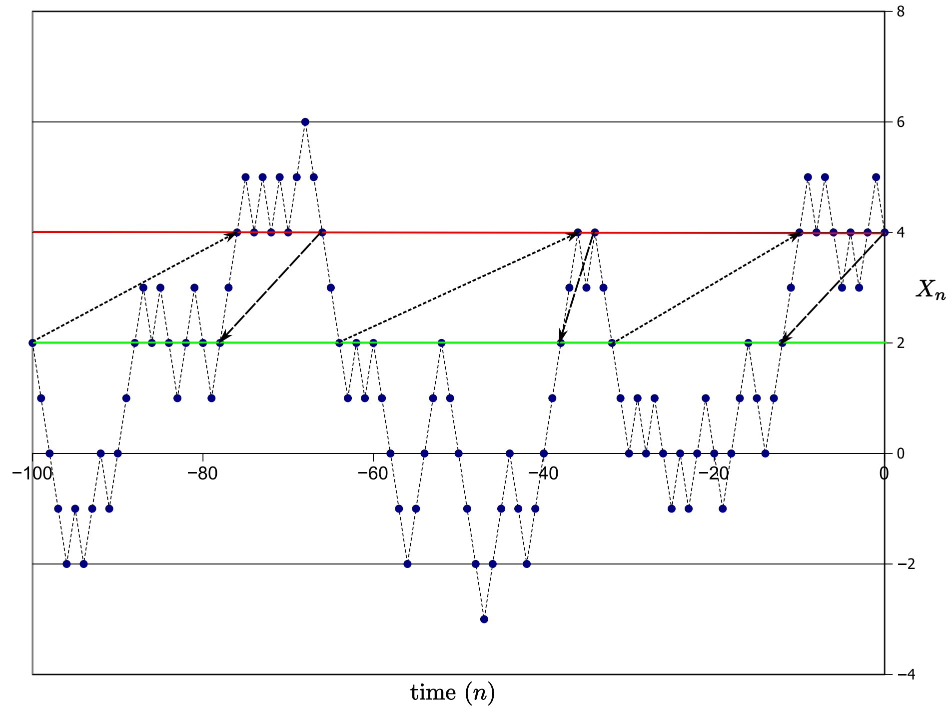



**Fig. 20.5.** A sample path of a backwards martingale on $[-100, 0]$ indicating the down crossings of $X_0$, $X_{-1}, \dots, X_{-100}$ and the upcrossings of $X_{-100}, X_{-99}, \dots, X_0$. The total number of each is the same.

*Remark 20.78.* For all $n \in \mathbb{Z}_-$ we have $X_n \le \mathbb{E}[X_0|\mathcal{B}_n]$ and therefore by the same argument in Lemma 20.52 we have $X_n^+ \le \mathbb{E}\left[X_0^+|\mathcal{B}_n\right]$. In particular we conclude that $\{X_n^+\}_{n\in\mathbb{Z}_-}$ is uniformly integrable. In general, backwards submartingales are even better behaved than forward submartingales. To see this even more clearly notice if $\{M_n\}_{n\le 0}$ is a backwards martingale, then $\mathbb{E}[M_n|\mathcal{B}_m] = M_{m\wedge n}$ for all $m, n \le 0$. Taking $n = 0$ in this equation implies that $M_m = \mathbb{E}[M_0|\mathcal{B}_m]$ and so the only backwards martingales are of the form $M_m = \mathbb{E}[M_0|\mathcal{B}_m]$ for some $M_0 \in L^1(P)$. We have seen in Example 20.7 that this need not be the case for forward martingales.

**Theorem 20.79 (Backwards (or reverse) submartingale convergence).** *Let $\{\mathcal{B}_n : n \le 0\}$ be a reverse filtration, $\{X_n\}_{n\le 0}$ is a backwards submartingale. Then $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. in $\{-\infty\} \cup \mathbb{R}$ and $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$. If we further assume that*[10]

[10] Since $X_n^- = X_n^+ - X_n$, it follows from Remark 20.78 that

$$C := \lim_{n\to-\infty} \mathbb{E}X_n = \inf_{n\leq 0} \mathbb{E}X_n > -\infty, \tag{20.68}$$

*then 1)* $X_n = M_n + A_n$ *where* $\{M_n\}_{-\infty<n\leq 0}$ *is a martingale,* $\{A_n\}_{-\infty<n\leq 0}$ *is a predictable process such that* $A_{-\infty} = \lim_{n\to-\infty} A_n = 0$*, 2)* $\{X_n\}_{n\leq 0}$ *is uniformly integrability, 3)* $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$ *, and 4)* $\lim_{n\to-\infty} \mathbb{E}|X_n - X_{-\infty}| = 0.$

**Proof.** The number of downcrossings of $(X_0, X_{-1}, \dots, X_{-(n-1)}, X_{-n})$ across $[a,b]$, (denoted by $D_n(a,b)$) is equal to the number of upcrossings, $(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0)$ across $[a,b]$, see Figure 20.5. Since $(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0)$ is a $\mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \dots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0$ submartingale, we may apply Doob's upcrossing inequality (Theorem 20.53) to find;

$$\begin{aligned}(b-a)\,\mathbb{E}[D_n(a,b)] &\leq \mathbb{E}(X_0 - a)_+ - \mathbb{E}(X_{-n} - a)_+ \\ &\leq \mathbb{E}(X_0 - a)_+ < \infty. \end{aligned}\tag{20.69}$$

Letting $D_\infty(a,b) :=\uparrow \lim_{n\to\infty} D_n(a,b)$ be the total number of downcrossing of $(X_0, X_{-1}, \dots, X_{-n}, \dots)$, using the MCT to pass to the limit in Eq. (20.69), we have

$$(b-a)\,\mathbb{E}[D_\infty(a,b)] \leq \mathbb{E}(X_0 - a)_+ < \infty.$$

In particular it follows that $D_\infty(a,b) < \infty$ a.s. for all $a < b$.

As in the proof of Corollary 20.56 (making use of the obvious downcrossing analogue of Lemma 20.55), it follows that $X_{-\infty} := \lim_{n\to-\infty} X_n$ exists in $\bar{\mathbb{R}}$ a.s. At the end of the proof, we will show that $X_{-\infty}$ takes values in $\{-\infty\} \cup \mathbb{R}$ almost surely, i.e. $X_{-\infty} < \infty$ a.s.

Now suppose that $C > -\infty$. We begin by computing the Doob decomposition of $X_n$ as $X_n = M_n + A_n$ with $A_n$ being predictable, increasing and satisfying, $A_{-\infty} = \lim_{n\to-\infty} A_n = 0$. If such an $A$ is to exist, following Lemma 20.17, we should define

$$A_n = \sum_{k\leq n} \mathbb{E}[\Delta_k X|\mathcal{B}_{k-1}] \text{ where } \Delta_k X := X_k - X_{k-1}.$$

This is a well defined increasing predictable process since the submartingale property implies $\mathbb{E}[\Delta_k X|\mathcal{B}_{k-1}] \geq 0$. Moreover we have

---

$$\mathbb{E}X_n^- = \mathbb{E}X_n^+ - \mathbb{E}X_n \leq \mathbb{E}X_0^+ - \inf_{n\leq 0} \mathbb{E}X_n = \mathbb{E}X_0^+ - C$$

and

$$\mathbb{E}|X_n| \leq \mathbb{E}X_n^+ + \mathbb{E}X_n^- \leq 2\mathbb{E}X_0^+ - C.$$

Therefore $C > -\infty$ iff $\sup_{n\leq\infty} \mathbb{E}|X_n| < \infty$.

$$\begin{aligned}\mathbb{E}A_0 &= \sum_{k\leq 0} \mathbb{E}[\mathbb{E}[\Delta_k X|\mathcal{B}_{k-1}]] = \sum_{k\leq 0} \mathbb{E}[\Delta_k X] \\ &= \lim_{N\to\infty} (\mathbb{E}X_0 - \mathbb{E}X_{-N}) = \mathbb{E}X_0 - \inf_{n\leq 0} \mathbb{E}X_n = \mathbb{E}X_0 - C < \infty.\end{aligned}$$

As $0 \leq A_n \leq A_n^* = A_0 \in L^1(P)$, it follows that $\{A_n\}_{n\leq 0}$ is uniformly integrable. Moreover if we define $M_n := X_n - A_n$, then

$$\mathbb{E}[\Delta_n M|\mathcal{B}_{n-1}] = \mathbb{E}[\Delta_n X - \Delta_n A|\mathcal{B}_{n-1}] = \mathbb{E}[\Delta_n X|\mathcal{B}_{n-1}] - \Delta_n A = 0 \text{ a.s.}$$

Thus $M$ is a martingale and therefore, $M_n = \mathbb{E}[M_0|\mathcal{B}_n]$ with $M_0 = X_0 - A_0 \in L^1(P)$. An application of Proposition 20.8 implies $\{M_n\}_{n\leq 0}$ is uniformly integrable and hence $X_n = M_n + A_n$ is uniformly integrable as well. (See Remark 20.80 for an alternate proof of the uniform integrability of $X$.) Therefore $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$ and $X_n \to X_{-\infty}$ in $L^1(\Omega, \mathcal{B}, P)$ as $n \to \infty$.

To finish the proof we must show without assuming $C > -\infty$ that $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$ which will then also imply $P(X_{-\infty} = \infty) = 0$. To prove this, notice that $X_{-\infty}^+ = \lim_{n\to-\infty} X_n^+$ and that (by Jensen's inequality) $\{X_n^+\}_{n=1}^\infty$ is a non-negative backwards submartingale. Since $\inf \mathbb{E}X_n^+ \geq 0 > -\infty$, it follows by what we have just proved that $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$. ■

*Remark 20.80 (*Not necessary to read.).* Let us give a direct proof of the fact that $X$ is uniformly integrable if $C > -\infty$. We begin with Jensen's inequality;

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - C = K < \infty, \tag{20.70}$$

which shows that $\{X_n\}_{n=1}^\infty$ is $L^1$ - bounded. For uniform integrability we will use the following identity;

$$\begin{aligned}\mathbb{E}[|X| : |X| \geq \lambda] &= \mathbb{E}[X : X \geq \lambda] - \mathbb{E}[X : X \leq -\lambda] \\ &= \mathbb{E}[X : X \geq \lambda] - (\mathbb{E}X - \mathbb{E}[X : X > -\lambda]) \\ &= \mathbb{E}[X : X \geq \lambda] + \mathbb{E}[X : X > -\lambda] - \mathbb{E}X.\end{aligned}$$

Taking $X = X_n$ and $k \geq n$, we find

$$\begin{aligned}\mathbb{E}[|X_n| : |X_n| \geq \lambda] =& \mathbb{E}[X_n : X_n \geq \lambda] + \mathbb{E}[X_n : X_n > -\lambda] - \mathbb{E}X_n \\ \leq& \mathbb{E}[X_k : X_n \geq \lambda] + \mathbb{E}[X_k : X_n > -\lambda] \\ & - \mathbb{E}X_k + (\mathbb{E}X_k - \mathbb{E}X_n) \\ =& \mathbb{E}[X_k : X_n \geq \lambda] - \mathbb{E}[X_k : X_n \leq -\lambda] + (\mathbb{E}X_k - \mathbb{E}X_n) \\ =& \mathbb{E}[|X_k| : |X_n| \geq \lambda] + (\mathbb{E}X_k - \mathbb{E}X_n).\end{aligned}$$

Given $\varepsilon > 0$ we may choose $k = k_\varepsilon < 0$ such that if $n \leq k$, $0 \leq \mathbb{E}X_k - \mathbb{E}X_n \leq \varepsilon$ and hence

$$\lim_{\lambda\uparrow\infty}\sup_{n\le k}\mathbb{E}\left[|X_n| : |X_n|\ge\lambda\right]\le\limsup_{\lambda\uparrow\infty}\mathbb{E}\left[|X_k| : |X_n|\ge\lambda\right]+\varepsilon\le\varepsilon$$

wherein we have used Eq. (20.70), Chebyschev's inequality to conclude $P(|X_n|\ge\lambda)\le K/\lambda$ and then the uniform integrability of the singleton set, $\{|X_k|\}\subset L^1(\Omega,\mathcal{B},P)$. From this it now easily follows that $\{X_n\}_{n\le 0}$ is a uniformly integrable.

**Corollary 20.81.** *Suppose $1\le p<\infty$ and $X_n=M_n$ in Theorem 20.79, where $M_n$ is an $L^p$ – bounded martingale on $-\mathbb{N}\cup\{0\}$. Then $M_{-\infty}:=\lim_{n\to\infty}M_n$ exists a.s. and in $L^p(P)$. Moreover $M_{-\infty}=\mathbb{E}[M_0|\mathcal{B}_{-\infty}]$, where $\mathcal{B}_{-\infty}=\cap_{n\le 0}\mathcal{B}_n$.*

**Proof.** Since $M_n=\mathbb{E}[M_0|\mathcal{B}_n]$ for all $n$, it follows by cJensen that $|M_n|^p\le\mathbb{E}[|M_0|^p|\mathcal{B}_n]$ for all $n$. By Proposition 20.8, $\{\mathbb{E}[|M_0|^p|\mathcal{B}_n]\}_{n\le 0}$ is uniformly integrable and so is $\{|M_n|^p\}_{n\le 0}$. By Theorem 20.79, $M_n\to M_{-\infty}$ a.s.. Hence we may now apply Theorem 14.51 to see that $M_n\to M_{-\infty}$ in $L^p(P)$. ■

*Example 20.82 (Kolmogorov's SLLN).* In this example we are going to give another proof of the strong law of large numbers in Theorem 18.10, also see Theorem 22.31 below for a third proof. Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $\mathbb{E}X_n=0$ and let $S_0=0$, $S_n:=X_1+\dots+X_n$ and $\mathcal{B}_{-n}=\sigma(S_n,S_{n+1},S_{n+2},\dots)$ so that $S_n$ is $\mathcal{B}_{-n}$ measurable for all $n\ge 0$. [The first three items below give a solution to Exercise 16.8.]

1. For any permutation $\sigma$ of the set $\{1,2,\dots,n\}$,
$$(X_1,\dots,X_n,S_n,S_{n+1},S_{n+2},\dots)\overset{d}{=}(X_{\sigma1},\dots,X_{\sigma n},S_n,S_{n+1},S_{n+2},\dots)$$
and in particular
$$(X_j,S_n,S_{n+1},S_{n+2},\dots)\overset{d}{=}(X_1,S_n,S_{n+1},S_{n+2},\dots)\ \text{ for all } j\le n.$$
2. By Exercise 16.7 we may conclude that
$$\mathbb{E}[X_j|\mathcal{B}_{-n}]=\mathbb{E}[X_1|\mathcal{B}_{-n}]\ \text{ a.s. for all } j\le n. \tag{20.71}$$
To see this directly notice that if $\sigma$ is any permutation of $\mathbb{N}$ leaving $\{n+1,n+2,\dots\}$ fixed, then
$$\mathbb{E}[g(X_1,\dots,X_n)\cdot f(S_n,S_{n+1},\dots)]=\mathbb{E}[g(X_{\sigma1},\dots,X_{\sigma n})\cdot f(S_n,S_{n+1},\dots)]$$
for all bounded measurable $f$ and $g$ such that $g(X_1,\dots,X_n)\in L^1(P)$. From this equation it follows that
$$\mathbb{E}[g(X_1,\dots,X_n)|\mathcal{B}_{-n}]=\mathbb{E}[g(X_{\sigma1},\dots,X_{\sigma n})|\mathcal{B}_{-n}]\ \text{ a.s.}$$
and then taking $g(x_1,\dots,x_n)=x_1$ give the desired result. The point is that $(X_1,X_2,\dots)\overset{d}{=}(X_{\sigma1},X_{\sigma2},\dots)$ and
$$S_k^\sigma:=\sum_{\ell=1}^k X_{\sigma\ell}=\sum_{\ell=1}^k X_\ell=S_k\ \text{ for all } k\ge n.$$
This argument generalizes to $\{X_j\}_{j=1}^\infty$ which are exchangeable, i.e. we only need $(X_1,X_2,\dots)\overset{d}{=}(X_{\sigma1},X_{\sigma2},\dots)$ for all permutations $\sigma$ such that $\sigma(k)=k$ for a.a. $k$.
3. Summing Eq. (20.71) over $j=1,2,\dots,n$ gives,
$$S_n=\mathbb{E}[S_n|S_n,S_{n+1},S_{n+2},\dots]=n\mathbb{E}[X_1|S_n,S_{n+1},S_{n+2},\dots]$$
from which it follows that
$$M_{-n}:=\frac{S_n}{n}:=\mathbb{E}[X_1|S_n,S_{n+1},S_{n+2},\dots] \tag{20.72}$$
and hence $\left\{M_{-n}=\frac{1}{n}S_n\right\}$ is a backwards martingale.
4. By Theorem 20.79 we know;
$$\lim_{n\to\infty}\frac{S_n}{n}=\lim_{n\to-\infty}M_{-n}=:M_{-\infty}\ \text{exists a.s. and in } L^1(P).$$
5. Since $M_{-\infty}=\lim_{n\to\infty}\frac{S_n}{n}$ is a $\{\sigma(X_1,\dots,X_n)\}_{n=1}^\infty$ – tail random variable it follows by Corollary 12.54 (basically by Kolmogorov's zero one law of Proposition 12.53) that $\lim_{n\to\infty}\frac{S_n}{n}=c$ a.s. for some constant $c$.
6. Since $\frac{S_n}{n}\to c$ in $L^1(P)$ we may conclude that
$$c=\lim_{n\to\infty}\mathbb{E}\frac{S_n}{n}=\mathbb{E}X_1.$$

Thus we have given another proof of Kolmogorov's strong law of large numbers.

The next Hilbert space exercise could be used as the basis of a proof the $L^2$ and then $L^1$ – convergence of backwards martingales.

**Exercise 20.15.** Suppose that $\{M_n\}_{n=1}^\infty$ is a decreasing sequence of closed subspaces of a Hilbert space, $H$. Let $M_\infty:=\cap_{n=1}^\infty M_n$. Show $\lim_{n\to\infty}P_{M_n}x=P_{M_\infty}x$ for all $x\in H$. [**Hint:** you might make use of Exercise 15.5.]

## 20.10 Some More Martingale Exercises

(The next four problems were taken directly from http://math.nyu.edu/˜sheff/martingalenote.pdf.)

**Exercise 20.16.** Suppose Harriet has 7 dollars. Her plan is to make one dollar bets on fair coin tosses until her wealth reaches either 0 or 50, and then to go home. What is the expected amount of money that Harriet will have when she goes home? What is the probability that she will have 50 when she goes home?

**Exercise 20.17.** Consider a contract that at time $N$ will be worth either 100 or 0. Let $S_n$ be its price at time $0 \le n \le N$. If $S_n$ is a martingale, and $S_0 = 47$, then what is the probability that the contract will be worth 100 at time $N$?

**Exercise 20.18.** Pedro plans to buy the contract in the previous problem at time 0 and sell it the first time $T$ at which the price goes above 55 or below 15. What is the expected value of $S_T$? You may assume that the value, $S_n$, of the contract is bounded – there is only a finite amount of money in the world up to time $N$. Also note, by assumption, $T \le N$.

**Exercise 20.19.** Suppose $S_N$ is with probability one either 100 or 0 and that $S_0 = 50$. Suppose further there is at least a 60% probability that the price will at some point dip to below 40 and then subsequently rise to above 60 before time $N$. Prove that $S_n$ cannot be a martingale. (I don't know if this problem is correct! but if we modify the 40 to a 30 the buy low sell high strategy will show that $\{S_n\}$ is not a martingale.)

**Exercise 20.20.** Let $(M_n)_{n=0}^{\infty}$ be a martingale with $M_0 = 0$ and $E[M_n^2] < \infty$ for all $n$. Show that for all $\lambda > 0$,

$$P\left(\max_{1\le m\le n} M_m \ge \lambda\right) \le \frac{E[M_n^2]}{E[M_n^2] + \lambda^2}.$$

**Hints:** First show that for any $c > 0$ that $\left\{X_n := (M_n + c)^2\right\}_{n=0}^{\infty}$ is a submartingale and then observe,

$$\left\{\max_{1\le m\le n} M_m \ge \lambda\right\} \subset \left\{\max_{1\le m\le n} X_n \ge (\lambda + c)^2\right\}.$$

Now use Doob' Maximal inequality (Proposition 20.43) to estimate the probability of the last set and then choose $c$ so as to optimize the resulting estimate you get for $P\left(\max_{1\le m\le n} M_m \ge \lambda\right)$. (Notice that this result applies to $-M_n$ as well so it also holds that;

$$P\left(\min_{1\le m\le n} M_m \le -\lambda\right) \le \frac{E[M_n^2]}{E[M_n^2] + \lambda^2} \text{ for all } \lambda > 0.$$

**Exercise 20.21.** Let $\{Z_n\}_{n=1}^{\infty}$ be independent random variables, $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$, and $f_n(\lambda) := \mathbb{E}\left[e^{i\lambda Z_n}\right]$. Suppose $\mathbb{E}e^{i\lambda S_n} = \prod_{n=1}^{N} f_n(\lambda)$ converges to a continuous function, $F(\lambda)$, as $N \to \infty$. Show for each $\lambda \in \mathbb{R}$ that

$$P\left(\lim_{n\to\infty} e^{i\lambda S_n} \text{ exists}\right) = 1. \tag{20.73}$$

**Hints:**

1. Show it is enough to find an $\varepsilon > 0$ such that Eq. (20.73) holds for $|\lambda| \le \varepsilon$.
2. Choose $\varepsilon > 0$ such that $|F(\lambda) - 1| < 1/2$ for $|\lambda| \le \varepsilon$. For $|\lambda| \le \varepsilon$, show $M_n(\lambda) := \frac{e^{i\lambda S_n}}{\mathbb{E}e^{i\lambda S_n}}$ is a bounded complex[11] martingale relative to the filtration, $\mathcal{B}_n = \sigma(Z_1, \dots, Z_n)$.

**Lemma 20.83 (Protter [36, See the lemma on p. 22.]).** *Let $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ such that $\left\{e^{iux_n}\right\}_{n=1}^{\infty}$ is convergent for Lebesgue almost every $u \in \mathbb{R}$. Then $\lim_{n\to\infty} x_n$ exists in $\mathbb{R}$.*

**Proof.** Let $U$ be a uniform random variable with values in $[0,1]$. By assumption, for any $t \in \mathbb{R}$, $\lim_{n\to\infty} e^{itUx_n}$ exists a.s. Thus if $n_k$ and $m_k$ are any increasing sequences we have

$$\lim_{k\to\infty} e^{itUx_{n_k}} = \lim_{n\to\infty} e^{itUx_n} = \lim_{k\to\infty} e^{itUx_{m_k}} \text{ a.s.}$$

and therefore,

$$e^{it\left(Ux_{n_k} - Ux_{m_k}\right)} = \frac{e^{itUx_{n_k}}}{e^{itUx_{m_k}}} \to 1 \text{ a.s. as } k \to \infty.$$

Hence by DCT it follows that

$$\mathbb{E}\left[e^{it\left(Ux_{n_k} - Ux_{m_k}\right)}\right] \to 1 \text{ as } k \to \infty$$

and therefore

$$(x_{n_k} - x_{m_k}) \cdot U = Ux_{n_k} - Ux_{m_k} \to 0$$

in distribution and hence in probability. But his can only happen if $(x_{n_k} - x_{m_k}) \to 0$ as $k \to \infty$. As $\{n_k\}$ and $\{m_k\}$ were arbitrary, this suffices to show $\{x_n\}$ is a Cauchy sequence. ∎

**Exercise 20.22 (Continuation of Exercise 20.21 – See Doob [8, Chapter VII.5]).** Let $\{Z_n\}_{n=1}^{\infty}$ be independent random variables. Use Exercise 20.21 an Lemma 20.83 to prove the series, $\sum_{n=1}^{\infty} Z_n$, converges in $\mathbb{R}$ a.s. iff $\prod_{n=1}^{N} f_n(\lambda)$ converges to a continuous function, $F(\lambda)$ as $N \to \infty$. Conclude from this that $\sum_{n=1}^{\infty} Z_n$ is a.s. convergent iff $\sum_{n=1}^{\infty} Z_n$ is convergent in distribution.

[11] Please use the obvious generalization of a martingale for complex valued processes. It will be useful to observe that the real and imaginary parts of a complex martingales are real martingales.

### 20.10.1 More Random Walk Exercises

For the next four exercises, let $\{Z_n\}_{n=1}^\infty$ be a sequence of Bernoulli random variables with $P(Z_n = \pm 1) = \frac{1}{2}$ and let $S_0 = 0$ and $S_n := Z_1 + \dots + Z_n$. Then $S$ becomes a martingale relative to the filtration, $\mathcal{B}_n := \sigma(Z_1, \dots, Z_n)$ with $\mathcal{B}_0 := \{\emptyset, \Omega\}$ – of course $S_n$ is the (fair) simple random walk on $\mathbb{Z}$. For any $a \in \mathbb{Z}$, let

$$\sigma_a := \inf\{n : S_n = a\}.$$

**Exercise 20.23.** For $a < 0 < b$ with $a, b \in \mathbb{Z}$, let $\tau = \sigma_a \wedge \sigma_b$. Explain why $\tau$ is regular for $S$. Use this to show $P(\tau = \infty) = 0$. **Hint:** make use of Remark 20.73 and the fact that $|S_n - S_{n-1}| = |Z_n| = 1$ for all $n$.

**Exercise 20.24.** In this exercise, you are asked to use the central limit Theorem 12.36 to prove again that $P(\tau = \infty) = 0$, Exercise 20.23. **Hints:** Use the central limit theorem to show

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-x^2/2} dx \ge f(0) P(\tau = \infty) \tag{20.74}$$

for all $f \in C^3(\mathbb{R} \to [0, \infty))$ with $M := \sup_{x \in \mathbb{R}} \left|f^{(3)}(x)\right| < \infty$. Use this inequality to conclude that $P(\tau = \infty) = 0$.

**Exercise 20.25.** Show

$$P(\sigma_b < \sigma_a) = \frac{|a|}{b + |a|} \tag{20.75}$$

and use this to conclude $P(\sigma_b < \infty) = 1$, i.e. every $b \in \mathbb{N}$ is almost surely visited by $S_n$. (This last result also follows by the Hewitt-Savage Zero-One Law, see Example 12.58 where it is shown $b$ is visited infinitely often.)

**Hint:** Using properties of martingales and Exercise 20.23, compute $\lim_{n\to\infty} \mathbb{E}[S_n^{\sigma_a \wedge \sigma_b}]$ in two different ways.

**Exercise 20.26.** Let $\tau := \sigma_a \wedge \sigma_b$. In this problem you are asked to show $\mathbb{E}[\tau] = |a| b$ with the aid of the following outline.

1. Use Exercise 20.4 above to conclude $N_n := S_n^2 - n$ is a martingale.
2. Now show
$$0 = \mathbb{E}N_0 = \mathbb{E}N_{\tau\wedge n} = \mathbb{E}S_{\tau\wedge n}^2 - \mathbb{E}[\tau \wedge n]. \tag{20.76}$$
3. Now use DCT and MCT along with Exercise 20.25 to compute the limit as $n \to \infty$ in Eq. (20.76) to find
$$\mathbb{E}[\sigma_a \wedge \sigma_b] = \mathbb{E}[\tau] = b|a|. \tag{20.77}$$
4. By considering the limit, $a \to -\infty$ in Eq. (20.77), show $\mathbb{E}[\sigma_b] = \infty$.

For the next group of exercise we are now going to suppose that $P(Z_n = 1) = p > \frac{1}{2}$ and $P(Z_n = -1) = q = 1 - p < \frac{1}{2}$. As before let $\mathcal{B}_n = \sigma(Z_1, \dots, Z_n)$, $S_0 = 0$ and $S_n = Z_1 + \dots + Z_n$ for $n \in \mathbb{N}$. Let us review the method above and what you did in Exercise 19.15 above.

In order to follow the procedures above, we start by looking for a function, $\varphi$, such that $\varphi(S_n)$ is a martingale. Such a function must satisfy,

$$\varphi(S_n) = \mathbb{E}_{\mathcal{B}_n}\varphi(S_{n+1}) = \varphi(S_n + 1) p + \varphi(S_n - 1) q,$$

and this then leads us to try to solve the following difference equation for $\varphi$;

$$\varphi(x) = p\varphi(x+1) + q\varphi(x-1) \text{ for all } x \in \mathbb{Z}. \tag{20.78}$$

Similar to the theory of second order ODE's this equation has two linearly independent solutions which could be found by solving Eq. (20.78) with initial conditions, $\varphi(0) = 1$ and $\varphi(1) = 0$ and then with $\varphi(0) = 0$ and $\varphi(1) = 0$ for example. Rather than doing this, motivated by second order constant coefficient ODE's, let us try to find solutions of the form $\varphi(x) = \lambda^x$ with $\lambda$ to be determined. Doing so leads to the equation, $\lambda^x = p\lambda^{x+1} + q\lambda^{x-1}$, or equivalently to the **characteristic equation,**

$$p\lambda^2 - \lambda + q = 0.$$

The solutions to this equation are

$$\begin{aligned}\lambda &= \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} \\ &= \frac{1 \pm \sqrt{4p^2 - 4p + 1}}{2p} = \frac{1 \pm \sqrt{(2p-1)^2}}{2p} = \{1, (1-p)/p\} = \{1, q/p\}.\end{aligned}$$

The most general solution to Eq. (20.78) is then given by

$$\varphi(x) = A + B(q/p)^x.$$

Below we will take $A = 0$ and $B = 1$. As before let $\sigma_a = \inf\{n \ge 0 : S_n = a\}$.

**Exercise 20.27.** Let $a < 0 < b$ and $\tau := \sigma_a \wedge \sigma_b$.

1. Apply the method in Exercise 20.23 with $S_n$ replaced by $M_n := (q/p)^{S_n}$ to show $P(\tau = \infty) = 0$. [Recall that $\{M_n\}_{n=1}^\infty$ is a martingale as explained in Example 20.14.]
2. Now use the method in Exercise 20.25 to show
$$P(\sigma_a < \sigma_b) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{20.79}$$

3. By letting $a \to -\infty$ in Eq. (20.79), conclude $P(\sigma_b = \infty) = 0$.
4. By letting $b \to \infty$ in Eq. (20.79), conclude $P(\sigma_a < \infty) = (q/p)^{|a|}$.

**Exercise 20.28.** Verify,

$$M_n := S_n - n(p-q)$$

and

$$N_n := M_n^2 - \sigma^2 n$$

are martingales, where $\sigma^2 = 1 - (p-q)^2$. (This should be simple; see either Exercise 20.4 or Exercise 20.3.)

**Exercise 20.29.** Using exercise 20.28, show

$$\mathbb{E}(\sigma_a \wedge \sigma_b) = \left( \frac{b\left[1-(q/p)^a\right] + a\left[(q/p)^b - 1\right]}{(q/p)^b - (q/p)^a} \right)(p-q)^{-1}. \tag{20.80}$$

By considering the limit of this equation as $a \to -\infty$, show

$$\mathbb{E}[\sigma_b] = \frac{b}{p-q}$$

and by considering the limit as $b \to \infty$, show $\mathbb{E}[\sigma_a] = \infty$.

## 20.11 Appendix: Some Alternate Proofs

This section may be safely omitted (for now).

**Proof. Alternate proof of Theorem 20.40.** Let $A \in \mathcal{B}_\sigma$. Then

$$\mathbb{E}[X_\tau - X_\sigma : A] = \mathbb{E}\left[\sum_{k=0}^{N-1} 1_{\sigma \le k < \tau} \Delta_{k+1} X : A\right] = \sum_{k=1}^{N} \mathbb{E}[\Delta_k X : A \cap \{\sigma \le k < \tau\}].$$

Since $A \in \mathcal{B}_\sigma$, $A \cap \{\sigma \le k\} \in \mathcal{B}_k$ and since $\{k < \tau\} = \{\tau \le k\}^c \in \mathcal{B}_k$, it follows that $A \cap \{\sigma \le k < \tau\} \in \mathcal{B}_k$. Hence we know that

$$\mathbb{E}[\Delta_{k+1} X : A \cap \{\sigma \le k < \tau\}] \begin{smallmatrix} \le \\ = \\ \ge \end{smallmatrix} 0 \text{ respectively.}$$

and hence that

$$\mathbb{E}[X_\tau - X_\sigma : A] \begin{smallmatrix} \le \\ = \\ \ge \end{smallmatrix} 0 \text{ respectively.}$$

Since this true for all $A \in \mathcal{B}_\sigma$, Eq. (20.23) follows. ∎

**Lemma 20.84.** *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. Then $\cup_{n=1}^\infty L^p(\Omega, \mathcal{B}_n, P)$ is dense in $L^p(\Omega, \mathcal{B}_\infty, P)$.*

**Proof.** Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$, then $M_n$ is an increasing sequence of closed subspaces of $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. Further let $\mathbb{A}$ be the algebra of functions consisting of those $f \in \cup_{n=1}^\infty M_n$ such that $f$ is bounded. As a consequence of the density Theorem 14.27, we know that $\mathbb{A}$ and hence $\cup_{n=1}^\infty M_n$ is dense in $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 14.27 in this context.

Let $\mathbb{H}$ denote those bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{A}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded $\mathcal{B}_\infty$ –measurable $\mathbb{R}$ – valued functions on $\Omega$, $1 \in \mathbb{H}$, $\mathbb{A} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated (or bounded) convergence theorem, $\lim_{n\to\infty} \|(f - f_n)\|_{L^p(P)} = 0$.[12] We may now choose $\varphi_n \in \mathbb{A}$ such that $\|\varphi_n - f_n\|_{L^p(P)} \le \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} \le \limsup_{n\to\infty} \|(f - f_n)\|_{L^p(P)} + \limsup_{n\to\infty} \|f_n - \varphi_n\|_{L^p(P)} = 0,$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 10.20, now shows $\mathbb{H}$ contains all bounded $\sigma(\mathbb{A}) = \mathcal{B}_\infty$ – measurable functions on $\Omega$. Since for any $f \in L^p(\Omega, \mathcal{B}, P)$, $f 1_{|f| \le n} \in \mathbb{H}$ there exists $\varphi_n \in \mathbb{A}$ such that $\|f_n - \varphi_n\|_p \le n^{-1}$. Using the DCT we know that $f_n \to f$ in $L^p$ and therefore by Minkowski's inequality it follows that $\varphi_n \to f$ in $L^p$. ∎

**Theorem 20.85.** *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. Then for every $X \in L^p(\Omega, \mathcal{B}, P)$, $X_n = \mathbb{E}[X|\mathcal{B}_n]$ is a martingale and $X_n \to X_\infty := \mathbb{E}[X|\mathcal{B}_\infty]$ in $L^p(\Omega, \mathcal{B}_\infty, P)$ as $n \to \infty$.*

**Proof.** We have already seen in Example 20.6 that $X_n = \mathbb{E}[X|\mathcal{B}_n]$ is always a martingale. Since conditional expectation is a contraction on $L^p$ it follows that $\mathbb{E}|X_n|^p \le \mathbb{E}|X|^p < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$. So to finish the proof we need to show $X_n \to X_\infty$ in $L^p(\Omega, \mathcal{B}, P)$ as $n \to \infty$.

Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$ and $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. If $X \in \cup_{n=1}^\infty M_n$, then $X_n = X$ for all sufficiently large $n$ and for $n = \infty$. Now suppose that $X \in M_\infty$ and $Y \in \cup_{n=1}^\infty M_n$. Then

[12] It is at this point that the proof would break down if $p = \infty$.

$$\begin{aligned}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p &\le \|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_\infty}Y\|_p+\|\mathbb{E}_{\mathcal{B}_\infty}Y-\mathbb{E}_{\mathcal{B}_n}Y\|_p+\|\mathbb{E}_{\mathcal{B}_n}Y-\mathbb{E}_{\mathcal{B}_n}X\|_p\\ &\le 2\|X-Y\|_p+\|\mathbb{E}_{\mathcal{B}_\infty}Y-\mathbb{E}_{\mathcal{B}_n}Y\|_p\end{aligned}$$

and hence

$$\limsup_{n\to\infty}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p\le 2\|X-Y\|_p.$$

Using the density Lemma 20.84 we may choose $Y\in\cup_{n=1}^\infty M_n$ as close to $X\in M_\infty$ as we please and therefore it follows that $\limsup_{n\to\infty}\|\mathbb{E}_{\mathcal{B}_\infty}X-\mathbb{E}_{\mathcal{B}_n}X\|_p=0.$

For general $X\in L^p(\Omega,\mathcal{B},P)$ it suffices to observe that $X_\infty:=\mathbb{E}[X|\mathcal{B}_\infty]\in L^p(\Omega,\mathcal{B}_\infty,P)$ and by the tower property of conditional expectations,

$$\mathbb{E}[X_\infty|\mathcal{B}_n]=\mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n]=\mathbb{E}[X|\mathcal{B}_n]=X_n.$$

So again $X_n\to X_\infty$ in $L^p$ as desired. ■

We are now ready to prove the converse of Theorem 20.85.

**Theorem 20.86.** *Suppose $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty,P)$ is a filtered probability space, $1\le p<\infty$, $\mathcal{B}_\infty:=\vee_{n=1}^\infty\mathcal{B}_n:=\sigma(\cup_{n=1}^\infty\mathcal{B}_n)$, and $\{X_n\}_{n=1}^\infty\subset L^p(\Omega,\mathcal{B},P)$ is a martingale. Further assume that $\sup_n\|X_n\|_p<\infty$ and that $\{X_n\}_{n=1}^\infty$ is uniformly integrable if $p=1$. Then there exists $X_\infty\in L^p(\Omega,\mathcal{B}_\infty,P)$ such that $X_n:=\mathbb{E}[X_\infty|\mathcal{B}_\infty]$. Moreover by Theorem 20.85 we know that $X_n\to X_\infty$ in $L^p(\Omega,\mathcal{B}_\infty,P)$ as $n\to\infty$ and hence $X_\infty$ is uniquely determined by $\{X_n\}_{n=1}^\infty$.*

**Proof.** By Theorems 15.21 and 15.23 exists $X_\infty\in L^p(\Omega,\mathcal{B}_\infty,P)$ and a subsequence, $Y_k=X_{n_k}$such that

$$\lim_{k\to\infty}\mathbb{E}[Y_kh]=\mathbb{E}[X_\infty h]\ \text{ for all } h\in L^q(\Omega,\mathcal{B}_\infty,P)$$

where $q:=p(p-1)^{-1}$. Using the martingale property, if $h\in(\mathcal{B}_n)_b$ for some $n$, it follows that $\mathbb{E}[Y_kh]=\mathbb{E}[X_nh]$ for all large $k$ and therefore that

$$\mathbb{E}[X_\infty h]=\mathbb{E}[X_nh]\ \text{ for all } h\in(\mathcal{B}_n)_b.$$

This implies that $X_n=\mathbb{E}[X_\infty|\mathcal{B}_n]$ as desired. ■

**Theorem 20.87 (Almost sure convergence).** *Suppose $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty,P)$ is a filtered probability space, $1\le p<\infty$, and let $\mathcal{B}_\infty:=\vee_{n=1}^\infty\mathcal{B}_n:=\sigma(\cup_{n=1}^\infty\mathcal{B}_n)$. Then for every $X\in L^1(\Omega,\mathcal{B},P)$, the martingale, $X_n=\mathbb{E}[X|\mathcal{B}_n]$, converges almost surely to $X_\infty:=\mathbb{E}[X|\mathcal{B}_\infty]$.*

Before starting the proof, recall from Proposition 1.5, if $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ are two bounded sequences, then

$$\begin{aligned}&\limsup_{n\to\infty}(a_n+b_n)-\liminf_{n\to\infty}(a_n+b_n)\\ &\quad\le\limsup_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\left(\liminf_{n\to\infty}a_n+\liminf_{n\to\infty}b_n\right)\\ &\quad=\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n.\end{aligned}\tag{20.81}$$

**Proof.** Since

$$X_n=\mathbb{E}[X|\mathcal{B}_n]=\mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n]=\mathbb{E}[X_\infty|\mathcal{B}_n],$$

there is no loss in generality in assuming $X=X_\infty$. If $X\in M_n:=L^1(\Omega,\mathcal{B}_n,P)$, then $X_m=X_\infty$ a.s. for all $m\ge n$ and hence $X_m\to X_\infty$ a.s. Therefore the theorem is valid for any $X$ in the dense (by Lemma 20.84) subspace $\cup_{n=1}^\infty M_n$ of $L^1(\Omega,\mathcal{B}_\infty,P)$.

For general $X\in L^1(\Omega,\mathcal{B}_\infty,P)$, let $Y_j\in\cup M_n$ such that $Y_j\to X\in L^1(\Omega,\mathcal{B}_\infty,P)$ and let $Y_{j,n}:=\mathbb{E}[Y_j|\mathcal{B}_n]$ and $X_n:=\mathbb{E}[X|\mathcal{B}_n]$. We know that $Y_{j,n}\to Y_{j,\infty}$ a.s. for each $j\in\mathbb{N}$ and our goal is to show $X_n\to X_\infty$ a.s. By Doob's inequality in Corollary 20.48 and the $L^1$ - contraction property of conditional expectation we know that

$$P(X_N^*\ge a)\le\frac{1}{a}\mathbb{E}|X_N|\le\frac{1}{a}\mathbb{E}|X|$$

and so passing to the limit as $N\to\infty$ we learn that

$$P\left(\sup_n|X_n|\ge a\right)\le\frac{1}{a}\mathbb{E}|X|\ \text{ for all } a>0.\tag{20.82}$$

Letting $a\uparrow\infty$ then shows $P(\sup_n|X_n|=\infty)=0$ and hence $\sup_n|X_n|<\infty$ a.s. Hence we may use Eq. (20.81) with $a_n=X_n-Y_{j,n}$ and $b_n:=Y_{j,n}$ to find

$$\begin{aligned}D&=\limsup_{n\to\infty}X_n-\liminf_{n\to\infty}X_n\\ &\le\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n+\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n\\ &=\limsup_{n\to\infty}a_n-\liminf_{n\to\infty}a_n\le 2\sup_n|a_n|\\ &=2\sup_n|X_n-Y_{j,n}|,\end{aligned}$$

wherein we have used $\limsup_{n\to\infty}b_n-\liminf_{n\to\infty}b_n=0$ a.s. since $Y_{j,n}\to Y_{j,\infty}$ a.s.

We now apply Doob's inequality one more time, i.e. use Eq. (20.82) with $X_n$ being replaced by $X_n-Y_{j,n}$ and $X$ by $X-Y_j$, to conclude,

$$P(D\ge a)\le P\left(\sup_n|X_n-Y_{j,n}|\ge\frac{a}{2}\right)\le\frac{2}{a}\mathbb{E}|X-Y_j|\to 0\text{ as } j\to\infty.$$

Since $a>0$ is arbitrary here, it follows that $D=0$ a.s., i.e. $\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n$ and hence $\lim_{n\to\infty} X_n$ exists in $\mathbb{R}$ almost surely. Since we already know that $X_n \to X_\infty$ in $L^1(\Omega,\mathcal{B},P)$, we may conclude that $\lim_{n\to\infty} X_n = X_\infty$ a.s.

**Alternative proof** – see Stroock [43, Corollary 5.2.7]. Let $\mathbb{H}$ denote those $X\in L^1(\Omega,\mathcal{B}_n,P)$ such that $X_n := \mathbb{E}[X|\mathcal{B}_n] \to X_\infty$ a.s. As we saw above $\mathbb{H}$ contains the dense subspace $\cup_{n=1}^\infty M_n$. It is also easy to see that $\mathbb{H}$ is a linear space. Thus it suffices to show that $\mathbb{H}$ is closed in $L^1(P)$. To prove this let $X^{(k)}\in\mathbb{H}$ with $X^{(k)}\to X$ in $L^1(P)$ and let $X_n^{(k)} := \mathbb{E}\left[X^{(k)}|\mathcal{B}_n\right]$. Then by the maximal inequality in Eq. (20.82),

$$P\left(\sup_n \left|X_n - X_n^{(k)}\right| \ge a\right) \le \frac{1}{a}\mathbb{E}\left|X - X^{(k)}\right| \text{ for all } a>0 \text{ and } k\in\mathbb{N}.$$

Therefore,

$$\begin{aligned}
&P\left(\sup_{n\ge N}|X-X_n|\ge 3a\right)\\
&\quad\le P\left(\left|X-X^{(k)}\right|\ge a\right) + P\left(\sup_{n\ge N}\left|X^{(k)}-X_n^{(k)}\right|\ge a\right)\\
&\qquad + P\left(\sup_{n\ge N}\left|X_n^{(k)}-X_n\right|\ge a\right)\\
&\quad\le \frac{2}{a}\mathbb{E}\left|X-X^{(k)}\right| + P\left(\sup_{n\ge N}\left|X^{(k)}-X_n^{(k)}\right|\ge a\right)
\end{aligned}$$

and hence

$$\limsup_{N\to\infty} P\left(\sup_{n\ge N}|X-X_n|\ge 3a\right) \le \frac{2}{a}\mathbb{E}\left|X-X^{(k)}\right| \to 0 \text{ as } k\to\infty.$$

Thus we have shown

$$\limsup_{N\to\infty} P\left(\sup_{n\ge N}|X-X_n|\ge 3a\right) = 0 \text{ for all } a>0.$$

Since

$$\left\{\limsup_{n\to\infty}|X-X_n|\ge 3a\right\} \subset \left\{\sup_{n\ge N}|X-X_n|\ge 3a\right\} \text{ for all } N,$$

it follows that

$$P\left(\limsup_{n\to\infty}|X-X_n|\ge 3a\right) = 0 \text{ for all } a>0$$

and therefore $\limsup_{n\to\infty}|X-X_n| = 0$ ($P$ a.s.) which shows that $X\in\mathbb{H}$. (This proof works equally as well in the case that $X$ is a Banach valued random variable. One only needs to replace the absolute values in the proof by the Banach norm.) ■

# 21

# Some Martingale Examples and Applications

**Exercise 21.1.** Let $S_n$ be the total assets of an insurance company in year $n \in \mathbb{N}_0$. Assume $S_0 > 0$ is a constant and that for all $n \geq 1$ that $S_n = S_{n-1} + \xi_n$, where $\xi_n = c - Z_n$ and $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables having the normal distribution with mean $\mu < c$ and variance $\sigma^2$. (The number $c$ is to be interpreted as the yearly premium.) Let $R = \{S_n \leq 0 \text{ for some } n\}$ be the event that the company eventually becomes bankrupt, i.e. is **R**uined. Show

$$P(\text{Ruin}) = P(R) \leq e^{-2(c-\mu)S_0/\sigma^2}.$$

**Outline:**

1. Show that $\lambda = -2(c-\mu)/\sigma^2 < 0$ satisfies, $\mathbb{E}\left[e^{\lambda \xi_n}\right] = 1$.
2. With this $\lambda$ show

$$Y_n := \exp(\lambda S_n) = e^{\lambda S_0} \prod_{j=1}^{n} e^{\lambda \xi_j} \tag{21.1}$$

is a non-negative $\mathcal{B}_n = \sigma(Z_1, \dots, Z_n)$ – martingale.

3. Use a martingale convergence theorem to argue that $\lim_{n\to\infty} Y_n = Y_\infty$ exists a.s. and then use Fatou's lemma to show $\mathbb{E}Y_\tau \leq e^{\lambda S_0}$.
4. Finally conclude that

$$P(R) \leq \mathbb{E}[Y_\tau : \tau < \infty] \leq \mathbb{E}Y_\tau \leq e^{\lambda S_0} = e^{-2(c-\mu)S_0/\sigma^2}.$$

Observe that by the strong law of large numbers that $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}\xi_1 = c - \mu > 0$ a.s. Thus for large $n$ we have $S_n \sim n(c-\mu) \to \infty$ as $n \to \infty$. The question we have addressed is what happens to the $S_n$ for intermediate values – in particular what is the likelyhood that $S_n$ makes a sufficiently "large deviation" from the "typical" value of $n(c-\mu)$ in order for the company to go bankrupt.

## 21.1 Aside on Large Deviations

The goal of this short section is to give a prototypical method of estimating the probabilities of unlikely events, i.e. of "**large deviations"** from the typical behaviors. For the example give here we are going to consider events of the form $\{S_n \geq n\ell\}$ where $S_n := Z_1 + \dots + Z_n$ with $\{Z_n\}_{n=1}^{\infty}$ being i.i.d. random variables. By the law of large numbers this event should be rare whenever $\ell > \mathbb{E}Z_1$. We begin with some basic notation.

**Definition 21.1.** *A real valued random variable, $Z$, is said to be* ***exponentially integrable*** *if $M(\theta) := \mathbb{E}\left[e^{\theta Z}\right] < \infty$ for all $\theta \in \mathbb{R}$. The function $M$ is called the* ***moment generating function*** *of $Z$ and we further let $\psi(\theta) := \ln M(\theta) = \ln \mathbb{E}\left[e^{\theta Z}\right]$ be the* ***log-moment generating function*** *of $Z$.*

**Theorem 21.2 (Large Deviation Upper Bound).** *Let, for $n \in \mathbb{N}$, $S_n := Z_1 + \dots + Z_n$ where $\{Z_n\}_{n=1}^{\infty}$ be i.i.d. exponentially integrable random variables such that $\mathbb{E}Z = 0$ where $Z \overset{d}{=} Z_n$. Then for all $\ell > 0$,*

$$P(S_n \geq n\ell) \leq e^{-nI(\ell)} \tag{21.2}$$

*where $I(\ell)$ is the "**Legendre transformation**" of the log-moment generating function, $\psi(\theta) = \ln \mathbb{E}\left[e^{\theta Z}\right]$, defined by*

$$I(\ell) = \sup_{\theta \geq 0} (\theta\ell - \psi(\theta)) = \sup_{\theta \in \mathbb{R}} (\theta\ell - \psi(\theta)) \geq 0. \tag{21.3}$$

*In particular,*

$$\limsup_{n\to\infty} \ln P(S_n \geq n\ell) \leq -I(\ell) \text{ for all } \ell > 0. \tag{21.4}$$

**Proof.** Let $\{Z_n\}_{n=1}^{\infty}$ be i.i.d. exponentially integrable random variables such that $\mathbb{E}Z = 0$ where $Z \overset{d}{=} Z_n$. Then for $\ell > 0$ we have for any $\theta \geq 0$ that

$$\begin{aligned} P(S_n \geq n\ell) &= P\left(e^{\theta S_n} \geq e^{\theta n\ell}\right) \\ &\leq e^{-\theta n\ell}\mathbb{E}\left[e^{\theta S_n}\right] = \left(e^{-\theta\ell}\mathbb{E}\left[e^{\theta Z}\right]\right)^n = \left(e^{-[\theta\ell - \psi(\theta)]}\right)^n \\ &= \exp(-n(\theta\ell - \psi(\theta))). \end{aligned}$$

Minimizing the far right member of this inequality over $\theta \geq 0$ gives the upper bound in Eq. (21.2) where $I(\ell)$ is given as in the first equality in Eq. (21.3).

To prove the second equality in Eq. (21.3), we use the fact that $e^{\theta x}$ is a convex function in $x$ for all $\theta \in \mathbb{R}$ and therefore by Jensen's inequality,

$$M(\theta) = \mathbb{E}\left[e^{\theta Z}\right] \geq e^{\theta \mathbb{E}X} = e^{\theta 0} = 1 \text{ for all } \theta \in \mathbb{R}.$$

This then implies that $\psi(\theta) = \ln M(\theta) \geq 0$ for all $\theta \in \mathbb{R}$. In particular, $\theta\ell - \psi(\theta) < 0$ for $\theta < 0$ while $[\theta\ell - \psi(\theta)]|_{\theta=0} = 0$ and therefore

$$\sup_{\theta\in\mathbb{R}} (\theta\ell - \psi(\theta)) = \sup_{\theta\geq 0} (\theta\ell - \psi(\theta)) \geq 0.$$

This completes the proof as Eq. (21.4) easily follows from Eq. (21.2). ∎

**Theorem 21.3 (Large Deviation Lower Bound).** *If there is a maximizer, $\theta_0$, for the the function $\theta \to \theta\ell - \psi(\theta)$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \ln P(S_n \geq n\ell) \geq -I(\ell) = \theta_0\ell - \psi(\theta_0). \tag{21.5}$$

*Combining this result with Eq. (21.4) then implies,*

$$\lim_{n\to\infty} \frac{1}{n} \ln P(S_n \geq n\ell) = -I(\ell) = \theta_0\ell - \psi(\theta_0).$$

**Proof.** If there is a maximizer, $\theta_0$, for the the function $\theta \to \theta\ell - \psi(\theta)$, then

$$0 = \ell - \psi'(\theta_0) = \ell - \frac{M'(\theta_0)}{M(\theta_0)} = \ell - \frac{\mathbb{E}\left[Ze^{\theta_0 Z}\right]}{M(\theta_0)}.$$

Thus if $W$ is a random variable with law determined by

$$\mathbb{E}[f(W)] = M(\theta_0)^{-1}\,\mathbb{E}\left[f(Z)\,e^{\theta_0 Z}\right]$$

for all non-negative functions $f : \mathbb{R} \to [0,\infty]$ then $\mathbb{E}[W] = \ell$.

Suppose that $\{W_n\}_{n=1}^\infty$ has been chosen to be a sequence of i.i.d. random variables such that $W_n \stackrel{d}{=} W$ for all $n$. Then, for all non-negative functions $f : \mathbb{R}^n \to [0,\infty]$ we have

$$\begin{aligned}\mathbb{E}[f(W_1,\dots,W_n)] &= M(\theta_0)^{-n}\,\mathbb{E}\left[f(Z_1,\dots,Z_n)\prod_{i=1}^n e^{\theta_0 Z_i}\right] \\ &= M(\theta_0)^{-n}\,\mathbb{E}\left[f(Z_1,\dots,Z_n)\,e^{\theta_0 S_n}\right].\end{aligned}$$

This is easily verified by showing the right side of this equation gives the correct expectations when $f$ is a product function. Replacing $f(z_1,\dots,z_n)$ by $M(\theta_0)^n e^{-\theta_0(z_1+\dots z_n)} f(z_1,\dots,z_n)$ in the previous equation then shows

$$\mathbb{E}[f(Z_1,\dots,Z_n)] = M(\theta_0)^n\,\mathbb{E}\left[f(W_1,\dots,W_n)\,e^{-\theta_0 T_n}\right] \tag{21.6}$$

where $T_n := W_1 + \dots + W_n$.

Taking $\delta > 0$ and $f(z_1,\dots,z_n) = 1_{z_1+\dots+z_n \geq n\ell}$ in Eq. (21.6) shows

$$\begin{aligned}P(S_n \geq n\ell) &= M(\theta_0)^n\,\mathbb{E}\left[e^{-\theta_0 T_n} : n\ell \leq T_n\right] \\ &\geq M(\theta_0)^n\,\mathbb{E}\left[e^{-\theta_0 T_n} : n\ell \leq T_n \leq n(\ell+\delta)\right] \\ &\geq M(\theta_0)^n\, e^{-n\theta_0(\ell+\delta)} P[n\ell \leq T_n \leq n(\ell+\delta)] \\ &= e^{-nI(\ell)} e^{-n\theta_0\delta} P[n\ell \leq T_n \leq n(\ell+\delta)].\end{aligned}$$

Taking logarithms of this equation, then dividing by $n$, then letting $n \to \infty$ we learn

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \ln P(S_n \geq n\ell) &\geq -I(\ell) - \theta_0\ell\delta + \lim_{n\to\infty} \frac{1}{n} \ln P[n\ell \leq T_n \leq n(\ell+\delta)] \\ &= -I(\ell) - \theta_0\ell\delta + 0\end{aligned} \tag{21.7}$$

wherein have used the central limit theorem to argue that

$$\begin{aligned}P[n\ell \leq T_n \leq n(\ell+\delta)] &= P[0 \leq T_n - n\ell \leq n\delta] \\ &= P\left[0 \leq \frac{T_n - n\ell}{\sqrt{n}} \leq \sqrt{n}\delta\right] \to \frac{1}{2} \text{ as } n \to \infty.\end{aligned}$$

Equation (21.5) now follows from Eq. (21.7) as $\delta > 0$ was arbitrary. ∎

*Example 21.4.* Suppose that $Z \stackrel{d}{=} N(0,\sigma^2) \stackrel{d}{=} \sigma N$ where $N \stackrel{d}{=} N(0,1)$, then

$$M(\theta) = \mathbb{E}\left[e^{\theta Z}\right] = \mathbb{E}\left[e^{\theta\sigma N}\right] = \exp\left(\frac{1}{2}(\sigma\theta)^2\right)$$

and therefore $\psi(\theta) = \ln M(\theta) = \frac{1}{2}\sigma^2\theta^2$. Moreover for $\ell > 0$,

$$\ell = \psi'(\theta) \implies \ell = \sigma^2\theta \implies \theta_0 = \frac{\ell}{\sigma^2}.$$

Thus it follows that

$$I(\ell) = \theta_0\ell - \psi(\theta_0) = \frac{\ell^2}{\sigma^2} - \frac{1}{2}\sigma^2\left(\frac{\ell}{\sigma^2}\right)^2 = \frac{1}{2}\frac{\ell^2}{\sigma^2}.$$

In this Gaussian case we actually know that $S_n \stackrel{d}{=} N(0, n\sigma^2)$ and therefore by Mill's ratio (see Lemma 9.71),

$$\begin{aligned}P(S_n \geq n\ell) &= P\left(\sqrt{n}\sigma N \geq n\ell\right) = P\left(N \geq \sqrt{n}\frac{\ell}{\sigma}\right) \\ &\sim \frac{1}{\sqrt{2\pi}n\frac{\ell}{\sigma}} e^{-n\frac{1}{2}\frac{\ell^2}{\sigma^2}} = \frac{\sigma}{\sqrt{2\pi}n\ell} e^{-nI(\ell)} \text{ as } n \to \infty.\end{aligned}$$

*Remark 21.5.* The technique used in the proof of Theorem 21.3 was to make a change of measure so that the large deviation (from the usual) event with small probability became typical behavior with substantial probability. One could imaging making other types of change of distribution of the form

$$\mathbb{E}[f(W)] = \frac{\mathbb{E}[f(Z)\rho(Z)]}{\mathbb{E}[\rho(Z)]}$$

where $\rho$ is some positive function. Under this change of measure the analogue of Eq. (21.6) is

$$\mathbb{E}[f(Z_1,\dots,Z_n)] = (\mathbb{E}[\rho(Z)])^n \cdot \mathbb{E}\left[f(W_1,\dots,W_n)\prod_{j=1}^n \frac{1}{\rho(W_i)}\right].$$

However to make this change of variable easy to deal with in the setting at hand we would like to further have

$$\prod_{j=1}^n \frac{1}{\rho(W_j)} = f_n(T_n) = f_n(W_1+\dots+W_n)$$

for some function $f_n$. Equivalently we would like, for some function $g_n$, that

$$\prod_{j=1}^n \rho(w_j) = g_n(w_1+\dots+w_n)$$

for all $w_i$. Taking logarithms of this equation and differentiating in the $w_j$ and $w_k$ variables shows,

$$\frac{\rho'(w_j)}{\rho(w_j)} = (\ln g_n)'(w_1+\dots+w_n) = \frac{\rho'(w_k)}{\rho(w_k)}.$$

From the extremes of this last equation we conclude that $\rho'(w_j)/\rho(w_j) = c$ (for some constant $c$) and therefore $\rho(w) = Ke^{cw}$ for some constant $K$. This helps to explain why the exponential function is used in the above proof.

## 21.2 A Polya Urn Model

In this section we are going to analyze the long run behavior of the Polya urn Markov process which was introduced in Exercise 19.4. Recall that if the urn contains $r$ red balls and $g$ green balls at a given time we draw one of these balls at random and replace it and add $c$ more balls of the same color drawn. Let $(r_n, g_n)$ be the number of red and green balls in the earn at time $n$. Then we have

$$P((r_{n+1},g_n) = (r+c,g)\,|\,(r_n,g_n)=(r,g)) = \frac{r}{r+g} \text{ and}$$
$$P((r_{n+1},g_n) = (r,g+c)\,|\,(r_n,g_n)=(r,g)) = \frac{g}{r+g}.$$

Let us observe that $r_n + g_n = r_0 + g_0 + nc$ and hence if we let $X_n$ be the fraction of green balls in the urn at time $n$,

$$X_n := \frac{g_n}{r_n+g_n},$$

then

$$X_n := \frac{g_n}{r_n+g_n} = \frac{g_n}{r_0+g_0+nc}.$$

We now claim that $\{X_n\}_{n=0}^\infty$ is a martingale relative to

$$\mathcal{B}_n := \sigma((r_k,g_k) : k\le n) = \sigma(X_k : k \le n).$$

Indeed,

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{B}_n] &= \mathbb{E}[X_{n+1}|X_n] \\ &= \frac{r_n}{r_n+g_n}\cdot\frac{g_n}{r_n+g_n+c} + \frac{g_n}{r_n+g_n}\cdot\frac{g_n+c}{r_n+g_n+c} \\ &= \frac{g_n}{r_n+g_n}\cdot\frac{r_n+g_n+c}{r_n+g_n+c} = X_n.\end{aligned}$$

Since $X_n \ge 0$ and $\mathbb{E}X_n = \mathbb{E}X_0 < \infty$ for all $n$ it follows by Corollary 20.56 that $X_\infty := \lim_{n\to\infty} X_n$ exists a.s. The distribution of $X_\infty$ is described in the next theorem.

**Theorem 21.6.** *Let $\gamma := g/c$ and $\rho := r/c$ and $\mu := \mathrm{Law}_P(X_\infty)$. Then $\mu$ is the* ***beta distribution on*** $[0,1]$ *with parameters, $\gamma, \rho$, i.e.*

$$d\mu(x) = \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} x^{\gamma-1}(1-x)^{\rho-1}\,dx \text{ for } x\in[0,1]. \tag{21.8}$$

**Proof.** We will begin by computing the distribution of $X_n$. As an example, the probability of drawing 3 greens and then 2 reds is

$$\frac{g}{r+g}\cdot\frac{g+c}{r+g+c}\cdot\frac{g+2c}{r+g+2c}\cdot\frac{r}{r+g+3c}\cdot\frac{r+c}{r+g+4c}.$$

More generally, the probability of first drawing $m$ greens and then $n-m$ reds is

$$\frac{g\cdot(g+c)\cdot\dots\cdot(g+(n-1)c)\cdot r\cdot(r+c)\cdot\dots\cdot(r+(n-m-1)c)}{(r+g)\cdot(r+g+c)\cdot\dots\cdot(r+g+(n-1)c)}.$$

Since this is the same probability for any of the $\binom{n}{m}$ – ways of drawing $m$ greens and $n-m$ reds in $n$ draws we have

$$\begin{aligned} &P\left(\text{Draw } m-\text{greens}\right) \\ &\quad=\binom{n}{m}\frac{g\cdot(g+c)\cdot\dots\cdot(g+(m-1)\,c)\cdot r\cdot(r+c)\cdot\dots\cdot(r+(n-m-1)\,c)}{(r+g)\cdot(r+g+c)\cdot\dots\cdot(r+g+(n-1)\,c)} \\ &\quad=\binom{n}{m}\frac{\gamma\cdot(\gamma+1)\cdot\dots\cdot(\gamma+(m-1))\cdot\rho\cdot(\rho+1)\cdot\dots\cdot(\rho+(n-m-1))}{(\rho+\gamma)\cdot(\rho+\gamma+1)\cdot\dots\cdot(\rho+\gamma+(n-1))}. \end{aligned} \tag{21.9}$$

Before going to the general case let us warm up with the special case, $g=r=c=1$. In this case Eq. (21.9) becomes,

$$P\left(\text{Draw } m-\text{greens}\right)=\binom{n}{m}\frac{1\cdot2\cdot\dots\cdot m\cdot1\cdot2\cdot\dots\cdot(n-m)}{2\cdot3\cdot\dots\cdot(n+1)}=\frac{1}{n+1}.$$

On the set, $\{\text{Draw } m-\text{greens}\}$, we have $X_n=\frac{1+m}{2+n}$ and hence it follows that for any $f\in C([0,1])$ that

$$\begin{aligned} \mathbb{E}\left[f\left(X_n\right)\right] &= \sum_{m=0}^{n} f\left(\frac{m+1}{n+2}\right)\cdot P\left(\text{Draw } m-\text{greens}\right) \\ &= \sum_{m=0}^{n} f\left(\frac{m+1}{n+2}\right)\frac{1}{n+1}. \end{aligned}$$

Therefore

$$\mathbb{E}\left[f\left(X\right)\right]=\lim_{n\to\infty}\mathbb{E}\left[f\left(X_n\right)\right]=\int_0^1 f\left(x\right)dx \tag{21.10}$$

and hence we may conclude that $X_\infty$ has the uniform distribution on $[0,1]$.

For the general case, recall from Example 9.50 that $n!=\Gamma(n+1)$, $\Gamma\left(t+1\right)=t\Gamma\left(t\right)$, and therefore for $m\in\mathbb{N}$,

$$\Gamma\left(x+m\right)=\left(x+m-1\right)\left(x+m-2\right)\dots\left(x+1\right)x\Gamma\left(x\right). \tag{21.11}$$

Also recall Stirling's formula in Eq. (9.61) (also see Theorem 9.72) that

$$\Gamma\left(x\right)=\sqrt{2\pi}x^{x-1/2}e^{-x}\left[1+r\left(x\right)\right] \tag{21.12}$$

where $\left|r\left(x\right)\right|\to0$ as $x\to\infty$. To finish the proof we will follow the strategy of the proof of Eq. (21.10) using Stirling's formula to estimate the expression for $P\left(\text{Draw } m-\text{greens}\right)$ in Eq. (21.9).

On the set, $\{\text{Draw } m-\text{greens}\}$, we have

$$X_n=\frac{g+mc}{r+g+nc}=\frac{\gamma+m}{\rho+\gamma+n}=:x_m,$$

where $\rho:=r/c$ and $\gamma:=g/c$. For later notice that $\Delta_m x=\frac{\gamma}{\rho+\gamma+n}$.

Using this notation we may rewrite Eq. (21.9) as

$$\begin{aligned} &P\left(\text{Draw } m-\text{greens}\right) \\ &\quad=\binom{n}{m}\frac{\frac{\Gamma(\gamma+m)}{\Gamma(\gamma)}\cdot\frac{\Gamma(\rho+n-m)}{\Gamma(\rho)}}{\frac{\Gamma(\rho+\gamma+n)}{\Gamma(\rho+\gamma)}} \\ &\quad=\frac{\Gamma\left(\rho+\gamma\right)}{\Gamma\left(\rho\right)\Gamma\left(\gamma\right)}\cdot\frac{\Gamma\left(n+1\right)}{\Gamma\left(m+1\right)\Gamma\left(n-m+1\right)}\frac{\Gamma\left(\gamma+m\right)\Gamma\left(\rho+n-m\right)}{\Gamma\left(\rho+\gamma+n\right)}. \end{aligned} \tag{21.13}$$

Now by Stirling's formula,

$$\begin{aligned} \frac{\Gamma\left(\gamma+m\right)}{\Gamma\left(m+1\right)} &= \frac{\left(\gamma+m\right)^{\gamma+m-1/2}e^{-(\gamma+m)}\left[1+r\left(\gamma+m\right)\right]}{\left(1+m\right)^{m+1-1/2}e^{-(m+1)}\left[1+r\left(1+m\right)\right]} \\ &= \left(\gamma+m\right)^{\gamma-1}\cdot\left(\frac{\gamma+m}{m+1}\right)^{m+1/2}e^{-(\gamma-1)}\frac{1+r\left(\gamma+m\right)}{1+r\left(m+1\right)}. \\ &= \left(\gamma+m\right)^{\gamma-1}\cdot\left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2}e^{-(\gamma-1)}\frac{1+r\left(\gamma+m\right)}{1+r\left(m+1\right)} \end{aligned}$$

We will keep $m$ fairly large, so that

$$\begin{aligned} \left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2} &= \exp\left(\left(m+1/2\right)\ln\left(\frac{1+\gamma/m}{1+1/m}\right)\right) \\ &\cong \exp\left(\left(m+1/2\right)\left(\gamma/m-1/m\right)\right)\cong e^{\gamma-1}. \end{aligned}$$

Hence we have

$$\frac{\Gamma\left(\gamma+m\right)}{\Gamma\left(m+1\right)}\asymp\left(\gamma+m\right)^{\gamma-1}.$$

Similarly, keeping $n-m$ fairly large, we also have

$$\begin{aligned} \frac{\Gamma\left(\rho+n-m\right)}{\Gamma\left(n-m+1\right)} &\asymp \left(\rho+n-m\right)^{\rho-1} \text{ and} \\ \frac{\Gamma\left(\rho+\gamma+n\right)}{\Gamma\left(n+1\right)} &\asymp \left(\rho+\gamma+n\right)^{\rho+\gamma-1}. \end{aligned}$$

Combining these estimates with Eq. (21.13) gives,

$$\begin{aligned}
&P\,(\text{Draw } m - \text{greens}) \\
&\quad \asymp \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\,\Gamma(\gamma)} \cdot \frac{(\gamma+m)^{\gamma-1}\cdot(\rho+n-m)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\
&\quad = \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\,\Gamma(\gamma)} \cdot \frac{\left(\frac{\gamma+m}{\rho+\gamma+n}\right)^{\gamma-1}\cdot\left(\frac{\rho+n-m}{\rho+\gamma+n}\right)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\
&\quad = \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\,\Gamma(\gamma)} \cdot (x_m)^{\gamma-1}\cdot(1-x_m)^{\rho-1}\,\Delta_m x.
\end{aligned}$$

Therefore, for any $f \in C([0,1])$, it follows that

$$\begin{aligned}
\mathbb{E}[f(X_\infty)] &= \lim_{n\to\infty}\mathbb{E}[f(X_n)] \\
&= \lim_{n\to\infty}\sum_{m=0}^{n} f(x_m)\,\frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\,\Gamma(\gamma)}\cdot(x_m)^{\gamma-1}\cdot(1-x_m)^{\rho-1}\,\Delta_m x \\
&= \int_0^1 f(x)\,\frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\,\Gamma(\gamma)} x^{\gamma-1}(1-x)^{\rho-1}\;dx.
\end{aligned}$$

■

## 21.3 Galton Watson Branching Process

This section is taken from [12, p. 245 –249]. Let $\{\xi_i^n : i, n \geq 1\}$ be a sequence of i.i.d. non-negative integer valued random variables. Suppose that $Z_n$ is the number of people in the $n^{\text{th}}$ – generation and $\xi_1^{n+1}, \dots, \xi_{Z_n}^{n+1}$ are the number of off spring of the $Z_n$ people of generation $n$. Then

$$\begin{aligned}
Z_{n+1} &= \xi_1^{n+1} + \dots + \xi_{Z_n}^{n+1} \\
&= \sum_{k=1}^{\infty}\left(\xi_1^{n+1} + \dots + \xi_k^{n+1}\right) 1_{Z_n = k}.
\end{aligned} \tag{21.14}$$

represents the number of people present in generation, $n+1$. We complete the description of the process, $Z_n$ by setting $Z_0 = 1$ and $Z_{n+1} = 0$ if $Z_n = 0$, i.e. once the population dies out it remains extinct forever after. The process $\{Z_n\}_{n\geq 0}$ is called a **Galton-Watson Branching** process, see Figure 21.1.

To understand $Z_n$ a bit better observe that

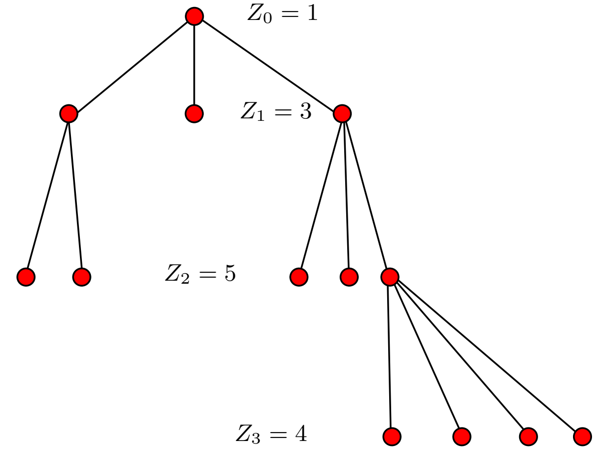



**Fig. 21.1.** A possible realization of a Galton Watson "tree."

$$\begin{aligned}
Z_0 &= 1, \\
Z_1 &= \xi_{Z_0}^1 = \xi_1^1, \\
Z_2 &= \xi_1^2 + \dots + \xi_{\xi_1^1}^2, \\
Z_3 &= \xi_1^3 + \dots + \xi_{Z_2}^3, \\
&\vdots
\end{aligned}$$

The sample path in Figure 21.1 corresponds to

$$\begin{aligned}
&\xi_1^1 = 3, \\
&\xi_1^2 = 2,\ \xi_2^2 = 0,\ \xi_3^2 = 3, \\
&\xi_1^3 = \xi_2^3 = \xi_3^3 = \xi_4^3 = 0,\ \xi_5^3 = 4, \text{ and} \\
&\xi_1^4 = \xi_2^4 = \xi_3^4 = \xi_4^4 = 0.
\end{aligned}$$

We will use later the intuitive fact that the different branches of the Galton-Watson tree evolve independently of one another – you will be asked to make this precise Exercise 21.4.

Let $\xi \overset{d}{=} \xi_i^m$, $p_k := P(\xi = k)$ be the **off-spring** distribution,

$$\mu := \mathbb{E}\xi = \sum_{k=0}^{\infty} k p_k,$$

which we assume to be finite.

Let $\mathcal{B}_0 = \{\emptyset, \Omega\}$ and

$$\mathcal{B}_n := \sigma\left(\xi_i^m : i \geq 1 \text{ and } 1 \leq m \leq n\right).$$

**Notation 21.7** *Given a bounded function $f : S = \mathbb{N}_0 \to \mathbb{C}$, let $Qf : S \to \mathbb{C}$ be defined by bounded or non-negative let*

$$Qf(0) := f(0) \text{ and}$$
$$Qf(k) := \mathbb{E}[f(Y_1 + \cdots + Y_k)] \text{ for all } k \geq 1.$$

*where $\{Y_i\}_{i=1}^{\infty}$ are i.i.d. with $P(Y_i = k) = p_k$ for all $k \in S$. Also, for $\lambda \in \mathbb{C}$ with $|\lambda| \leq 1$ let*

$$\varphi(\lambda) := \mathbb{E}\left[\lambda^{Y_1}\right] = \sum_{k \geq 0} p_k \lambda^k \tag{21.15}$$

*be the moment generating function of $\{p_k\}_{k=0}^{\infty}$.*

Let us evaluate $Qf$ for a couple of $f$. If $f(k) = k$, then

$$Qf(k) = \mathbb{E}[Y_1 + \cdots + Y_k] = k \cdot \mu \implies Qf = \mu f. \tag{21.16}$$

If$f(k) = \lambda^k$ for some $|\lambda| \leq 1$, then

$$\left(Q\lambda^{(\cdot)}\right)(k) = \mathbb{E}\left[\lambda^{Y_1 + \cdots + Y_k}\right] = \varphi(\lambda)^k. \tag{21.17}$$

*Remark 21.8.* Notice that

$$Qf(k) = \sum_{l \in S} f(l)\, p_l^{*k}$$

where

$$p_l^{*k} := P(Y_1 + \cdots + Y_k = l) = \sum_{l_1 + \cdots + l_k = l} p_{l_1} \dots p_{l_k}$$

with the convention that $p_n^{*0} = \delta_{0,n}$. For example

**Exercise 21.2.** Show that $\{Z_n\}_{n=0}^{\infty}$ is a time homogeneous Markov process with one step transition kernel being $Q$, i.e. show

$$\mathbb{E}[f(Z_{n+1})|\mathcal{B}_n] = (Qf)(Z_n) \tag{21.18}$$

for all bounded or non-negative functions $f : S = \mathbb{N}_0 \to \mathbb{C}$. In particular verify that

$$P(Z_n = j | Z_{n-1} = k) = p_j^{*k} \text{ for all } j, k \in S \text{ and } n \geq 1$$

and

$$\mathbb{E}\left[\lambda^{Z_n}|\mathcal{B}_{n-1}\right] = \varphi(\lambda)^{Z_{n-1}} \text{ a.s.} \tag{21.19}$$

for all $\lambda \in \mathbb{C}$ with $|\lambda| \leq 1$.

**Corollary 21.9.** *Continuing the notation used above, $M_n := Z_n/\mu^n$ is a positive martingale and in particular*

$$\mathbb{E}Z_n = \mu^n < \infty \text{ for all } n \in \mathbb{N}_0. \tag{21.20}$$

**Proof.** If $f(n) = n$ for all $n$, then $Qf = \mu f$ by Eq. (21.16) and therefore

$$\mathbb{E}[Z_{n+1}|\mathcal{B}_n] = Qf(Z_n) = \mu \cdot f(Z_n) = \mu Z_n.$$

Dividing this equation by $\mu^{n+1}$ then shows $\mathbb{E}[M_{n+1}|\mathcal{B}_n] = M_n$ as desired. As $M_0 = 1$ it then follows that $\mathbb{E}M_n = 1$ for all $n$ and this gives Eq. (21.20). ■

**Theorem 21.10.** *If $\mu < 1$, then, almost surely, $Z_n = 0$ for a.a. $n$.*

**Proof.** When $\mu < 1$, we have

$$\mathbb{E}\sum_{n=0}^{\infty} Z_n = \sum_{n=0}^{\infty} \mu^n = \frac{1}{1-\mu} < \infty$$

and therefore $\sum_{n=0}^{\infty} Z_n < \infty$ a.s. As $Z_n \in \mathbb{N}_0$ for all $n$, this can only happen if $Z_n = 0$ for almost all $n$ a.s. ■

**Theorem 21.11.** *If $\mu = 1$ and $P(\xi_i^m = 1) < 1$,*[1] *then again, almost surely, $Z_n = 0$ for a.a. $n$.*

**Proof.** In this case $\{Z_n\}_{n=1}^{\infty}$ is a martingale which, being positive, is $L^1$ – bounded. Therefore, $\lim_{n\to\infty} Z_n =: Z_\infty$ exists with $\mathbb{E}Z_\infty \leq 1 < \infty$. Because $Z_n$ is integer valued, it must happen that $Z_n = Z_\infty$ a.a. If $k \in \mathbb{N}$, Since

$$\{Z_\infty = k\} = \{Z_n = k \text{ a.a. } n\} = \cup_{N=1}^{\infty}\{Z_n = k \text{ for all } n \geq N\},$$

we have

$$P(Z_\infty = k) = \lim_{N\to\infty} P(Z_n = k \text{ for all } n \geq N).$$

However, if $Z_{n-1} = k$ then

$$Z_n = \xi_1^n + \cdots + \xi_{Z_{n-1}}^n = \xi_1^n + \cdots + \xi_k^n$$

and so

$$\begin{aligned} P(Z_n = k \text{ for all } n \geq N-1) &\leq P(\xi_1^n + \cdots + \xi_k^n = k \text{ for all } n \geq N) \\ &= [P(\xi_1^n + \cdots + \xi_k^n = k)]^\infty = 0, \end{aligned}$$

because, $P(\xi_1^n + \cdots + \xi_k^n = k) < 1$. Indeed, since $p_1 = P(\xi = 1) < 1$ and $\mu = 1$, it follows that $p_0 = P(\xi = 0) > 0$ and therefore

[1] The assumption here is equivalent to $p_0 > 0$ and $\mu = 1$.

$$P\left(\xi_1^n+\cdots+\xi_k^n=0\right)=\prod_{i=1}^k P\left(\xi_i^n=0\right)=p_0^k>0.$$

which then implies $P\left(\xi_1^n+\cdots+\xi_k^n=k\right)<1$. Therefore we have shown $P\left(Z_\infty=k\right)=0$ for all $k>0$ and therefore, $Z_\infty=0$ a.s. and hence almost surely, $Z_n=0$ for a.a. $n$.

[What if $k=0$, what goes wrong in the argument. Answer, now we have empty sums which are taken to be zero by definition of the process $\{Z_n\}_{n=0}^\infty$.] ■

*Remark 21.12.* By the way, the branching process, $\{Z_n\}_{n=0}^\infty$ with $\mu=1$ and $P(\xi=1)<1$ gives a nice example of a non regular martingale. Indeed, if $Z$ were regular, we would have

$$Z_n=\mathbb{E}\left[\lim_{m\to\infty}Z_m|\mathcal{B}_n\right]=\mathbb{E}\left[0|\mathcal{B}_n\right]=0$$

which is clearly false.

We now wish to consider the case where $\mu:=\mathbb{E}\left[\xi_i^m\right]>1$. Let $\varphi(\lambda)$ be as in Eq. (21.15) be the moment generating function for $\{p_k\}_{k=0}^\infty$. Notice that $\varphi(1)=1$ and for $\lambda=s\in(-1,1)$ we have

$$\varphi'(s)=\sum_{k\geq 0}kp_ks^{k-1} \text{ and } \varphi''(s)=\sum_{k\geq 0}k(k-1)p_ks^{k-2}\geq 0$$

with

$$\lim_{s\uparrow 1}\varphi'(s)=\sum_{k\geq 0}kp_k=\mathbb{E}[\xi]=:\mu \text{ and}$$
$$\lim_{s\uparrow 1}\varphi''(s)=\sum_{k\geq 0}k(k-1)p_k=\mathbb{E}[\xi(\xi-1)].$$

Therefore $\varphi$ is convex with $\varphi(0)=p_0$, $\varphi(1)=1$ and $\varphi'(1)=\mu$.

**Lemma 21.13.** *If $\mu=\varphi'(1)>1$, there exists a unique $\rho<1$ so that $\varphi(\rho)=\rho$.*

**Proof.** See Figure 21.2 below. ■

**Theorem 21.14 (See Durrett [12], p. 247-248.).** *If $\mu>1$, then*

$$P(\text{Extinction})=P\left(\left\{\lim_{n\to\infty}Z_n=0\right\}\right)=P\left(\{Z_n=0\text{ for some }n\}\right)=\rho.$$



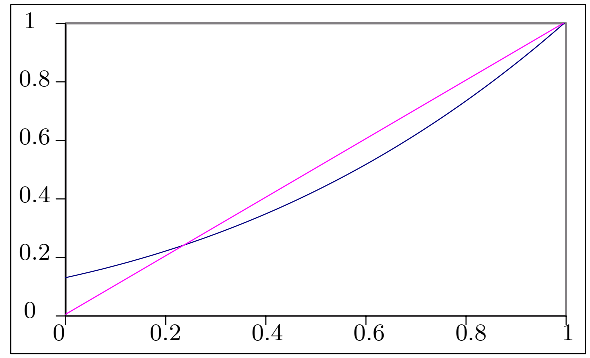



**Fig. 21.2.** Figure associated to $\varphi(s)=\frac{1}{8}\left(1+3s+3s^2+s^3\right)$ which is relevant for Exercise 3.13 of Durrett on p. 249. In this case $\rho\cong 0.236\,07$.

**Proof.** Since $\{Z_m=0\}\subset\{Z_{m+1}=0\}$, it follows that $\{Z_m=0\}\uparrow\{Z_n=0\text{ for some }n\}$ and therefore if

$$\theta_m:=P(Z_m=0),$$

then

$$P(\{Z_n=0\text{ for some }n\})=\lim_{m\to\infty}\theta_m.$$

We now show; $\theta_m=\varphi(\theta_{m-1})$. To see this, conditioned on the set $\{Z_1=k\}$, $Z_m=0$ iff all $k$ – families die out in the remaining $m-1$ time units. Since each family evolves independently, the probability[2] of this event is $\theta_{m-1}^k$. Combining this with, $P(\{Z_1=k\})=P\left(\xi_1^1=k\right)=p_k$, allows us to conclude,

$$\theta_m=P(Z_m=0)=\sum_{k=0}^\infty P(Z_m=0,Z_1=k)$$
$$=\sum_{k=0}^\infty P(Z_m=0|Z_1=k)P(Z_1=k)=\sum_{k=0}^\infty\theta_{m-1}^kp_k=\varphi(\theta_{m-1}).$$

It is now easy to see that $\theta_m\uparrow\rho$ as $m\uparrow\infty$, again see Figure 21.3. ■

**Exercise 21.3.** In the notation used in this section (Section 21.3), show for all $n\in\mathbb{N}$ and $\lambda_i\in\mathbb{C}$ with $|\lambda_i|\leq 1$ that

$$\mathbb{E}\left[\prod_{j=1}^n\lambda_i^{Z_i}\right]=\varphi\left(\lambda_1\varphi\left(\ldots\lambda_{n-2}\varphi\left(\lambda_{n-1}\varphi(\lambda_n)\right)\right)\right).$$

[2] This argument is made precise with the aid of Exercise 21.4.

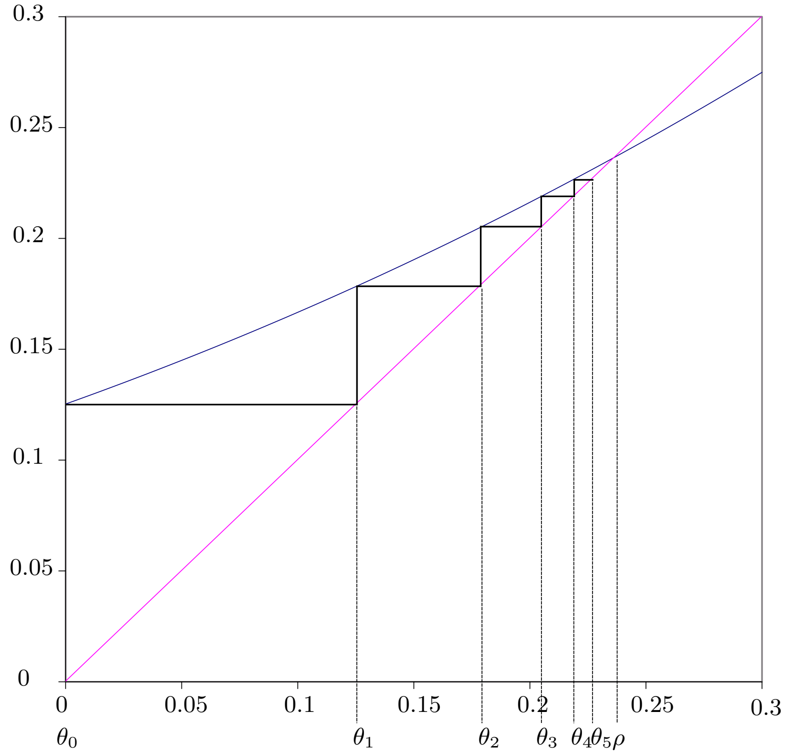



**Fig. 21.3.** The graphical interpretation of $\theta_m = \varphi(\theta_{m-1})$ starting with $\theta_0 = 0$.

For example you should show,

$$\mathbb{E}\left[\lambda_1^{Z_1}\lambda_2^{Z_2}\lambda_3^{Z_3}\right] = \varphi(\lambda_1\varphi(\lambda_2\varphi(\lambda_3)))$$

and

$$\mathbb{E}\left[\lambda_1^{Z_1}\lambda_2^{Z_2}\lambda_3^{Z_3}\lambda_4^{Z_4}\right] = \varphi(\lambda_1\varphi(\lambda_2\varphi(\lambda_3\varphi(\lambda_4)))).$$

**Exercise 21.4.** Suppose that $n \geq 2$ and $f : \mathbb{N}_0^{n-1} \to \mathbb{C}$ is a bounded function or a non-negative function. Show for all $k \geq 1$ that

$$\mathbb{E}[f(Z_2,\dots,Z_n)|Z_1 = k] = \mathbb{E}\left[f\left(\sum_{l=1}^k \left(Z_1^l,\dots,Z_{n-1}^l\right)\right)\right] \tag{21.21}$$

where $\left\{Z_n^l\right\}_{n=0}^\infty$ for $1 \leq l \leq k$ are i.i.d. Galton-Watson Branching processes such that $\left\{Z_n^l\right\}_{n=0}^\infty \overset{d}{=} \{Z_n\}_{n=0}^\infty$ for each $l$.

**Suggestion:** it suffices to prove Eq. (21.21) for $f$ of the form,

$$f(k_2,\dots,k_n) = \prod_{j=2}^n \lambda_i^{k_i}. \tag{21.22}$$

## 21.4 Kakutani's Theorem

For broad generalizations of the results in this section, see [21, Chapter IV.] or [22].

**Proposition 21.15.** *Suppose that $\mu$ and $\nu$ are $\sigma$ – finite positive measures on $(X,\mathcal{M})$, $\nu = \nu_a + \nu_s$ is the Lebesgue decomposition of $\nu$ relative to $\mu$, and $\rho : X \to [0,\infty)$ is a measurable function such that $d\nu_a = \rho d\mu$ so that*

$$d\nu = d\nu_a + d\nu_s = \rho d\mu + d\nu_s.$$

*If $g : X \to [0,\infty)$ is another measurable function such that $gd\mu \leq d\nu$, (i.e. $\int_B gd\mu \leq \nu(B)$ for all $B \in \mathcal{M}$), then $g \leq \rho$, $\mu$ – a.e.*

**Proof.** Let $A \in \mathcal{M}$ be chosen so that $\mu(A^c) = 0$ and $\nu_s(A) = 0$. Then, for all $B \in \mathcal{M}$,

$$\int_B gd\mu = \int_{B\cap A} gd\mu \leq \nu(B\cap A) = \int_{B\cap A}\rho d\mu = \int_B \rho d\mu.$$

So by the comparison Lemma 9.24, $g \leq \rho$. ■

*Example 21.16.* This example generalizes Example 20.9. Suppose $(\Omega,\mathcal{B},\{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $Q$ is **any** another probability measure on $(\Omega,\mathcal{B})$. By the Raydon-Nikodym Theorem 17.8, for each $n \in \bar{\mathbb{N}}$ we may write

$$dQ|_{\mathcal{B}_n} = X_n dP|_{\mathcal{B}_n} + dR_n \tag{21.23}$$

where $R_n$ is a measure on $(\Omega,\mathcal{B}_n)$ which is singular relative to $P|_{\mathcal{B}_n}$ and $0 \leq X_n \in L^1(\Omega,\mathcal{B}_n,P)$. In this case the most we can say in general is that $X := \{X_n\}_{n\leq\infty}$ is a positive supermartingale. To verify this assertion, for $B \in \mathcal{B}_n$ and $n \leq m \leq \infty$, we have

$$Q(B) = \mathbb{E}[X_m : B] + R_m(B) \geq \mathbb{E}[X_m : B] = \mathbb{E}[\mathbb{E}_{\mathcal{B}_n}(X_m) : B]$$

from which it follows that $\mathbb{E}_{\mathcal{B}_n}(X_m)\cdot dP|_{\mathcal{B}_n} \leq dQ|_{\mathcal{B}_n}$. So according to Proposition 21.15,

$$\mathbb{E}_{\mathcal{B}_n}(X_m) \leq X_n \ (P\text{ – a.s.}) \text{ for all } n \leq m \leq \infty. \tag{21.24}$$

**Proposition 21.17.** *Keeping the assumptions and notation used in Example 21.16, then* $\lim_{n\to\infty} X_n = X_\infty$ *a.s. and in particular the Lebesgue decomposition of* $Q|_{\mathcal{B}_\infty}$ *relative to* $P|_{\mathcal{B}_\infty}$ *may be written as*

$$dQ|_{\mathcal{B}_\infty} = \left(\lim_{n\to\infty} X_n\right) \cdot dP|_{\mathcal{B}_\infty} + dR_\infty. \tag{21.25}$$

**Proof.** By Example 21.16, we know that $\{X_n\}_{n\leq\infty}$ is a positive supermartingale and by letting $m = \infty$ in Eq. (21.24), we know

$$\mathbb{E}_{\mathcal{B}_n} X_\infty \leq X_n \text{ a.s.} \tag{21.26}$$

By the supermartingale convergence Corollary 20.65 or by the submartingale convergence Corollary 20.56 applied to $-X_n$ we know that $Y := \lim_{n\to\infty} X_n$ exists almost surely. To finish the proof it suffices to show that $Y = X_\infty$ a.s. where $X_\infty$ is defined so that Eq. (21.23) holds for $n = \infty$.

From the regular martingale convergence Theorem 20.67 we also know that $\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}_n} X_\infty = X_\infty$ a.s. as well. So passing to the limit in Eq. (21.26) implies $X_\infty \leq Y$ a.s. To prove the reverse inequality, $Y \leq X_\infty$ a.s., let $B \in \mathcal{B}_m$ and $n \geq m$. Then

$$Q(B) = \mathbb{E}[X_n : B] + R_n(B) \geq \mathbb{E}[X_n : B]$$

and so by Fatou's lemma,

$$\mathbb{E}[Y : B] = \mathbb{E}\left[\liminf_{n\to\infty} X_n : B\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n : B] \leq Q(B). \tag{21.27}$$

Since $m \in \mathbb{N}$ was arbitrary, we have proved $\mathbb{E}[Y : B] \leq Q(B)$ for all $B$ in the algebra, $\mathcal{A} := \cup_{m\in\mathbb{N}} \mathcal{B}_m$. As a consequence of the regularity Theorem 6.24 or of the monotone class Lemma 6.23, or of Theorem[3] 6.44, it follows that $\mathbb{E}[Y : B] \leq Q(B)$ for all $B \in \sigma(\mathcal{A}) = \mathcal{B}_\infty$. An application of Proposition 21.15 then implies $Y \leq X_\infty$ a.s. ■

**Theorem 21.18.** $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ *be a filtered probability space and* $Q$ *be a probability measure on* $(\Omega, \mathcal{B})$ *such that* $Q|_{\mathcal{B}_n} \ll P|_{\mathcal{B}_n}$ *for all* $n \in \mathbb{N}$. *Let* $M_n := \frac{dQ|_{\mathcal{B}_n}}{dP|_{\mathcal{B}_n}}$ *be a version of the Raydon-Nikodym derivative of* $Q|_{\mathcal{B}_n}$ *relative to* $P|_{\mathcal{B}_n}$, *see Theorem 17.8. Recall from Example 20.9 that* $\{M_n\}_{n=1}^\infty$ *is a positive martingale and let* $M_\infty = \lim_{n\to\infty} M_n$ *which exists a.s. Then the following are equivalent;*

1. $Q|_{\mathcal{B}_\infty} \ll P|_{\mathcal{B}_\infty}$,
2. $\mathbb{E}_P M_\infty = 1$,
3. $M_n \to M_\infty$ *in* $L^1(P)$, *and*
4. $\{M_n\}_{n=1}^\infty$ *is uniformly integrable.*

**Proof.** Recall from Proposition 21.17 (where $X_n$ is now $M_n$) that in general,

$$dQ|_{\mathcal{B}_\infty} = M_\infty \cdot dP|_{\mathcal{B}_\infty} + dR_\infty \tag{21.28}$$

where $R_\infty$ is singular relative to $P|_{\mathcal{B}_\infty}$. Therefore, $Q|_{\mathcal{B}_\infty} \ll P|_{\mathcal{B}_\infty}$ iff $R_\infty = 0$ which happens iff $R_\infty(\Omega) = 0$, i.e. iff

$$1 = Q(\Omega) = \int_\Omega M_\infty \cdot dP|_{\mathcal{B}_\infty} = \mathbb{E}_P M_\infty.$$

This proves the equivalence of items 1. and 2. If item 2. holds, then $M_n \to M_\infty$ by the DCT, Corollary 14.9, with $g_n = f_n = M_n$ and $g = f = M_\infty$ and so item 3. holds. The implication of 3. $\implies$ 2. is easy and the equivalence of items 3. and 4. follows from Theorem 14.51 for simply see Theorem 20.67. ■

*Remark 21.19.* Recall from Exercise 12.8, that if $0 < a_n \leq 1$, $\prod_{n=1}^\infty a_n > 0$ iff $\sum_{n=1}^\infty (1 - a_n) < \infty$. Indeed, $\prod_{n=1}^\infty a_n > 0$ iff

$$-\infty < \ln\left(\prod_{n=1}^\infty a_n\right) = \sum_{n=1}^\infty \ln a_n = \sum_{n=1}^\infty \ln(1 - (1 - a_n))$$

and $\sum_{n=1}^\infty \ln(1 - (1 - a_n)) > -\infty$ iff $\sum_{n=1}^\infty (1 - a_n) < \infty$. Recall that $\ln(1 - (1 - a_n)) \cong (1 - a_n)$ for $a_n$ near 1.

**Theorem 21.20 (Kakutani's Theorem).** *Let* $\{X_n\}_{n=1}^\infty$ *be independent non-negative random variables with* $\mathbb{E}X_n = 1$ *for all* $n$. *Further, let* $M_0 = 1$ *and* $M_n := X_1 \cdot X_2 \cdot \dots \cdot X_n$ *– a martingale relative to the filtration,* $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$ *as was shown in Example 20.11. According to Corollary 20.65,* $M_\infty := \lim_{n\to\infty} M_n$ *exists a.s. and* $\mathbb{E}M_\infty \leq 1$. *The following statements are equivalent;*

1. $\mathbb{E}M_\infty = 1$,
2. $M_n \to M_\infty$ *in* $L^1(\Omega, \mathcal{B}, P)$,
3. $\{M_n\}_{n=1}^\infty$ *is uniformly integrable,*
4. $\prod_{n=1}^\infty \mathbb{E}\left(\sqrt{X_n}\right) > 0$,
5. $\sum_{n=1}^\infty \left(1 - \mathbb{E}\left(\sqrt{X_n}\right)\right) < \infty$.

*Moreover, if any one, and hence all of the above statements,* ***fails*** *to hold, then* $P(M_\infty = 0) = 1$.

[3] This theorem implies that for $B \in \mathcal{B}$,

$$\mathbb{E}[X_0 : B] = \inf\{\mathbb{E}[X_0 : A] : A \in \mathcal{A}_\sigma\} \text{ and}$$
$$Q(B) = \inf\{Q(A) : A \in \mathcal{A}_\sigma\}$$

and since, by MCT, $\mathbb{E}[X_0 : A] \leq Q(A)$ for all $A \in \mathcal{A}_\sigma$ it follows that Eq. (21.27) holds for all $B \in \mathcal{B}$.

**Proof.** If $a_n := \mathbb{E}\left(\sqrt{X_n}\right)$, then $0 < a_n$ and $a_n^2 \le \mathbb{E}X_n = 1$ with equality iff $X_n = 1$ a.s. So Remark 21.19 gives the equivalence of items 4. and 5.

The equivalence of items 1., 2. and 3. follow by the same techniques used in the proof of Theorem 21.18 above. We will now complete the proof by showing 4. $\implies$ 3. and not(4.) $\implies$ $P(M_\infty = 0) = 1$ which clearly implies not(1.). For both pars of the argument, let $N_0 = 1$ and $N_n$ be the martingale (again see Example 20.11) defined by

$$N_n := \prod_{k=1}^n \frac{\sqrt{X_k}}{a_k} = \frac{\sqrt{M_n}}{\prod_{k=1}^n a_k}. \tag{21.29}$$

Further observe that, in all cases, $N_\infty = \lim_{n\to\infty} N_n$ exists in $[0,\infty)$ $\mu$ – a.s., see Corollary 20.56 or Corollary 20.65.

4. $\implies$ 3. Since

$$N_n^2 = \prod_{k=1}^n \frac{X_k}{a_k^2} = \frac{M_n}{\left(\prod_{k=1}^n a_k\right)^2},$$

$$\mathbb{E}\left[N_n^2\right] = \frac{\mathbb{E}M_n}{\left(\prod_{k=1}^n a_k\right)^2} = \frac{1}{\left(\prod_{k=1}^n a_k\right)^2} \le \frac{1}{\left(\prod_{k=1}^\infty a_k\right)^2} < \infty,$$

and hence $\{N_n\}_{n=1}^\infty$ is bounded in $L^2$. Therefore, using

$$M_n = \left(\prod_{k=1}^n a_k\right)^2 N_n^2 \le N_n^2 \tag{21.30}$$

and Doob's inequality in Corollary 20.48, we find

$$\mathbb{E}\left[\sup_n M_n\right] = \mathbb{E}\left[\sup_n N_n^2\right] \le 4 \sup_n \mathbb{E}\left[N_n^2\right] < \infty. \tag{21.31}$$

Equation Eq. (21.31) certainly implies $\{M_n\}_{n=1}^\infty$ is uniformly integrable, see Proposition 14.48.

Not(4.) $\implies$ $P(M_\infty = 0) = 1$. If

$$\prod_{n=1}^\infty \mathbb{E}\left(\sqrt{X_n}\right) = \lim_{n\to\infty} \prod_{k=1}^n a_k = 0,$$

we may pass to the limit in Eq. (21.30) to find

$$M_\infty = \lim_{n\to\infty} M_n = \lim_{n\to\infty}\left[\left(\prod_{k=1}^n a_k\right)^2 \cdot N_n^2\right] = 0 \cdot \left(\lim_{n\to\infty} N_n\right)^2 = 0 \text{ a.s.}.$$

■

**Lemma 21.21.** *Given two probability measures, $\mu$ and $\nu$ on a measurable space, $(\Omega,\mathcal{B})$, there exists a a positive measure $\rho$ such that $d\rho := \sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda$, where $\lambda$ is any other $\sigma$ – finite measure on $(\Omega,\mathcal{B})$ such that $\mu \ll \lambda$ and $\nu \ll \lambda$. We will write $\sqrt{d\mu\cdot d\nu}$ for $d\rho$ in the future.*

**Proof.** The main point is to show that $\rho$ is well defined. So suppose $\lambda_1$ and $\lambda_2$ are two $\sigma$ – finite measures such that $\mu \ll \lambda_i$ and $\nu \ll \lambda_i$ for $i = 1,2$. Further let $\lambda := \lambda_1 + \lambda_2$ so that $\lambda_i \ll \lambda$ for $i = 1,2$. Observe that

$$d\lambda_1 = \frac{d\lambda_1}{d\lambda}d\lambda,$$
$$d\mu = \frac{d\mu}{d\lambda_1}d\lambda_1 = \frac{d\mu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}d\lambda, \text{ and}$$
$$d\nu = \frac{d\nu}{d\lambda_1}d\lambda_1 = \frac{d\nu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}d\lambda.$$

So

$$\begin{aligned}\sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda &= \sqrt{\frac{d\mu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}\cdot\frac{d\nu}{d\lambda_1}\frac{d\lambda_1}{d\lambda}}d\lambda \\ &= \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}\frac{d\lambda_1}{d\lambda}d\lambda = \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}d\lambda_1\end{aligned}$$

and by symmetry,

$$\sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda = \sqrt{\frac{d\mu}{d\lambda_2}\cdot\frac{d\nu}{d\lambda_2}}d\lambda_2.$$

This shows

$$\sqrt{\frac{d\mu}{d\lambda_2}\cdot\frac{d\nu}{d\lambda_2}}d\lambda_2 = \sqrt{\frac{d\mu}{d\lambda_1}\cdot\frac{d\nu}{d\lambda_1}}d\lambda_1$$

and hence $d\rho = \sqrt{d\mu\cdot d\nu}$ is well defined. ■

**Definition 21.22.** *Two probability measures, $\mu$ and $\nu$ on a measure space, $(\Omega,\mathcal{B})$ are said to be **equivalent** (written $\mu \sim \nu$) if $\mu \ll \nu$ and $\nu \ll \mu$, i.e. if $\mu$ and $\nu$ are absolutely continuous relative to one another. The **Hellinger integral of** $\mu$ **and** $\nu$ is defined as*

$$H(\mu,\nu) := \int_\Omega \sqrt{d\mu\cdot d\nu} = \int_\Omega \sqrt{\frac{d\mu}{d\lambda}\cdot\frac{d\nu}{d\lambda}}d\lambda \tag{21.32}$$

*where $\lambda$ is any measure (for example $\lambda = \frac{1}{2}(\mu+\nu)$ would work) on $(\Omega,\mathcal{B})$ such that there exists, $\frac{d\mu}{d\lambda}$ and $\frac{d\nu}{d\lambda}$ in $L^1(\Omega,\mathcal{B},\lambda)$ such that $d\mu = \frac{d\mu}{d\lambda}d\lambda$ and $d\nu = \frac{d\nu}{d\lambda}d\lambda$. Lemma 21.21 guarantees that $H(\mu,\nu)$ is well defined.*

**Proposition 21.23.** *The Hellinger integral, $H(\mu,\nu)$, of two probability measures, $\mu$ and $\nu$, is well defined. Moreover $H(\mu,\nu)$ satisfies;*

1. $0 \le H(\mu,\nu) \le 1$,
2. $H(\mu,\nu) = 1$ *iff* $\mu = \nu$,
3. $H(\mu,\nu) = 0$ *iff* $\mu \perp \nu$, *and*
4. *If* $\mu \sim \nu$ *or more generally if* $\nu \ll \mu$, *then* $H(\mu,\nu) > 0$.

*Furthermore*[4],

$$H(\mu,\nu) = \inf\left\{\sum_{i=1}^{n}\sqrt{\mu(A_i)\,\nu(A_i)} : \Omega = \sum_{i=1}^{n} A_i \text{ and } n \in \mathbb{N}\right\}. \tag{21.33}$$

**Proof.** Items 1. and 2. are both an easy consequence of the Schwarz inequality and its converse. For item 3., if $H(\mu,\nu) = 0$, then $\frac{d\mu}{d\lambda}\cdot\frac{dv}{d\lambda} = 0$, $\lambda -$ a.e.. Therefore, if we let

$$A := \left\{\frac{d\mu}{d\lambda} \neq 0\right\},$$

then $\frac{d\mu}{d\lambda} = 1_A\frac{d\mu}{d\lambda} - \lambda$ –a.e. and $\frac{dv}{d\lambda}1_{A^c} = \frac{dv}{d\lambda} - \lambda -$ a.e. Hence it follows that $\mu(A^c) = 0$ and $\nu(A) = 0$ and hence $\mu \perp \nu$.

If $\nu \sim \mu$ and in particular, $v \ll \mu$, then

$$H(\mu,\nu) = \int_\Omega \sqrt{\frac{d\nu}{d\mu}\frac{d\mu}{d\mu}}d\mu = \int_\Omega \sqrt{\frac{d\nu}{d\mu}}d\mu.$$

For sake of contradiction, if $H(\mu,\nu) = 0$ then $\sqrt{\frac{d\nu}{d\mu}} = 0$ and hence $\frac{d\nu}{d\mu} = 0$, $\mu -$ a.e. The later would imply $\nu = 0$ which is impossible. Therefore, $H(\mu,\nu) > 0$ if $\nu \ll \mu$. The last statement is left to the reader as Exercise 21.6. ■

**Exercise 21.5.** Find a counter example to the statement that $H(\mu,\nu) > 0$ implies $\nu \ll \mu$.

**Exercise 21.6.** Prove Eq. (21.33).

**Corollary 21.24 (Kakutani [25]).** *Let $\Omega = \mathbb{R}^{\mathbb{N}}$, $Y_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, and $\mathcal{B} := \mathcal{B}_\infty = \sigma(Y_n : n \in \mathbb{N})$ be the product $\sigma$ – algebra on $\Omega$. Further, let $\mu := \otimes_{n=1}^\infty \mu_n$ and $\nu := \otimes_{n=1}^\infty \nu_n$ be product measures on $(\Omega, \mathcal{B}_\infty)$ associated to two sequences of probability measures, $\{\mu_n\}_{n=1}^\infty$ and $\{\nu_n\}_{n=1}^\infty$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, see Theorem 12.65 (take $\mu := P \circ (Y_1, Y_2, \dots)^{-1}$). Let us further assume that $\nu_n \ll \mu_n$ for all $n$ so that*

$$0 < H(\mu_n,\nu_n) = \int_\mathbb{R}\sqrt{\frac{d\nu_n}{d\mu_m}}d\mu_n \le 1.$$

*Then precisely one of the two cases below hold;*

1. $\sum_{n=1}^\infty (1 - H(\mu_n,\nu_n)) < \infty$ *which happens iff* $\prod_{n=1}^\infty H(\mu_n,\nu_n) > 0$ *which happens iff* $\nu \ll \mu$
*or*
2. $\sum_{n=1}^\infty (1 - H(\mu_n,\nu_n)) = \infty$ *which happens iff* $\prod_{n=1}^\infty H(\mu_n,\nu_n) = 0$ *which happens iff* $\mu \perp \nu$.

*In case 1. where $\nu \ll \mu$ we have*

$$\frac{d\nu}{d\mu} = \prod_{n=1}^\infty \frac{d\nu_n}{d\mu_n}(Y_n) \quad \mu\text{-a.s.} \tag{21.34}$$

*and in all cases we have*

$$H(\mu,\nu) = \prod_{n=1}^\infty H(\mu_n,\nu_n).$$

**Proof.** Let $P = \mu$, $Q = \nu$, $\mathcal{B}_n := \sigma(Y_1,\dots,Y_n)$, $X_n := \frac{d\nu_n}{d\mu_n}(Y_n)$, and

$$M_n := X_1 \dots X_n = \frac{d\nu_1}{d\mu_1}(Y_1)\dots\frac{d\nu_n}{d\mu_n}(Y_n).$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded measurable function, then

$$\begin{aligned}
\mathbb{E}_\nu(f(Y_1,\dots,Y_n)) &= \int_{\mathbb{R}^n} f(y_1,\dots,y_n)\,d\nu_1(y_1)\dots d\nu_n(y_n) \\
&= \int_{\mathbb{R}^n} f(y_1,\dots,y_n)\frac{d\nu_1}{d\mu_1}(y_1)\dots\frac{d\nu_n}{d\mu_n}(y_n)\;d\mu_1(y_1)\dots d\mu_n(y_n) \\
&= \mathbb{E}_\mu\left[f(Y_1,\dots,Y_n)\frac{d\nu_1}{d\mu_1}(Y_1)\dots\frac{d\nu_n}{d\mu_n}(Y_n)\right] \\
&= \mathbb{E}_\mu[f(Y_1,\dots,Y_n)M_n]
\end{aligned}$$

from which it follows that

$$d\nu|_{\mathcal{B}_n} = M_n d\mu|_{\mathcal{B}_n}.$$

Hence by Theorem 21.18, $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and the Lebesgue decomposition of $\nu$ is given by

$$d\nu = M_\infty d\mu + dR_\infty$$

[4] This statement and its proof may be safely omitted.

where $R_\infty \perp \mu$. Moreover $\nu \ll \mu$ iff $R_\infty = 0$ which happens iff $\mathbb{E}M_\infty = 1$ and $\nu \perp \mu$ iff $R_\infty = \nu$ which happens iff $M_\infty = 0$. From Theorem 21.20,

$$\mathbb{E}_\mu M_\infty = 1 \text{ iff } 0 < \prod_{n=1}^{\infty} \mathbb{E}_\mu \left(\sqrt{X_n}\right) = \prod_{n=1}^{\infty} \int_{\mathbb{R}} \sqrt{\frac{d\nu_n}{d\mu_n}} d\mu_n = \prod_{n=1}^{\infty} H(\mu_n, \nu_n)$$

and in this case

$$d\nu = M_\infty d\mu = \left(\prod_{k=1}^{\infty} X_k\right) \cdot d\mu = \left(\prod_{n=1}^{\infty} \frac{d\nu_n}{d\mu_n}(Y_n)\right) \cdot d\mu.$$

On the other hand, if

$$\prod_{n=1}^{\infty} \mathbb{E}_\mu \left(\sqrt{X_n}\right) = \prod_{n=1}^{\infty} H(\mu_n, \nu_n) = 0,$$

Theorem 21.20 implies $M_\infty = 0$, $\mu -$ a.s. in which case Theorem 21.18 implies $\nu = R_\infty$ and so $\nu \perp \mu$.

(The rest of the argument may be safely omitted.) For the last assertion, if $\prod_{n=1}^\infty H(\mu_n, \nu_n) = 0$ then $\mu \perp \nu$ and hence $H(\mu, \nu) = 0$. Conversely if $\prod_{n=1}^\infty H(\mu_n, \nu_n) > 0$, then $M_n \to M_\infty$ in $L^1(\mu)$ and therefore

$$\begin{aligned}\mathbb{E}_\mu \left[\left|\sqrt{M_n} - \sqrt{M_\infty}\right|^2\right] &\le \mathbb{E}_\mu \left[\left|\sqrt{M_n} - \sqrt{M_\infty}\right| \cdot \left|\sqrt{M_n} + \sqrt{M_\infty}\right|\right] \\ &= \mathbb{E}_\mu \left[|M_n - M_\infty|\right] \to 0 \text{ as } n \to \infty.\end{aligned}$$

Since $d\nu = M_\infty d\mu$ in this case, it follows that

$$H(\mu, \nu) = \mathbb{E}_\mu \left[\sqrt{M_\infty}\right] = \lim_{n\to\infty} \mathbb{E}_\mu \left[\sqrt{M_n}\right] = \lim_{n\to\infty} \prod_{k=1}^{n} H(\mu_k, \nu_k) = \prod_{k=1}^{\infty} H(\mu_k, \nu_k).$$

■

*Example 21.25.* Suppose that $\nu_n = \delta_1$ for all $n$ and $\mu_n = \left(1 - p_n^2\right)\delta_0 + p_n^2 \delta_1$ with $p_n \in (0, 1)$. Then $\nu_n \ll \mu_n$ with

$$\frac{d\nu_n}{d\mu_n} = 1_{\{1\}} p_n^{-2}$$

and

$$H(\mu_n, \nu_n) = \int_{\mathbb{R}} \sqrt{1_{\{1\}} p_n^{-2}} d\mu_n = \sqrt{p_n^{-2}} \cdot p_n^2 = p_n.$$

So in this case $\nu \ll \mu$ iff $\sum_{n=1}^\infty (1 - p_n) < \infty$. Observe that $\mu$ is never absolutely continuous relative to $\nu$.

On the other hand; if we further assume in Corollary 21.24 that $\mu_n \sim \nu_n$, then either; $\mu \sim \nu$ or $\mu \perp \nu$ depending on whether $\prod_{n=1}^\infty H(\mu_n, \nu_n) > 0$ or $\prod_{n=1}^\infty H(\mu_n, \nu_n) = 0$ respectively.

In the next group of problems you will be given probability measures, $\mu_n$ and $\nu_n$ on $\mathbb{R}$ and you will be asked to decide if $\mu := \otimes_{n=1}^\infty \mu_n$ and $\nu := \otimes_{n=1}^\infty \nu_n$ are equivalent. For the solutions of these problems you will want to make use of the following Gaussian integral formula;

$$\begin{aligned}\int_{\mathbb{R}} \exp\left(-\frac{a}{2}x^2 + bx\right) dx &= \int_{\mathbb{R}} \exp\left(-\frac{a}{2}\left(x - \frac{b}{a}\right)^2 + \frac{b^2}{2a}\right) dx \\ &= e^{\frac{b^2}{2a}} \int_{\mathbb{R}} \exp\left(-\frac{a}{2}x^2\right) dx = \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a}}\end{aligned}$$

which is valid for all $a > 0$ and $b \in \mathbb{R}$.

**Exercise 21.7 (A Discrete Cameron-Martin Theorem).** Suppose $t > 0$, $\{a_n\} \subset \mathbb{R}$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-(x+a_n)^2/2t} dx$ . Show $\mu \sim \nu$ iff $\sum_{k=1}^\infty a_k^2 < \infty$.

**Exercise 21.8.** Suppose $s, t > 0$, $\{a_n\} \subset \mathbb{R}$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi s}} e^{-(x+a_n)^2/2s} dx$. Show $\mu \perp \nu$ if $s \neq t$.

**Exercise 21.9.** Suppose $\{t_n\} \subset (0, \infty)$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi t_n}} e^{-x^2/2t_n} dx$. If $\sum_{n=1}^\infty (t_n - 1)^2 < \infty$ then $\mu \sim \nu$.

# Part IV

# (Weak) Convergence of Random Sums

# 22

# Random Sums

As usual let $(\Omega, \mathcal{B}, P)$ be a probability space. The general theme of this chapter is to consider arrays of random variables, $\{X_k^n\}_{k=1}^n$, for each $n \in \mathbb{N}$. We are going to look for conditions under which $\lim_{n\to\infty} \sum_{k=1}^n X_k^n$ exists almost surely or in $L^p$ for some $0 \le p < \infty$. Typically we will start with a sequence of random variables, $\{X_k\}_{k=1}^\infty$ and consider the convergence of

$$S_n = \frac{X_1 + \dots + X_n}{b_n} - a_n$$

for appropriate choices of sequence of numbers, $\{a_n\}$ and $\{b_n\}$. This fits into our general scheme by taking $X_k^n = X_k^n/b_n - a_n/n$.

## 22.1 Weak Laws of Large Numbers

**Theorem 22.1 (An $L^2$ – Weak Law of Large Numbers).** *Let $\{X_n\}_{n=1}^\infty$ be a sequence of uncorrelated square integrable random variables, $\mu_n = \mathbb{E}X_n$ and $\sigma_n^2 = \operatorname{Var}(X_n)$. If there exists an increasing positive sequence, $\{a_n\}$ and $\mu \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{j=1}^n \mu_j = \mu \text{ and } \lim_{n\to\infty} \frac{1}{a_n^2} \sum_{j=1}^n \sigma_j^2 = 0,$$

*then $\frac{S_n}{a_n} \to \mu$ in $L^2(P)$ (and hence also in probability).*

**Exercise 22.1.** Prove Theorem 22.1.

*Example 22.2.* Suppose that $\{X_k\}_{k=1}^\infty \subset L^2(P)$ are uncorrelated identically distributed random variables. Then

$$\frac{S_n}{n} \overset{L^2(P)}{\to} \mu = \mathbb{E}X_1 \text{ as } n \to \infty.$$

To see this, simply apply Theorem 22.1 with $a_n = n$. More generally if $b_n \uparrow \infty$ such that $\lim_{n\to\infty} (n/b_n^2) = 0$, then

$$\operatorname{Var}\left(\frac{S_n}{b_n}\right) = \frac{1}{b_n^2} \cdot n \operatorname{Var}(X_1) \to 0 \text{ as } n \to \infty$$

and therefore

$$(S_n - n\mu)/b_n \to 0 \text{ in } L^2(P).$$

**Note well:** since $L^2(P)$ convergence implies $L^p(P)$ – convergence for $0 \le p \le 2$, where by $L^0(P)$ – **convergence** we mean convergence in probability. The remainder of this chapter is mostly devoted to proving *a.s.* convergence for the quantities in Theorem 14.25 and Proposition 22.10 under various assumptions. These results will be described in the next section.

**Theorem 22.3 (Weak Law of Large Numbers).** *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent random variables. Let and*

$$S_n := \sum_{j=1}^n X_j \text{ and } a_n := \sum_{k=1}^n \mathbb{E}(X_k : |X_k| \le n).$$

*If*

$$\lim_{n\to\infty} \sum_{k=1}^n P(|X_k| > n) = 0 \text{ and} \tag{22.1}$$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k| \le n\right) = 0, \tag{22.2}$$

*then*

$$\frac{S_n - a_n}{n} \overset{P}{\to} 0. \tag{22.3}$$

**Proof.** A key ingredient in this proof and proofs of other versions of the law of large numbers is to introduce truncations of the $\{X_k\}$. In this case we consider

$$S_n' := \sum_{k=1}^n X_k 1_{|X_k| \le n}.$$

Since $\{S_n \ne S_{n'}\} \subset \cup_{k=1}^n \{|X_k| > n\}$,

$$P\left(\left|\frac{S_n - a_n}{n} - \frac{S_n' - a_n}{n}\right| > \varepsilon\right) = P\left(\left|\frac{S_n - S_n'}{n}\right| > \varepsilon\right)$$
$$\le P(S_n \ne S_{n'}) \le \sum_{k=1}^n P(|X_k| > n) \to 0 \text{ as } n \to \infty.$$

Hence it suffices to show $\frac{S_n'-a_n}{n} \xrightarrow{P} 0$ as $n\to\infty$ and for this it suffices to show, $\frac{S_n'-a_n}{n} \overset{L^2(P)}{\to} 0$ as $n\to\infty$.

Observe that $\mathbb{E}S_n' = a_n$ and therefore,

$$\mathbb{E}\left(\left[\frac{S_n'-a_n}{n}\right]^2\right) = \frac{1}{n^2}\operatorname{Var}\left(S_n'\right) = \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}\left(X_k 1_{|X_k|\le n}\right)$$
$$\le \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 1_{|X_k|\le n}\right) \to 0 \text{ as } n\to\infty,$$

wherein we have used $\operatorname{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \le \mathbb{E}Y^2$ in the last inequality. ■

We are now going to use this result to prove Feller's weak law of large numbers which will be valid with an assumption which is weaker than first moments existing.

*Remark 22.4.* If $X \in L^1(P)$, Chebyschev's inequality along with the dominated convergence theorem implies

$$\tau(x) := xP(|X|\ge x) \le \mathbb{E}[|X| : |X|\ge x] \to 0 \text{ as } x\to\infty.$$

If $X$ is a random variable such that $\tau(x) = xP(|X|\ge x)\to 0$ as $x\to\infty$, we say that $X$ is in "weak $L^1$."

**Exercise 22.2.** Let $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$ be the Borel $\sigma$ – algebra, $P = m$ be Lebesgue measure on $(\Omega,\mathcal{B})$, and $X(y) := (y|\ln y|)^{-1}\cdot 1_{y\le 1/2}$ for $y\in\Omega$. Show that $X\notin L^1(P)$ yet $\lim_{x\to\infty} xP(|X|\ge x) = 0$.

**Lemma 22.5.** *Let $X$ be a random variable such that $\tau(x) := xP(|X|\ge x)\to 0$ as $x\to\infty$, then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[|X|^2 : |X|\le n\right] = 0. \tag{22.4}$$

**Proof.** To prove this we observe that

$$\mathbb{E}\left[|X|^2 : |X|\le n\right] = \mathbb{E}\left[2\int 1_{0\le x\le |X|\le n}\, x dx\right] = 2\int P(0\le x\le |X|\le n)\, x dx$$
$$\le 2\int_0^n xP(|X|\ge x)\,dx = 2\int_0^n \tau(x)\,dx$$

so that

$$\frac{1}{n}\mathbb{E}\left[|X|^2 : |X|\le n\right] = \frac{2}{n}\int_0^n \tau(x)\,dx.$$

It is now easy to check (we leave it to the reader) that

$$\lim_{n\to\infty}\frac{1}{n}\int_0^n \tau(x)\,dx = 0.$$

■

**Corollary 22.6 (Feller's WLLN).** *If $\{X_n\}_{n=1}^\infty$ are i.i.d. and $\tau(x) := xP(|X_1| > x)\to 0$ as $x\to\infty$, then the hypothesis of Theorem 22.3 are satisfied so that*

$$\frac{S_n}{n} - \mathbb{E}(X_1 : |X_1|\le n) \xrightarrow{P} 0.$$

**Proof.** Since

$$\sum_{k=1}^n P(|X_k| > n) = nP(|X_1| > n) = \tau(n)\to 0 \text{ as } n\to\infty,$$

Eq. (22.1) is satisfied. Equation (22.2) follows from Lemma 22.5 and the identity,

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k|\le n\right) = \frac{1}{n}\mathbb{E}\left[|X_1|^2 : |X_1|\le n\right].$$

■

As a direct corollary of Feller's WLLN and Remark 22.4 we get Khintchin's weak law of large numbers.

**Corollary 22.7 (Khintchin's WLLN).** *If $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^1(P)$ – random variables, then $\frac{1}{n}S_n \xrightarrow{P} \mu = \mathbb{E}X_1$. This convergence holds in $L^1(P)$ as well since $\left\{\frac{1}{n}S_n\right\}_{n=1}^\infty$ is uniformly integrable under these hypothesis.*

This result is also clearly a consequence of Komogorov's strong law of large numbers.

### 22.1.1 A WLLN Example

**Theorem 22.8 (Shannon's Theorem).** *Let $\{X_i\}_{i=1}^\infty$ be a sequence of i.i.d. random variables with values in $\{1,2,\dots,r\}\subset\mathbb{N}$, $p(k) := P(X_i = k) > 0$ for $1\le k\le r$, and*

$$H(p) := -\mathbb{E}[\ln p(X_1)] = -\sum_{k=1}^r p(k)\ln p(k)$$

*be the entropy of $p = \{p_k\}_{k=1}^r$. If we define $\pi_n(\omega) := p(X_1(\omega))\dots p(X_n(\omega))$ to be the "probability of the realization" $(X_1(\omega),\dots,X_n(\omega))$, then for all $\varepsilon > 0$,*

$$P\left(e^{-n(H(p)+\varepsilon)} \le \pi_n \le e^{-n(H(p)-\varepsilon)}\right) \to 1 \text{ as } n\to\infty.$$

Thus the probability, $\pi_n$, that the random sample $\{X_1,\dots,X_n\}$ should occur is approximately $e^{-nH(p)}$ with high probability. The number $H(p)$ is called the entropy of the distribution, $\{p(k)\}_{k=1}^r$.

**Proof.** Since $\{\ln p(X_i)\}_{i=1}^\infty$ are i.i.d. it follows by the weak law of large numbers that

$$-\frac{1}{n}\ln \pi_n = -\frac{1}{n}\sum_{i=1}^n \ln p(X_i) \xrightarrow{P} -\mathbb{E}[\ln p(X_1)] = -\sum_{k=1}^r p(k)\ln p(k) =: H(p),$$

i.e. for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|H(p) + \frac{1}{n}\ln \pi_n\right| > \varepsilon\right) = 0.$$

Since

$$\begin{aligned}\left\{\left|H(p) + \frac{1}{n}\ln \pi_n\right| > \varepsilon\right\} &= \left\{H(p) + \frac{1}{n}\ln \pi_n > \varepsilon\right\} \cup \left\{H(p) + \frac{1}{n}\ln \pi_n < -\varepsilon\right\} \\ &= \left\{\frac{1}{n}\ln \pi_n > -H(p) + \varepsilon\right\} \cup \left\{\frac{1}{n}\ln \pi_n < -H(p) - \varepsilon\right\} \\ &= \left\{\pi_n > e^{n(-H(p)+\varepsilon)}\right\} \cup \left\{\pi_n < e^{n(-H(p)-\varepsilon)}\right\}\end{aligned}$$

it follows that

$$\begin{aligned}\left\{\left|H(p) + \frac{1}{n}\ln \pi_n\right| > \varepsilon\right\}^c &= \left\{\pi_n \le e^{n(-H(p)+\varepsilon)}\right\} \cap \left\{\pi_n \ge e^{n(-H(p)-\varepsilon)}\right\} \\ &= \left\{e^{-n(H(p)+\varepsilon)} \le \pi_n \le e^{-n(H(p)-\varepsilon)}\right\},\end{aligned}$$

and therefore

$$P\left(e^{-n(H(p)+\varepsilon)} \le \pi_n \le e^{-n(H(p)-\varepsilon)}\right) \to 1 \text{ as } n \to \infty.$$

■

For our next example, let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables with common distribution function, $F(x) := P(X_n \le x)$. For $x \in \mathbb{R}$ let $F_n(x)$ be the **empirical distribution function** defined by,

$$F_n(x) := \frac{1}{n}\sum_{j=1}^n 1_{X_j \le x} = \left(\frac{1}{n}\sum_{j=1}^n \delta_{X_j}\right)((-\infty, x]).$$

Since $\mathbb{E}1_{X_j \le x} = F(x)$ and $\left\{1_{X_j \le x}\right\}_{j=1}^\infty$ are Bernoulli random variables, the weak law of large numbers implies $F_n(x) \xrightarrow{P} F(x)$ as $n \to \infty$. As usual, for $p \in (0,1)$ let

$$F^{\leftarrow}(p) := \inf\{x : F(x) \ge p\}$$

and recall that $F^{\leftarrow}(p) \le x$ iff $F(x) \ge p$. Let us notice that

$$\begin{aligned}F_n^{\leftarrow}(p) = \inf\{x : F_n(x) \ge p\} &= \inf\left\{x : \sum_{j=1}^n 1_{X_j \le x} \ge np\right\} \\ &= \inf\{x : \#\{j \le n : X_j \le x\} \ge np\}.\end{aligned}$$

Recall from Definition 13.11 that the **order statistic of** $(X_1, \dots, X_n)$ is the finite sequence, $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$, where $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$ denotes $(X_1, \dots, X_n)$ arranged in increasing order with possible repetitions. It follows from the formula in Definition 13.11 that $X_k^{(n)}$ are all random variables for $k \le n$ but it will be useful to give another proof. Indeed, $X_k^{(n)} \le x$ iff $\#\{j \le n : X_j \le x\} \ge k$ iff $\sum_{j=1}^n 1_{X_j \le x} \ge k$, i.e.

$$\left\{X_k^{(n)} \le x\right\} = \left\{\sum_{j=1}^n 1_{X_j \le x} \ge k\right\} \in \mathcal{B}.$$

Moreover, if we let $\lceil x \rceil = \min\{n \in \mathbb{Z} : n \ge x\}$, the reader may easily check that $F_n^{\leftarrow}(p) = X_{\lceil np \rceil}^{(n)}$.

**Proposition 22.9.** *Keeping the notation above. Suppose that $p \in (0,1)$ is a point where*

$$F(F^{\leftarrow}(p) - \varepsilon) < p < F(F^{\leftarrow}(p) + \varepsilon) \text{ for all } \varepsilon > 0$$

*then $X_{\lceil np \rceil}^{(n)} = F_n^{\leftarrow}(p) \xrightarrow{P} F^{\leftarrow}(p)$ as $n \to \infty$. Thus we can recover, with high probability, the $p^{th}$ – quantile of the distribution $F$ by observing $\{X_i\}_{i=1}^n$.*

**Proof.** Let $\varepsilon > 0$. Then

$$\begin{aligned}\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\}^c = \{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\} &= \{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\} \\ &= \{F_n(\varepsilon + F^{\leftarrow}(p)) \ge p\}\end{aligned}$$

so that

$$\begin{aligned}\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\} &= \{F_n(F^{\leftarrow}(p) + \varepsilon) < p\} \\ &= \{F_n(\varepsilon + F^{\leftarrow}(p)) - F(\varepsilon + F^{\leftarrow}(p)) < p - F(F^{\leftarrow}(p) + \varepsilon)\}.\end{aligned}$$

Letting $\delta_\varepsilon := F(F^{\leftarrow}(p) + \varepsilon) - p > 0$, we have, as $n \to \infty$, that

$$P(\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\}) = P(F_n(\varepsilon + F^{\leftarrow}(p)) - F(\varepsilon + F^{\leftarrow}(p)) < -\delta_\varepsilon) \to 0.$$

Similarly, let $\delta_\varepsilon := p - F(F^{\leftarrow}(p) - \varepsilon) > 0$ and observe that

$$\{F^{\leftarrow}(p) - F_n^{\leftarrow}(p) \ge \varepsilon\} = \{F_n^{\leftarrow}(p) \le F^{\leftarrow}(p) - \varepsilon\} = \{F_n(F^{\leftarrow}(p) - \varepsilon) \ge p\}$$

and hence,

$$\begin{aligned}&P\left(F^{\leftarrow}(p)-F_n^{\leftarrow}(p)\geq\varepsilon\right)\\&\quad=P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right)-F\left(F^{\leftarrow}(p)-\varepsilon\right)\geq p-F\left(F^{\leftarrow}(p)-\varepsilon\right)\right)\\&\quad=P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right)-F\left(F^{\leftarrow}(p)-\varepsilon\right)\geq\delta_\varepsilon\right)\to 0\text{ as }n\to\infty.\end{aligned}$$

Thus we have shown that $X_{\lceil np\rceil}^{(n)}\overset{P}{\to}F^{\leftarrow}(p)$ as $n\to\infty$. ■

## 22.2 Kolmogorov's Convergence Criteria

**Proposition 22.10 ($L^2$ - Convergence of Random Sums).** *Suppose that $\{Y_k\}_{k=1}^\infty\subset L^2(P)$ are uncorrelated. If $\sum_{k=1}^\infty\operatorname{Var}(Y_k)<\infty$ then*

$$\sum_{k=1}^\infty(Y_k-\mu_k)\text{ converges in }L^2(P).$$

*where $\mu_k:=\mathbb{E}Y_k$.*

**Proof.** Letting $S_n:=\sum_{k=1}^n(Y_k-\mu_k)$, it suffices by the completeness of $L^2(P)$ (see Theorem 14.25) to show $\|S_n-S_m\|_2\to 0$ as $m,n\to\infty$. Supposing $n>m$, we have

$$\begin{aligned}\|S_n-S_m\|_2^2&=\mathbb{E}\left(\sum_{k=m+1}^n(Y_k-\mu_k)\right)^2\\&=\sum_{k=m+1}^n\operatorname{Var}(Y_k)=\sum_{k=m+1}^n\sigma_k^2\to 0\text{ as }m,n\to\infty.\end{aligned}$$

■

**Theorem 22.11 (Kolmogorov's Convergence Criteria).** *Suppose that $\{Y_n\}_{n=1}^\infty$ are independent square integrable random variables. If $\sum_{j=1}^\infty\operatorname{Var}(Y_j)<\infty$, then $\sum_{j=1}^\infty(Y_j-\mathbb{E}Y_j)$ converges a.s. In particular if $\sum_{j=1}^\infty\operatorname{Var}(Y_j)<\infty$ and $\sum_{j=1}^\infty\mathbb{E}Y_j$ is convergent, then $\sum_{j=1}^\infty Y_j$ converges a.s. and in $L^2(P)$.*

**Proof.** This is a special case of Theorem 20.69. Indeed, let $S_n:=\sum_{j=1}^n(Y_j-\mathbb{E}Y_j)$ with $S_0=0$. Then $\{S_n\}_{n=0}^\infty$ is a martingale relative to the filtration, $\mathcal{B}_n=\sigma(S_0,\dots S_n)$. By assumption we have

$$\mathbb{E}S_n^2=\sum_{j=1}^n\operatorname{Var}(Y_j)\leq\sum_{j=1}^\infty\operatorname{Var}(Y_j)<\infty$$

so that $\{S_n\}_{n=0}^\infty$ is bounded in $L^2(P)$. Therefore by Theorem 20.69, $\sum_{j=1}^\infty(Y_j-\mathbb{E}Y_j)=\lim_{n\to\infty}S_n$ exists a.s. and in $L^2(P)$.

Another way to prove this is to appeal Proposition 22.10 above and Lévy's Theorem 22.50 below. As second method is to make use of Kolmogorov's inequality and we will give this proof below. ■

**Exercise 22.3 (Resnik 7.1).** Does $\sum_n 1/n$ converge? Does $\sum_n(-1)^n/n$ converge? Let $\{X_n\}$ be iid with $P[X_n=\pm1]=1/2$ Does $\sum_n X_n/n$ converge? [See Example 22.41 below for a more thorough investigation of this sort.]

*Example 22.12 (Brownian Motion).* Let $\{N_n\}_{n=1}^\infty$ be i.i.d. standard normal random variable, i.e.

$$P(N_n\in A)=\int_A\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx\text{ for all }A\in\mathcal{B}_{\mathbb{R}}.$$

Let $\{\omega_n\}_{n=1}^\infty\subset\mathbb{R}$, $\{a_n\}_{n=1}^\infty\subset\mathbb{R}$, and $t\in\mathbb{R}$, then

$$\sum_{n=1}^\infty a_nN_n\sin\omega_nt\text{ converges a.s.}$$

provided $\sum_{n=1}^\infty a_n^2<\infty$. This is a simple consequence of Kolmogorov's convergence criteria, Theorem 22.11, and the facts that $\mathbb{E}[a_nN_n\sin\omega_nt]=0$ and

$$\operatorname{Var}(a_nN_n\sin\omega_nt)=a_n^2\sin^2\omega_nt\leq a_n^2.$$

As a special case, if we take $\omega_n=(2n-1)\frac{\pi}{2}$ and $a_n=\frac{\sqrt{2}}{\pi(2n-1)}$, then it follows that

$$B_t:=\frac{2\sqrt{2}}{\pi}\sum_{k=1,3,5,\dots}\frac{N_k}{k}\sin\left(k\frac{\pi}{2}t\right)\tag{22.5}$$

is a.s. convergent for all $t\in\mathbb{R}$. The factor $\frac{2\sqrt{2}}{\pi k}$ has been determined by requiring,

$$\int_0^1\left[\frac{d}{dt}\frac{2\sqrt{2}}{\pi k}\sin(k\pi t)\right]^2dt=1$$

as seen by,

$$\begin{aligned}\int_0^1\left[\frac{d}{dt}\sin\left(\frac{k\pi}{2}t\right)\right]^2dt&=\frac{k^2\pi^2}{2^2}\int_0^1\left[\cos\left(\frac{k\pi}{2}t\right)\right]^2dt\\&=\frac{k^2\pi^2}{2^2}\frac{2}{k\pi}\left[\frac{k\pi}{4}t+\frac{1}{4}\sin k\pi t\right]_0^1=\frac{k^2\pi^2}{2^3}.\end{aligned}$$

**Fact:** Wiener in 1923 showed the series in Eq. (22.5) is in fact almost surely uniformly convergent. Given this, the process, $t\to B_t$ is almost surely continuous. The process $\{B_t:0\leq t\leq1\}$ is **Brownian Motion.**

Kolmogorov's convergence criteria becomes a powerful tool when combined with the following real variable lemma.

**Lemma 22.13 (Kronecker's Lemma).** *Suppose that* $\{x_k\}_{k=1}^\infty \subset \mathbb{R}^1$ *and* $\{b_k\}_{k=1}^\infty \subset (0,\infty)$ *are sequences such that* $b_k \uparrow \infty$ *and* $\sum_{k=1}^\infty \frac{x_k}{b_k}$ *is convergent in* $\mathbb{R}$. *Then*

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{k=1}^n x_k = 0.$$

**Proof.** (We are going to use summation by parts.) Let $y_k := \frac{x_k}{b_k}$, $S_0 = 0$, $S_n := \sum_{k=1}^n y_k = \sum_{k=1}^n \frac{x_k}{b_k}$ for $n \in \mathbb{N}$, and

$$s = \sum_{k=1}^\infty y_k := \lim_{n\to\infty} S_n \in \mathbb{C}.$$

Then

$$\begin{aligned}\sum_{k=1}^n x_k &= \sum_{k=1}^n b_k y_k = \sum_{k=1}^n b_k\left(S_k - S_{k-1}\right) = \sum_{k=1}^n b_k S_k - \sum_{k=0}^{n-1} b_{k+1} S_k \\ &= b_n S_n + \sum_{k=1}^{n-1}\left(b_k - b_{k+1}\right) S_k\end{aligned}$$

and hence

$$\begin{aligned}\frac{1}{b_n}\sum_{k=1}^n x_k &= S_n - \frac{1}{b_n}\sum_{k=1}^{n-1}(b_{k+1}-b_k)S_k \\ &= S_n - \frac{1}{b_n}\sum_{k=1}^{n-1}(b_{k+1}-b_k)s + \frac{1}{b_n}\sum_{k=1}^{n-1}(b_{k+1}-b_k)\left(s - S_k\right) \\ &= S_n - \frac{1}{b_n}(b_n - b_1)s + R_n,\end{aligned}$$

where

$$R_n := \frac{1}{b_n}\sum_{k=1}^{n-1}(b_{k+1}-b_k)\left(s-S_k\right).$$

As

$$\lim_{n\to\infty}\left[S_n - \frac{1}{b_n}(b_n - b_1)s\right] = s - s = 0,$$

it suffices to show $\lim_{n\to\infty} R_n = 0$.

[1] In fact, one could replace $\mathbb{R}$ by any normed space in the Lemma.



To this end, for $N \in \mathbb{N}$ let $\varepsilon_N = \sup_{n\ge N}|S_n - s| \to 0$ so that $\varepsilon_N \downarrow 0$ as $N \uparrow \infty$. Then for $n, N \in \mathbb{N}$ with $n > N$ we have (using $b_{k+1} \ge b_k$ for all $k$) that

$$\begin{aligned}|R_n| &\le \frac{1}{b_n}\sum_{k=1}^{n-1}(b_{k+1}-b_k)\left|s - S_k\right| \\ &\le \frac{1}{b_n}\sum_{k=1}^{N}(b_{k+1}-b_k)\varepsilon_1 + \varepsilon_N \frac{1}{b_n}\sum_{k=N+1}^{n-1}(b_{k+1}-b_k) \\ &= \frac{1}{b_n}(b_{N+1}-b_1)\varepsilon_1 + \varepsilon_N\frac{1}{b_n}(b_n - b_{N+1}).\end{aligned}$$

Letting $n\to\infty$ in this last inequality then shows

$$\limsup_{n\to\infty}|R_n| \le 0 + \varepsilon_N \to 0 \text{ as } N \to \infty.$$

■

*Remark 22.14.* Here is another proof of Lemma 22.13 when $b_n = n^c$ for some $c > 0$. For this proof again let $y_k := \frac{x_k}{b_k}$, $S_n := \sum_{k=1}^n y_k$, $s := \sum_{k=1}^\infty y_k$ and further let $f_n(u) := \sum 1_{1\le k\le nu} y_k$. Then $|f_n(u)| \le M := \sup_{N\in\mathbb{N}}|S_N| < \infty$ and $f_n(u) \to s$ as $n\to\infty$ for all $u > 0$. Making use of the fundamental theorem of calculus we learn,

$$\begin{aligned}S_n - \frac{1}{b_n}\sum_{k=1}^n b_k y_k &= \sum_{k=1}^n\left(1 - \frac{b_k}{b_n}\right)y_k \\ &= \sum_{k=1}^n\left(1 - \left(\frac{k}{n}\right)^c\right)y_k \\ &= c\sum\left[1_{1\le k\le n}\int_{k/n}^1 u^{c-1}du\right]y_k \\ &= c\int_0^1 du u^{c-1}\sum\left[1_{1\le k\le n}\cdot 1_{\frac{k}{n}\le u}\right]y_k \\ &= c\int_0^1 du u^{c-1} f_n(u) \to c\int_0^1 du u^{c-1} s = s \text{ as } n\to\infty.\end{aligned}$$

Since $S_n \to s$ as $n\to\infty$, the result is proved.

As an immediate corollary we have the following corollary.

**Corollary 22.15 ($L^2$ − SSLN).** *Let* $\{X_n\}$ *be a sequence of independent random variables such that* $\sigma^2 = \mathbb{E}X_n^2 < \infty$ *and* $\mu = \mathbb{E}X_n$ *are independent of* $n$. *As above let* $S_n = \sum_{k=1}^n X_k$. *If* $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ *is a sequence such that* $b_n \uparrow \infty$ *and* $\sum_{n=1}^\infty \frac{1}{b_n^2} < \infty$, *then*

$$\frac{1}{b_n}(S_n - n\mu) \to 0 \ a.s. \ and \ in \ L^2(P) \tag{22.6}$$

We may rewrite Eq. (22.6) as

$$S_n = n\mu + o(1)\, b_n \text{ or } \frac{S_n}{n} = \mu + o(1)\frac{b_n}{n}.$$

*Example 22.16.* For example, we could take $b_n = n$ or $b_n = n^p$ for an $p > 1/2$, or $b_n = n^{1/2}(\ln n)^{1/2+\varepsilon}$ for any $\varepsilon > 0$. The idea here is that

$$\sum_{n=2}^{\infty} \frac{1}{\left(n^{1/2}(\ln n)^{1/2+\varepsilon}\right)^2} = \sum_{n=2}^{\infty} \frac{1}{n(\ln n)^{1+2\varepsilon}}$$

which may be analyzed by comparison with the integral

$$\int_2^{\infty} \frac{1}{x \ln^{1+2\varepsilon} x} dx = \int_{\ln 2}^{\infty} \frac{1}{e^y y^{1+2\varepsilon}} e^y dy = \int_{\ln 2}^{\infty} \frac{1}{y^{1+2\varepsilon}} dy < \infty,$$

wherein we have made the change of variables, $y = \ln x$. When $b_n = n^{1/2}(\ln n)^{1/2+\varepsilon}$ we may conclude that

$$\frac{S_n}{n} = \mu + o(1)\frac{(\ln n)^{1/2+\varepsilon}}{n^{1/2}},$$

i.e. the fluctuations of $\frac{S_n}{n}$ about the mean, $\mu$, have order smaller than $n^{-1/2}(\ln n)^{1/2+\varepsilon}$.

**Fact 22.17 (Missing Reference)** *Under the hypothesis in Corollary 22.15,*

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n^{1/2}(\ln \ln n)^{1/2}} = \sqrt{2}\sigma \ a.s.$$

We end this section with another example of using Kolmogorov's convergence criteria in conjunction with Kronecker's Lemma 22.13.

**Lemma 22.18.** *Let $\{X_n\}_{n=1}^{\infty}$ be independent square integrable random variables such that $\mathbb{E}S_n \uparrow \infty$ as $n \to \infty$. Then*

$$\sum_{n=1}^{\infty} \mathrm{Var}\left(\frac{X_n}{\mathbb{E}S_n}\right) = \sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{(\mathbb{E}S_n)^2} < \infty \implies \frac{S_n}{\mathbb{E}S_n} \to 1 \ a.s.$$

**Proof.** Kolmogorov's convergence criteria, Theorem 22.11 we know that

$$\sum_{n=1}^{\infty} \frac{X_n - \mathbb{E}X_n}{\mathbb{E}S_n} \text{ is a.s. convergent.}$$

It then follows by Kronecker's Lemma 22.13 that

$$0 = \lim_{n\to\infty} \frac{1}{\mathbb{E}S_n} \sum_{i=1}^{n} (X_n - \mathbb{E}X_n) = \lim_{n\to\infty} \frac{S_n}{\mathbb{E}S_n} - 1 \text{ a.s.}$$

■

*Example 22.19.* Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. square integrable random variables with $\mu := \mathbb{E}X_n > 0$ and $\sigma^2 := \mathrm{Var}(X_n) < \infty$. Since $\mathbb{E}S_n = \mu n \uparrow \infty$ and

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{(\mathbb{E}S_n)^2} = \sum_{n=1}^{\infty} \frac{\sigma^2}{\mu^2 n^2} < \infty,$$

we may conclude that $\lim_{n\to\infty} \frac{S_n}{\mu n} = 1$ a.s., i.e. $S_n/n \to \mu$ a.s. as we already know.

We now assume that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables with a continuous distribution function and let $A_j$ denote the event when $X_j$ is a record, i.e.

$$A_j := \{X_j > \max\{X_1, X_2, \dots, X_{k-1}\}\}.$$

Recall from Renyi Theorem 12.23 that $\{A_j\}_{j=1}^{\infty}$ are independent and $P(A_j) = \frac{1}{j}$ for all $j$.

**Proposition 22.20.** *Keeping the preceding notation and let $S_n := \sum_{j=1}^{n} 1_{A_j}$ denote the number of records in the first $n$ observations. Then $\lim_{n\to\infty} \frac{S_n}{\ln n} = 1$ a.s.*

**Proof.** In this case

$$\mathbb{E}S_n = \sum_{j=1}^{n} \mathbb{E}1_{A_j} = \sum_{j=1}^{n} \frac{1}{j} \sim \int_1^n \frac{1}{x} dx = \ln n \uparrow \infty$$

and

$$\mathrm{Var}(1_{A_n}) = \mathbb{E}1_{A_n}^2 - (\mathbb{E}1_{A_n})^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$$

so by that

$$\begin{aligned}\sum_{n=1}^{\infty} \mathrm{Var}\left(\frac{1_{A_n}}{\mathbb{E}S_n}\right) &= \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n^2}\right) \frac{1}{\left(\sum_{j=1}^{n} \frac{1}{j}\right)^2} \\ &\le \sum_{n=1}^{\infty} \frac{1}{\left(\sum_{j=1}^{n} \frac{1}{j}\right)^2} \frac{1}{n} \\ &\lessapprox 1 + \int_2^{\infty} \frac{1}{\ln^2 x} \frac{1}{x} dx = 1 + \int_{\ln 2}^{\infty} \frac{1}{y^2} dy < \infty.\end{aligned}$$

Therefore by Lemma 22.18 we may conclude that $\lim_{n\to\infty}\frac{S_n}{\mathbb{E}S_n}=1$ a.s.

So to finish the proof it only remains to show

$$\lim_{n\to\infty}\frac{\mathbb{E}S_n}{\ln n}\lim_{n\to\infty}\frac{\sum_{j=1}^n\frac{1}{j}}{\ln n}=1. \tag{22.7}$$

To see this write

$$\begin{aligned}\ln(n+1)&=\int_1^{n+1}\frac{1}{x}dx=\sum_{j=1}^n\int_j^{j+1}\frac{1}{x}dx\\&=\sum_{j=1}^n\int_j^{j+1}\left(\frac{1}{x}-\frac{1}{j}\right)dx+\sum_{j=1}^n\frac{1}{j}\\&=\rho_n+\sum_{j=1}^n\frac{1}{j}\end{aligned} \tag{22.8}$$

where

$$|\rho_n|=\sum_{j=1}^n\left|\ln\frac{j+1}{j}-\frac{1}{j}\right|=\sum_{j=1}^n\left|\ln(1+1/j)-\frac{1}{j}\right|\sim\sum_{j=1}^n\frac{1}{j^2}$$

and hence we conclude that $\lim_{n\to\infty}\rho_n<\infty$. So dividing Eq. (22.8) by $\ln n$ and letting $n\to\infty$ gives the desired limit in Eq. (22.7). ■

## 22.3 The Strong Law of Large Numbers Revisited

*Remark 22.21.* Here is a brief summary of the main results of this section. Suppose that $\{X_k\}_{k=1}^\infty$ are i.i.d. random variables, then the following are equivalent:

1. $\lim_{n\to\infty}\frac{S_n}{n}$ exists a.s. as an $\mathbb{R}$ – valued random variable
2. there exists $c\in\mathbb{R}$ such that $\lim_{n\to\infty}\frac{S_n}{n}=c$ a.s.
3. $\lim_{n\to\infty}\frac{X_n}{n}=0$ a.s.
4. $\mathbb{E}|X_1|<\infty$.

Indeed, 1. $\implies$ 2. by Kolmogorov's zero one law, see Corollary 12.54. 2. $\implies$ 3. $\implies$ 4. is contained in Proposition 22.27 and Corollary 22.28. 4. $\implies$ 2.($\implies$ 1.) is Theorem 22.31.

**Definition 22.22.** *Two sequences, $\{X_n\}$ and $\{X_n'\}$, of random variables are* ***tail equivalent*** *if*

$$\mathbb{E}\left[\sum_{n=1}^\infty 1_{X_n\neq X_n'}\right]=\sum_{n=1}^\infty P(X_n\neq X_n')<\infty.$$

**Proposition 22.23.** *Suppose $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Then*

1. *$\sum(X_n-X_n')$ converges a.s.*
2. *The sum $\sum X_n$ is convergent a.s. iff the sum $\sum X_n'$ is convergent a.s. More generally we have*
$$P\left(\left\{\sum X_n \text{ is convergent}\right\}\triangle\left\{\sum X_n' \text{ is convergent}\right\}\right)=0$$
3. *If there exists a random variable, $X$, and a sequence $a_n\uparrow\infty$ such that*
$$\lim_{n\to\infty}\frac{1}{a_n}\sum_{k=1}^n X_k=X \ \ a.s$$
*then*
$$\lim_{n\to\infty}\frac{1}{a_n}\sum_{k=1}^n X_k'=X \ \ a.s$$

**Proof.** If $\{X_n\}$ and $\{X_n'\}$ are tail equivalent, we know by the first Borel - Cantelli Lemma 9.14 that $P(X_n=X_n' \text{ for a.a. } n)=1$. The proposition is an easy consequence of this observation. ■

*Remark 22.24.* In what follows we will typically have a sequence, $\{X_n\}_{n=1}^\infty$, of independent random variables and $X_n'=f_n(X_n)$ for some "cutoff" functions, $f_n:\mathbb{R}\to\mathbb{R}$. In this case the collection of sets, $\{A_n:=\{X_n\neq X_n'\}\}_{n=1}^\infty$ are independent and so by the Borel zero one law (Lemma 12.44) we will have

$$P(X_n\neq X_n' \text{ i.o. } n)=0\iff\sum_{n=1}^\infty P(X_n\neq X_n')<\infty.$$

So in this case $\{X_n\}$ and $\{X_n'\}$ are tail equivalent iff $P(X_n=X_n' \text{ a.a. } n)=1$. For example if $\{k_n\}_{n=1}^\infty\subset(0,\infty)$ and $X_n':=X_n\cdot 1_{|X_n|\leq k_n}$ then the following are equivalent;

1. $P(|X_n|\leq k_n \text{ a.a. } n)=1$,
2. $P(|X_n|>k_n \text{ i.o. } n)=0$,
3. $\sum_{n=1}^\infty P(X_n\neq X_n')=\sum_{n=1}^\infty P(|X_n|>k_n)<\infty$,
4. $\{X_n\}$ and $\{X_n'\}$ are tail equivalent.

**Lemma 22.25.** *Suppose that $X:\Omega\to\mathbb{R}$ is a random variable, then*

$$\mathbb{E}|X|^p=\int_0^\infty ps^{p-1}P(|X|\geq s)\,ds=\int_0^\infty ps^{p-1}P(|X|>s)\,ds.$$

**Proof.** By the fundamental theorem of calculus,

$$|X|^p = \int_0^{|X|} ps^{p-1}ds = p\int_0^\infty 1_{s\le|X|}\cdot s^{p-1}ds = p\int_0^\infty 1_{s<|X|}\cdot s^{p-1}ds.$$

Taking expectations of this identity along with an application of Tonelli's theorem completes the proof. ∎

**Lemma 22.26.** *If $X$ is a random variable and $\varepsilon>0$, then*

$$\sum_{n=1}^\infty P(|X|\ge n\varepsilon)\le\frac{1}{\varepsilon}\mathbb{E}|X|\le\sum_{n=0}^\infty P(|X|\ge n\varepsilon). \tag{22.9}$$

**Proof.** First observe that for all $y\ge0$ we have,

$$\sum_{n=1}^\infty 1_{n\le y}\le y\le\sum_{n=1}^\infty 1_{n\le y}+1=\sum_{n=0}^\infty 1_{n\le y}. \tag{22.10}$$

Taking $y=|X|/\varepsilon$ in Eq. (22.10) and then take expectations gives the estimate in Eq. (22.9). ∎

**Proposition 22.27.** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables, then the following are equivalent:*

1. $\mathbb{E}|X_1|<\infty$.
2. *There exists $\varepsilon>0$ such that $\sum_{n=1}^\infty P(|X_1|\ge\varepsilon n)<\infty$.*
3. *For all $\varepsilon>0$, $\sum_{n=1}^\infty P(|X_1|\ge\varepsilon n)<\infty$.*
4. *$\{X_n\}_{n=1}^\infty$ and $\{X_n 1_{|X_n|\le n}\}_{n=1}^\infty$ are tail equivalent.*
5. $\lim_{n\to\infty}\frac{|X_n|}{n}=0$ *a.s.*

**Proof.** The equivalence of items 1., 2., and 3. easily follows from Lemma 22.26 and their equivalence with item 4. is explained in Remark 22.24. So to finish the proof it suffices to show 3. is equivalent to 5. To this end we start by noting that $\lim_{n\to\infty}\frac{|X_n|}{n}=0$ a.s. iff

$$0=P\left(\frac{|X_n|}{n}\ge\varepsilon\text{ i.o.}\right)=P(|X_n|\ge n\varepsilon\text{ i.o.})\text{ for all }\varepsilon>0. \tag{22.11}$$

Because $\{|X_n|\ge n\varepsilon\}_{n=1}^\infty$ are independent sets, the Borel zero-one law (Lemma 12.44) shows the statement in Eq. (22.11) is equivalent to $\sum_{n=1}^\infty P(|X_n|\ge n\varepsilon)<\infty$ for all $\varepsilon>0$. ∎

**Corollary 22.28.** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables such that $\frac{1}{n}S_n\to c\in\mathbb{R}$ a.s., then $X_n\in L^1(P)$ and $\mu:=\mathbb{E}X_n=c$.*

**Proof.** If $\frac{1}{n}S_n\to c$ a.s. then

$$\frac{X_{n+1}}{n+1}=\frac{1}{n+1}S_{n+1}-\frac{n}{n+1}\frac{1}{n}S_n\to c-1\cdot c=0\text{ a.s. as }n\to\infty.$$

Hence an application of Proposition 22.27 shows $X_n\in L^1(P)$. Moreover by Exercise 14.6, $\left\{\frac{1}{n}S_n\right\}_{n=1}^\infty$ is a uniformly integrable sequenced and therefore,

$$\mu=\mathbb{E}\left[\frac{1}{n}S_n\right]\to\mathbb{E}\left[\lim_{n\to\infty}\frac{1}{n}S_n\right]=\mathbb{E}[c]=c.$$

∎

**Lemma 22.29.** *For all $x>0$,*

$$\varphi(x):=\sum_{n=1}^\infty\frac{1}{n^2}1_{x\le n}=\sum_{n\ge x}\frac{1}{n^2}\le2\frac{1}{x\vee1}=2\cdot\min\left(\frac{1}{x},1\right).$$

**Proof.** First office notice that

$$\sum_{n=2}^\infty\frac{1}{n^2}1_{n\le t<n+1}\le\frac{1}{(t-1)^2}\text{ for }t\ge2.$$

Therefore for any $x>1$ we have

$$\begin{aligned}\sum_{n\ge x}\frac{1}{n^2}&\le\int_x^\infty dt\sum_{n=2}^\infty\frac{1}{n^2}1_{n\le t<n+1}\\&\le\int_x^\infty\frac{1}{(t-1)^2}dt\le\frac{1}{x-1}\le\frac{2}{x}.\end{aligned}$$

The last inequality also holds for $x=1$ as

$$\sum_{n\ge1}\frac{1}{n^2}=1+\sum_{n\ge2}\frac{1}{n^2}\le1+\frac{2}{2}=2=\frac{2}{1}.$$

∎

**Lemma 22.30.** *Suppose that $X:\Omega\to\mathbb{R}$ is a random variable, then*

$$\sum_{n=1}^\infty\frac{1}{n^2}\mathbb{E}\left[|X|^2:1_{|X|\le n}\right]\le2\mathbb{E}|X|.$$

**Proof.** This is a simple application of Lemma 22.29;

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\left[|X|^2 : 1_{|X| \le n}\right] = \mathbb{E}\left[|X|^2 \sum_{n=1}^{\infty} \frac{1}{n^2} 1_{|X| \le n}\right] = \mathbb{E}\left[|X|^2 \varphi(|X|)\right]$$
$$\le 2\mathbb{E}\left[|X|^2 \left(\frac{1}{|X|} \wedge 1\right)\right] \le 2\mathbb{E}|X|.$$

■

With this as preparation we are now in a position to give another proof of the Kolmogorov's strong law of large numbers which has already appeared in Theorem 18.10 and Example 20.82.

**Theorem 22.31 (Kolmogorov's Strong Law of Large Numbers).** *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. Then there exists $\mu \in \mathbb{R}$ such that $\frac{1}{n} S_n \to \mu$ a.s. iff $X_n$ is integrable and in which case $\mathbb{E}X_n = \mu$.*

**Proof.** The implication, $\frac{1}{n} S_n \to \mu$ a.s. implies $X_n \in L^1(P)$ and $\mathbb{E}X_n = \mu$ has already been proved in Corollary 22.28. So let us now assume $X_n \in L^1(P)$ and let $\mu := \mathbb{E}X_n$.

Let $X_n' := X_n 1_{|X_n| \le n}$. By Lemma 22.26,

$$\sum_{n=1}^{\infty} P(X_n' \ne X_n) = \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \le \mathbb{E}|X_1| < \infty,$$

and hence $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Therefore, by Proposition 22.23, it suffices to show $\lim_{n\to\infty} \frac{1}{n} S_n' = \mu$ a.s. where $S_n' := X_1' + \cdots + X_n'$. But by Lemma 22.30,

$$\sum_{n=1}^{\infty} \frac{\operatorname{Var}(X_n')}{n^2} \le \sum_{n=1}^{\infty} \frac{\mathbb{E}|X_n'|^2}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[|X_n|^2 1_{|X_n| \le n}\right]}{n^2}$$
$$= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[|X_1|^2 1_{|X_1| \le n}\right]}{n^2} \le 2\mathbb{E}|X_1| < \infty. \quad (22.12)$$

Therefore by Kolmogorov's convergence criteria, Theorem 22.11,

$$\sum_{n=1}^{\infty} \frac{X_n' - \mathbb{E}X_n'}{n} \text{ is almost surely convergent.}$$

Kronecker's Lemma 22.13 then implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (X_k' - \mathbb{E}X_k') = 0 \text{ a.s.}$$

So to finish the proof, it only remains to observe

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}X_k' = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_k 1_{|X_k| \le k}\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_1 1_{|X_1| \le k}\right] = \mu.$$

Here we have used the dominated convergence theorem to see that $a_k := \mathbb{E}\left[X_1 1_{|X_1| \le k}\right] \to \mu$ as $k \to \infty$ from which it is easy (and standard) to check that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = \mu$. ■

*Remark 22.32.* If $\mathbb{E}|X_1| = \infty$ but $\mathbb{E}X_1^- < \infty$, then $\frac{1}{n} S_n \to \infty$ a.s. To prove this, for $M > 0$ let $X_n^M := X_n \wedge M$ and $S_n^M := \sum_{i=1}^{n} X_i^M$. It follows from Theorem 22.31 that $\frac{1}{n} S_n^M \to \mu^M := \mathbb{E}X_1^M$ a.s.. Since $S_n \ge S_n^M$, we may conclude that

$$\liminf_{n\to\infty} \frac{S_n}{n} \ge \liminf_{n\to\infty} \frac{1}{n} S_n^M = \mu^M \text{ a.s.}$$

Since $\mu^M \to \infty$ as $M \to \infty$, it follows that $\liminf_{n\to\infty} \frac{S_n}{n} = \infty$ a.s. and hence that $\lim_{n\to\infty} \frac{S_n}{n} = \infty$ a.s.

**Exercise 22.4 (Resnik 7.9).** Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = 0$. Following the ideas in the proof of Theorem 22.31, show for any bounded sequence $\{c_n\}_{n=1}^{\infty}$ of real numbers that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} c_k X_k = 0 \text{ a.s.}$$

### 22.3.1 Strong Law of Large Number Examples

*Example 22.33 (Renewal Theory).* Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. non-negative integrable random variables such that $P(X_i > 0) > 0$. Think of the $X_i$ as the life time of bulb number $i$, $\mu := \mathbb{E}X_i$ is the mean life time of each bulb, and $S_n := X_1 + \cdots + X_n$ is the time that the $n^{\text{th}}$ – bulb burns out. (We assume the bulbs are replaced immediately on burning out.) By convention, we set $S_0 = 0$.

Let
$$N_t := \sup\{n \ge 0 : S_n \le t\}$$

denote the number of bulbs which have burned out up to time $t$. Since $\mathbb{E}X_i < \infty$, $X_i < \infty$ a.s. and therefore $S_n < \infty$ a.s. for all $n$. From this observation it follows that $N_t \uparrow \infty$ on the set, $\Omega_1 := \cap_{i=1}^{\infty} \{X_i < \infty\}$ – a subset of $\Omega$ with full measure.

It is reasonable to guess that $N_t \sim t/\mu$ and indeed we will show;

$$\lim_{t\uparrow\infty} \frac{1}{t} N_t = \frac{1}{\mu} \text{ a.s.} \quad (22.13)$$

To prove Eq. (22.13), by the SSLN, if $\Omega_0 := \left\{\lim_{n\to\infty} \frac{1}{n} S_n = \mu\right\}$ then $P(\Omega_0) = 1$. From the definition of $N_t$, $S_{N_t} \le t < S_{N_t+1}$ and so

$$\frac{S_{N_t}}{N_t} \le \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t}.$$

For $\omega \in \Omega_0 \cap \Omega_1$ we have

$$\mu = \lim_{t\to\infty} \frac{S_{N_t(\omega)}(\omega)}{N_t(\omega)} \le \liminf_{t\to\infty} \frac{t}{N_t(\omega)} \le \limsup_{t\to\infty} \frac{t}{N_t(\omega)} \le \lim_{t\to\infty}\left[\frac{S_{N_t(\omega)+1}(\omega)}{N_t(\omega)+1} \frac{N_t(\omega)+1}{N_t(\omega)}\right] = \mu.$$

*Example 22.34 (Renewal Theory II).* Let $\{X_i\}_{i=1}^\infty$ be i.i.d. and $\{Y_i\}_{i=1}^\infty$ be i.i.d. non-negative integrable random variables with $\{X_i\}_{i=1}^\infty$ being independent of the $\{Y_i\}_{i=1}^\infty$ and let $\mu = \mathbb{E}X_1$ and $\nu = \mathbb{E}Y_1$. Again assume that $P(X_i > 0) > 0$. We will interpret $Y_i$ to be the amount of time the $i^{\text{th}}$ – bulb remains out after burning out before it is replaced by bulb number $i+1$. Let $R_t$ be the amount of time that we have a working bulb in the time interval $[0,t]$. We are now going to show

$$\lim_{t\uparrow\infty} \frac{1}{t} R_t = \frac{\mathbb{E}X_1}{\mathbb{E}X_1 + \mathbb{E}Y_1} = \frac{\mu}{\mu+\nu}.$$

To prove this, let $S_n := \sum_{i=1}^n (X_i + Y_i)$ be the time that the $n^{\text{th}}$ – bulb is replaced and

$$N_t := \sup\{n \ge 0 : S_n \le t\}$$

denote the number of bulbs which have burned out up to time $n$. By Example 22.33 we know that

$$\lim_{t\uparrow\infty} \frac{1}{t} N_t = \frac{1}{\mu+\nu} \text{ a.s., i.e. } N_t = \frac{1}{\mu+\nu} t + o(t) \text{ a.s.}$$

Let us now set $\tilde{R}_t = \sum_{i=1}^{N_t} X_i$ and observe that

$$\tilde{R}_t \le R_t \le \tilde{R}_t + X_{N_t+1}.$$

By Proposition 22.27 we know that $X_n/n \to 0$ a.s. and therefore,

$$\lim_{t\uparrow\infty} \frac{X_{N_t+1}}{t} = \lim_{t\uparrow\infty}\left[\frac{X_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{t}\right] = 0 \cdot \frac{1}{\mu+\nu} = 0 \text{ a.s.}$$

Thus it follows that $\lim_{t\uparrow\infty} \frac{1}{t} R_t = \lim_{t\uparrow\infty} \frac{1}{t} \tilde{R}_t$ a.s. and the latter limit may be computed using the strong law of large numbers;

$$\frac{1}{t}\tilde{R}_t = \frac{1}{t}\sum_{i=1}^{N_t} X_i = \frac{N_t}{t} \cdot \frac{1}{N_t}\sum_{i=1}^{N_t} X_i \to \frac{1}{\mu+\nu} \cdot \mu \text{ a.s.}$$

**Theorem 22.35 (Glivenko-Cantelli Theorem).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables and $F(x) := P(X_i \le x)$. Further let $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the **empirical distribution** with **empirical distribution function,***

$$F_n(x) := \mu_n((-\infty, x]) = \frac{1}{n}\sum_{i=1}^n 1_{X_i \le x}.$$

*Then*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \text{ } a.s.$$

**Proof.** Since $\{1_{X_i \le x}\}_{i=1}^\infty$ are i.i.d random variables with $\mathbb{E}1_{X_i\le x} = P(X_i \le x) = F(x)$, it follows by the strong law of large numbers that

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ a.s. for all } x \in \mathbb{R}. \tag{22.14}$$

Our goal is to now show that this convergence is uniform.[2] To do this we will use another application of the strong law of large numbers applied to $\{1_{X_i < x}\}$ in order to conclude that, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(x-) = F(x-) \text{ a.s. for all } x \in \mathbb{R}. \tag{22.15}$$

Keep in mind that the exceptional set of probability zero depend on $x$.

Given $k \in \mathbb{N}$, let $\Lambda_k := \left\{\frac{i}{k} : i = 1, 2, \dots, k-1\right\}$ and let $x_i := \inf\{x : F(x) \ge i/k\}$ for $i = 1, 2, \dots, k-1$, see Figure 22.1.Let us further set $x_k = \infty$ and $x_0 = -\infty$ and let $\Omega_k$ denote the subset of $\Omega$ of full measure where Eqs. (22.14) and (22.15) hold for $x \in \{x_i : 1 \le i \le k-1\}$. For $\omega \in \Omega_k$ we may find $N(\omega) \in \mathbb{N}$ ($N$ is random) so that

$$|F_n(x_i) - F(x_i)| < 1/k \text{ and } |F_n(x_i-) - F(x_i-)| < 1/k$$

for $n \ge N(\omega)$, $1 \le i \le k-1$, and $\omega \in \Omega_k$ with $P(\Omega_k) = 1$.

[2] Observation. If $F$ is continuous then, by what we have just shown, there is a set $\Omega_0 \subset \Omega$ such that $P(\Omega_0) = 1$ and on $\Omega_0$, $F_n(r) \to F(r)$ for all $r \in \mathbb{Q}$. Moreover on $\Omega_0$, if $x \in \mathbb{R}$ and $r \le x \le s$ with $r, s \in \mathbb{Q}$, we have

$$F(r) = \lim_{n\to\infty} F_n(r) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} F_n(s) = F(s).$$

We may now let $s \downarrow x$ and $r \uparrow x$ to conclude, on $\Omega_0$, on

$$F(x) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x) \text{ for all } x \in \mathbb{R},$$

i.e. on $\Omega_0$, $\lim_{n\to\infty} F_n(x) = F(x)$. Thus, in this special case we have shown that off a fixed null set independent of $x$ that $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$.

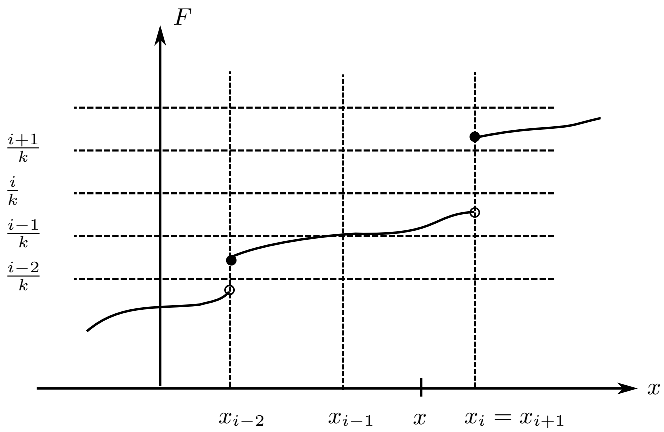



**Fig. 22.1.** Constructing the sequence of points $\{x_i\}_{i=0}^k$.

Observe that it is possible that $x_i = x_{i+1}$ for some of the $i$. This can occur when $F$ has jumps of size greater than $1/k$,[3] see Figure 22.1. Now suppose $i$ has been chosen so that $x_{i-1} < x_i$ and let $x \in (x_{i-1}, x_i)$. We then have for $\omega \in \Omega_k$ and $n \ge N(\omega)$ that

$$F_n(x) \le F_n(x_i-) \le F(x_i-) + 1/k \le F(x) + 2/k$$

and

$$F_n(x) \ge F_n(x_{i-1}) \ge F(x_{i-1}) - 1/k \ge F(x_i-) - 2/k \ge F(x) - 2/k.$$

From this it follows on $\Omega_k$ that $|F(x) - F_n(x)| \le 2/k$ for $n \ge N$ and therefore,

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \le 2/k.$$

Hence it follows on $\Omega_0 := \cap_{k=1}^\infty \Omega_k$ (a set with $P(\Omega_0) = 1$) that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0.$$

■

## 22.4 Kolmogorov's Three Series Theorem

In this section, $\{X_k\}_{k=1}^\infty$ be a collection of independent random variables. The goal of this section is Theorem 22.42 which gives necessary and sufficient conditions for the almost sure convergence of $\sum_{k=1}^\infty X_k$. We first are going to explore the special case where there exists $c < \infty$ such that $|X_k| \le c$ a.s. for all $k \in \mathbb{N}$. The main result here is Corollary 22.40 which states that $\sum_{k=1}^\infty X_k$ is a.s. convergent iff $\sum_{k=1}^\infty \mathbb{E}X_k$ and $\sum_{k=1}^\infty \operatorname{Var}(X_k)$ are convergent in $\mathbb{R}$. Kolmogorov's convergence criteria in Theorem 22.11 gives one direction. The converse direction will be based on the following (sub-martingale) inequality.

[3] In fact if $F(x) = \delta_0((-\infty, x]) = 1_{x\ge 0}$, then $x_1 = \cdots = x_{k-1} = 0$ for all $k$.

**Theorem 22.36.** *Suppose that $S_n := \sum_{k=1}^n X_k$ where $\{X_k\}_{k=1}^\infty$ are independent random variables such that $|X_k| \le c < \infty$ and $\mathbb{E}X_k = 0$ for all $k \in \mathbb{N}$. Then for all $\lambda > 0$,*

$$P\left(\sup_n |S_n| \le \lambda\right) \cdot \sum_{k=1}^\infty \operatorname{Var}(X_k) \le (\lambda + c)^2. \tag{22.16}$$

*In particular if $P(\sup_n |S_n| < \infty) > 0$, then $\sum_{k=1}^\infty \operatorname{Var}(X_k) < \infty$ and $\sum_{k=1}^\infty X_k = \lim_{n\to\infty} S_n$ exists in $\mathbb{R}$ a.s. and in $L^2(P)$.*

**Proof.** Let $S_0 = 0$, $\mathcal{B}_n := \sigma(S_0, \dots, S_n)$, and $\tau := \min\{n \in \mathbb{N}_0 : |S_n| > \lambda\}$ Recall that $\{S_n\}_{n=0}^\infty$ is a martingale and $Y_n := S_n^2$ is a sub-martingale (by conditional Jensen's inequality) whose Doob's decomposition is given by $Y_n = M_n + A_n$ where $A_n = \mathbb{E}S_n^2 = \sum_{k=1}^n \operatorname{Var}(X_k)$, see Examples 20.3 and 20.19. For completeness, here is a proof that $\{M_n := S_n^2 - \mathbb{E}S_n^2\}$ is a martingale;

$$\begin{aligned}
\mathbb{E}_{\mathcal{B}_n}[M_{n+1}] &= \mathbb{E}_{\mathcal{B}_n}\left[S_{n+1}^2 - A_{n+1}\right] = \mathbb{E}_{\mathcal{B}_n}\left[(S_n + X_{n+1})^2\right] - \mathbb{E}S_{n+1}^2 \\
&= S_n^2 + 2S_n\mathbb{E}_{\mathcal{B}_n}X_{n+1} + \mathbb{E}_{\mathcal{B}_n}X_{n+1}^2 - \mathbb{E}S_{n+1}^2 \\
&= S_n^2 + 2S_n\mathbb{E}X_{n+1} + \mathbb{E}X_{n+1}^2 - \mathbb{E}S_{n+1}^2 \\
&= S_n^2 - \mathbb{E}S_n^2 = M_n.
\end{aligned}$$

We now have

$$\begin{aligned}
\mathbb{E}\left[S_{\tau\wedge N}^2\right] &= \mathbb{E}\left[S_N^2 : \tau > N\right] + \mathbb{E}\left[S_\tau^2 : \tau \le N\right] \\
&\le \lambda^2 P(\tau > N) + \mathbb{E}\left[(|S_{\tau-1}| + c)^2 : \tau \le N\right] \\
&\le \lambda^2 P(\tau > N) + (\lambda + c)^2 \lambda P(\tau \le N) \le (\lambda + c)^2.
\end{aligned}$$

On the other hand, by the optional sampling theorem,

$$\mathbb{E}\left[S_{\tau\wedge N}^2\right] = \mathbb{E}[M_{\tau\wedge N}] + \mathbb{E}[A_{\tau\wedge N}] = \mathbb{E}[M_0] + \mathbb{E}[A_{\tau\wedge N}] = \mathbb{E}[A_{\tau\wedge N}].$$

Putting these together shows

$$A_\infty P(\tau = \infty) \le \mathbb{E}[A_\tau] = \lim_{N\uparrow\infty} \mathbb{E}[A_{\tau\wedge N}] \le (\lambda + c)^2$$

which is equivalent to Eq. (22.16).

If we now further assume $P(\sup_n |S_n| < \infty) > 0$ then $P(\sup_n |S_n| \le \lambda) > 0$ for sufficiently large $\lambda$ and it follows from Eq. (22.16) that $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty.$ Kolmogorov's convergence criteria then shows $\sum_{k=1}^{\infty} X_k = \lim_{n\to\infty} S_n$ exists in $\mathbb{R}$ a.s. and in $L^2(P)$. ■

The next theorem gives a natural generalization of this result to general sub-martingales.

**Theorem 22.37.** *Let $\{Y_n\}_{n=0}^{\infty}$ be a non-negative integrable submartingale such $Y_0 = 0$ and having the property that there exists an increasing function $f : [0,\infty) \to [0,\infty)$ such that $Y_{n+1} \le f(Y_n)$ a.s. for all $n \in \mathbb{N}_0$. [Thus we are assuming some uniform control over the jump sizes of the sequence $\{Y_n\}_{n=0}^{\infty}$.] If $Y_n = M_n + A_n$ is the Doob - decomposition of $\{Y_n\}$, then*

$$\mathbb{E}\left[A_\infty : \sup_n Y_n \le \lambda\right] \le \lambda \vee f(\lambda) \tag{22.17}$$

*and in particular $A_\infty < \infty$ a.s. on $\{\sup_n Y_n < \infty\}$.*

**Proof.** Let $\tau := \min\{n \in \mathbb{N}_0 : Y_n > \lambda\}$, then using the optional sampling theorem,

$$\mathbb{E}X_{\tau\wedge N} = \mathbb{E}M_{\tau\wedge N} + \mathbb{E}A_{\tau\wedge N} = \mathbb{E}M_0 + \mathbb{E}A_{\tau\wedge N} = \mathbb{E}A_{\tau\wedge N}.$$

Using $X_N \le \lambda$ on $\{\tau > N\}$ and $X_\tau \le f(X_{\tau-1}) \le f(\lambda)$ on $\{\tau \le N\}$ allows us to find the estimate,

$$\begin{aligned}\mathbb{E}A_{\tau\wedge N} = \mathbb{E}X_{\tau\wedge N} &= \mathbb{E}[X_\tau : \tau \le N] + \mathbb{E}[X_N : \tau > N] \\ &\le f(\lambda) P(\tau \le N) + \lambda P(\tau > N) \le \lambda \vee f(\lambda).\end{aligned}$$

Letting $N \uparrow \infty$ in this estimate the shows

$$\mathbb{E}[A_\infty : \tau = \infty] \le \mathbb{E}A_\tau \le \lambda \vee f(\lambda)$$

which gives Eq. (22.17) since $\{\tau = \infty\} = \{\sup_n Y_n \le \lambda\}$. ■

*Remark 22.38.* The above proof also shows

$$\mathbb{E}\left[A_N : \max_{n\le N} Y_n \le \lambda\right] \le \mathbb{E}A_{\tau\wedge N} \le \lambda \vee f(\lambda) \ \forall\, N \in \mathbb{N}.$$

**Corollary 22.39.** *Suppose that $\{X_n\}_{n=0}^{\infty}$ are independent random variables which are bounded by some $c < \infty$ and have mean zero, $\mathbb{E}X_n = 0$ for all $n$. Then $\sum_{k=1}^{\infty} X_k$ exists a.s. iff $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty$.*

**Proof.** We need only prove the forward direction since Theorem 22.11 proves the converse direction. As usual let $S_0 = 0$ and $S_n = \sum_{k=1}^{n} X_k$ and recall that $\{S_n\}_{n=0}^{\infty}$ is a martingale and $Y_n := S_n^2$ is a sub-martingale (by conditional Jensen's inequality) whose Doob decomposition is given by $Y_n = M_n + A_n$ where $A_n = \mathbb{E}S_n^2 = \sum_{k=1}^{n} \mathrm{Var}(X_k)$, see Examples 20.3 and 20.19.[4] Moreover

$$Y_{n+1} = S_{n+1}^2 \le (|S_n| + c)^2 = \left(\sqrt{Y_n} + c\right)^2 = f(Y_n)$$

where $f(y) := \left(\sqrt{y} + c\right)^2$. Since we are assuming

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} S_n^2 = \left(\sum_{k=1}^{\infty} X_k\right)^2 \text{ exists a.s.,}$$

it follows that $P(\{\sup_n Y_n < \infty\}) = 1$ and so by Theorem 22.37, $A_\infty = \sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty$ a.s. This completes the proof since $A_\infty = \sum_{k=1}^{\infty} \mathrm{Var}(X_k)$ is not random. ■

**Corollary 22.40.** *Suppose that $\{X_n\}_{n=0}^{\infty}$ are independent random variables which are bounded by some $c < \infty$. Then $\sum_{k=1}^{\infty} X_k$ exists a.s. iff $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty$ and $\sum_{k=1}^{\infty} \mathbb{E}X_k$ exists a.s.*

**Proof.** As mentioned at the start of this section, in light of Theorem 22.11, it suffices to prove the forward direction. Our goal is to make use of Corollary 22.39. In order to do this we will use the trick of doubling the probability space. In detail, let $\hat{X}_n : \Omega \times \Omega \to \mathbb{R}$ be defined by

$$\hat{X}_n(\omega, \omega') := X_n(\omega) - X_n(\omega')$$

thought of as random variables on $(\Omega \times \Omega, \mathcal{B} \otimes \mathcal{B}, P \otimes P)$. We then have $\left|\hat{X}_n\right| \le 2c$, $\mathbb{E}\hat{X}_n = 0$, and $\mathrm{Var}\left(\hat{X}_k\right) = 2\,\mathrm{Var}(X_n)$. Moreover if $\sum_{k=1}^{\infty} X_k$ exists $P-$a.s. then $\sum_{k=1}^{\infty} \hat{X}_k$ exists $P \otimes P-$a.s. and so by Theorem 22.36 or Corollary 22.39,

$$\sum_{k=1}^{\infty} \mathrm{Var}(X_k) = \frac{1}{2}\sum_{k=1}^{\infty} \mathrm{Var}\left(\hat{X}_k\right) < \infty.$$

---

[4] Here is again the proof that $\{M_n\}$ is a martingale;

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n}\left[S_{n+1}^2 - S_n^2\right] &= \mathbb{E}_{\mathcal{B}_n}\left[(S_{n+1} + S_n)(S_{n+1} - S_n)\right] \\ &= \mathbb{E}_{\mathcal{B}_n}\left[(S_{n+1} - S_n)(S_{n+1} - S_n)\right] \\ &= \mathbb{E}X_{n+1}^2 = \mathbb{E}S_{n+1}^2 - \mathbb{E}S_n^2.\end{aligned}$$

We may now apply Theorem 22.11 to conclude that $\sum_{k=1}^{\infty}(X_k - \mathbb{E}X_k)$ exists a.s. and therefore

$$\sum_{k=1}^{\infty} \mathbb{E}X_k = \sum_{k=1}^{\infty} \mathbb{E}X_k - \sum_{k=1}^{\infty}(X_k - \mathbb{E}X_k) \text{ exists a.s.}$$

■

*Example 22.41.* Let $\{Z_n\}_{n=1}^{\infty}$ be i.i.d. mean zero random variables, $|Z_n| \le c < \infty$ for some $n$. Further assume $\sigma^2 = \mathbb{E}Z_n^2 = \mathrm{Var}(Z_n) > 0$. We wish to decide for which $0 < p < \infty$ the sum $\sum_{n=1}^{\infty} \frac{1}{n^p} Z_n$ is almost surely convergent. Taking $X_n := \frac{1}{n^p} Z_n$ we have $|X_n| \le c$ and $\mathbb{E}X_n = 0$ for all $n$. Moreover $\mathrm{Var}(X_n) = \frac{1}{n^{2p}}\sigma^2$ and so

$$\sum_{n=1}^{\infty} \mathrm{Var}(X_n) = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^{2p}}$$

which is convergent iff $2p > 1$, i.e. $p > 1/2$. Thus according to Corollary 22.40, $\sum_{n=1}^{\infty} \frac{1}{n^p} Z_n$ is a.s. convergent iff $p > 1/2$. [The same results would hold if we only assumed there exists $c \in (0,\infty)$ such that $c \le \mathbb{E}Z_n^2 \le c^{-1}$ for all $n$.]

We are now ready to state and prove Kolmogorov's Three Series Theorem.

**Theorem 22.42 (Kolmogorov's Three Series Theorem).** *Suppose that $\{X_n\}_{n=1}^{\infty}$ are independent random variables and for $c > 0$ any $c > 0$ let $X_n^c := X_n 1_{|X_n| \le c}$. Then the random series, $\sum_{n=1}^{\infty} X_n$, is almost surely convergent in $\mathbb{R}$ iff there exists $c > 0$ such that following three sums converge;*

1. $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^{\infty} \mathrm{Var}(X_n^c) < \infty$, *and*
3. $\sum_{n=1}^{\infty} \mathbb{E}(X_n^c)$.

*Moreover, if the three series above converge for some $c > 0$ then they converge for all values of $c > 0$.*

**Proof.** Recall that $\{X_n\}_{n=1}^{\infty}$ and $\{X_n^c\}_{n=1}^{\infty}$ are tail equivalent iff $0 = P(X_n \neq X_n^c \text{ i.o.}) = P(|X_n| > c \text{ i.o.})$ which by the Borel Cantelli Lemma happens iff $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$.

($\Longleftarrow$) Suppose that $c > 0$ and the three series in items 1. – 3. converge. Then as above $\{X_n\}_{n=1}^{\infty}$ and $\{X_n^c\}_{n=1}^{\infty}$ are tail equivalent and therefore $\sum_{n=1}^{\infty} X_n$, is almost surely convergent in $\mathbb{R}$ iff $\sum_{n=1}^{\infty} X_n^c$, is almost surely convergent in $\mathbb{R}$ which according to Corollary 22.40 happens iff the sums in items 2. and 3. of the theorem are convergent.

($\Longrightarrow$) Now suppose that $\sum_{n=1}^{\infty} X_n$, is almost surely convergent in $\mathbb{R}$ and $c > 0$ is **any** positive number. Since $\sum_{n=1}^{\infty} X_n$ is convergent it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and from this we conclude that $\{X_n\}_{n=1}^{\infty}$ and $\{X_n^c\}_{n=1}^{\infty}$ are tail equivalent and hence $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$ as mentioned above. Moreover the tail equivalence also implies $\sum_{n=1}^{\infty} X_n^c$ is a.s. surely convergent and therefore the sums in items 2. and 3. of the theorem are convergent as well by Corollary 22.40. [Another proof of this direction may be found in Chapter 25, see Theorem 25.17.] ■



*Remark 22.43.* We have seen another necessary and sufficient condition in Exercise 20.22, namely $\sum_{n=1}^{\infty} X_n$, is almost surely convergent in $\mathbb{R}$ iff $\sum_{n=1}^{\infty} X_n$ is convergent in distribution. We will also see below that $\sum_{n=1}^{\infty} X_n$, is almost surely convergent in $\mathbb{R}$ iff $\sum_{n=1}^{\infty} X_n$, is convergent in probability, see Lévy's Theorem 22.50 below.

**Exercise 22.5 (Two Series Theorem – Resnik 7.15).** Prove that the three series theorem reduces to a two series theorem when the random variables are positive. That is, if $X_n \ge 0$ are independent, then $\sum_n X_n < \infty$ a.s. iff for any $c > 0$ we have

$$\sum_n P(X_n > c) < \infty \text{ and} \tag{22.18}$$

$$\sum_n \mathbb{E}[X_n 1_{X_n \le c}] < \infty, \tag{22.19}$$

that is it is unnecessary to verify the convergence of the second series in Theorem 22.42 involving the variances.

### 22.4.1 Examples

**Lemma 22.44.** *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent square integrable random variables such that $Y_n \overset{d}{=} N(\mu_n, \sigma_n^2)$. Then $\sum_{j=1}^{\infty} Y_j$ converges a.s. iff $\sum_{j=1}^{\infty} \sigma_j^2 < \infty$ and $\sum_{j=1}^{\infty} \mu_j$ converges.*

**Proof.** The implication "$\Longleftarrow$" is true without the assumption that the $Y_n$ are normal random variables as pointed out in Theorem 22.11. To prove the converse directions we will make use of the Kolmogorov's three series Theorem 22.42. Namely, if $\sum_{j=1}^{\infty} Y_j$ converges a.s. then the three series in Theorem 22.42 converge for all $c > 0$.

1. Since $Y_n \overset{d}{=} \sigma_n N + \mu_n$, we have for any $c > 0$ that

$$\infty > \sum_{n=1}^{\infty} P(|\sigma_n N + \mu_n| > c). \tag{22.20}$$

If $\lim_{n\to\infty} \mu_n \neq 0$ then there is a $c > 0$ such that either $\mu_n \ge 2c$ for infinitely many $n$ or $\mu_n \le -2c$ for infinitely many $n$. It then follows that either $\{N > 0\} \subset$

$\{|\sigma_n N + \mu_n| > c\}$ $n$ i.o. or $\{N < 0\} \subset \{|\sigma_n N + \mu_n| > c\}$ $n$ i.o. In either case we would have $P(|\sigma_n N + \mu_n| > c) \geq 1/2$ $n-$i.o. which would violate Eq. (22.20) and so we may concluded that $\lim_{n\to\infty} \mu_n = 0$. Similarly if $\lim_{n\to\infty} \sigma_n \neq 0$, then there exists $\alpha < \infty$ such that

$$\{N \geq \alpha\} \subset \{|\sigma_n N + \mu_n| > 1\} \;\; n - \text{i.o.}$$

which would imply $P(|\sigma_n N + \mu_n| > 1) \geq P(N \geq \alpha) > 0$ for infinitely many $n$. This again violate Eq. (22.20) and thus we may conclude that $\lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \sigma_n = 0$.

2. Let $\chi_n := 1_{|\sigma_n N + \mu_n| \leq c} \in \{0, 1\}$. The convergence of the second series for all $c > 0$ implies

$$\infty > \sum_{n=1}^{\infty} \operatorname{Var}\left(Y_n 1_{|Y_n| \leq c}\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right). \tag{22.21}$$

If we can show

$$\operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right) \geq \frac{1}{2}\sigma_n^2 \text{ for large } n, \tag{22.22}$$

it would then follow from Eq. (22.21) that $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. We may now use Kolmogorov's convergence criteria (Theorem 22.11) to infer that $\sum_{n=1}^{\infty} (Y_n - \mu_n)$ is almost surely convergent which then implies that $\sum_{n=1}^{\infty} \mu_n$ is convergent as $\mu_n = Y_n - (Y_n - \mu_n)$ and $\sum_{n=1}^{\infty} Y_n$ and $\sum_{n=1}^{\infty} (Y_n - \mu_n)$ are both convergent a.s. So to finish the proof we need to prove the estimate in Eq. (22.22).

Let $\alpha_n := \operatorname{Var}(N\chi_n)$ and $\beta_n := P(\chi_n = 1)$ so that $\operatorname{Var}(\chi_n) = \beta_n (1 - \beta_n)$ and

$$\varepsilon_n := \operatorname{Cov}(N\chi_n, \chi_n) = \mathbb{E}[N\chi_n \cdot \chi_n] - \mathbb{E}[N\chi_n]\,\mathbb{E}[\chi_n] = \mathbb{E}[N\chi_n](1 - \beta_n).$$

Therefore, using $\operatorname{Var}(\sigma X + \mu Y) = \sigma^2 \operatorname{Var}(X) + \mu^2 \operatorname{Var}(Y) + 2\sigma\mu \operatorname{Cov}(X, Y)$, we find

$$\begin{aligned}\operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right) &= \operatorname{Var}\left(\sigma_n N \chi_n + \mu_n \chi_n\right) \\ &= \sigma_n^2 \alpha_n + \mu_n^2 \beta_n (1 - \beta_n) + 2\sigma_n \mu_n \varepsilon_n.\end{aligned}$$

Making use of the estimate, $2ab \leq a^2 + b^2$ valid for all $a, b \geq 0$, it follows that

$$\begin{aligned}\operatorname{Var}\left([\sigma_n N + \mu_n] \chi_n\right) &\geq \sigma_n^2 \alpha_n + \mu_n^2 \beta_n (1 - \beta_n) - 2|\varepsilon_n| \sigma_n |\mu_n| \\ &\geq \sigma_n^2 (\alpha_n - |\varepsilon_n|) + \mu_n^2 (\beta_n (1 - \beta_n) - |\varepsilon_n|) \\ &= \sigma_n^2 (\alpha_n - |\varepsilon_n|) + (1 - \beta_n)(\beta_n - |\mathbb{E}[N\chi_n]|)\,\mu_n^2.\end{aligned}$$

This estimate along with the observations that $1 - \beta_n \geq 0$, $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \beta_n = 1$, $\lim_{n\to\infty} \mathbb{E}[N\chi_n] = 0$ (use DCT) and $\lim_{n\to\infty} \varepsilon_n = 0$ easily implies Eq. (22.22).

**An alternative proof** that $\sum_{n=1}^{\infty} \mu_n$ is convergent using the the third series in Theorem 22.42. For all $c > 0$ the third series implies

$$\begin{aligned}&\sum_{n=1}^{\infty} \mathbb{E}\left([\sigma_n N + \mu_n]\, 1_{|\sigma_n N + \mu_n| \leq c}\right) \text{ is convergent, i.e.} \\ &\sum_{n=1}^{\infty} [\sigma_n \delta_n + \mu_n \beta_n] \text{ is convergent.}\end{aligned}$$

where $\delta_n := \mathbb{E}\left(N \cdot 1_{|\sigma_n N + \mu_n| \leq c}\right)$ and $\beta_n := \mathbb{E}\left(1_{|\sigma_n N + \mu_n| \leq c}\right)$. With a little effort one can show,

$$\delta_n \sim e^{-k/\sigma_n^2} \text{ and } 1 - \beta_n \sim e^{-k/\sigma_n^2} \text{ for large } n.$$

Since $e^{-k/\sigma_n^2} \leq C\sigma_n^2$ for large $n$, it follows that $\sum_{n=1}^{\infty} |\sigma_n \delta_n| \leq C \sum_{n=1}^{\infty} \sigma_n^3 < \infty$ so that $\sum_{n=1}^{\infty} \mu_n \beta_n$ is convergent. Moreover,

$$\sum_{n=1}^{\infty} |\mu_n (\beta_n - 1)| \leq C \sum_{n=1}^{\infty} |\mu_n| \sigma_n^2 < \infty$$

and hence

$$\sum_{n=1}^{\infty} \mu_n = \sum_{n=1}^{\infty} \mu_n \beta_n - \sum_{n=1}^{\infty} \mu_n (\beta_n - 1)$$

must also be convergent. ■

*Example 22.45.* As another simple application of Theorem 22.42, let us use it to give a proof of Theorem 22.11. We will apply Theorem 22.42 with $X_n := Y_n - \mathbb{E}Y_n$. We need to then check the three series in the statement of Theorem 22.42 converge. For the first series we have by the Markov inequality,

$$\sum_{n=1}^{\infty} P(|X_n| > c) \leq \sum_{n=1}^{\infty} \frac{1}{c^2} \mathbb{E}|X_n|^2 = \frac{1}{c^2} \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty.$$

For the second series, observe that

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \leq c}\right) \leq \sum_{n=1}^{\infty} \mathbb{E}\left[\left(X_n 1_{|X_n| \leq c}\right)^2\right] \leq \sum_{n=1}^{\infty} \mathbb{E}\left[X_n^2\right] = \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty$$

and we estimate the third series as;

$$\sum_{n=1}^{\infty} \left|\mathbb{E}\left(X_n 1_{|X_n| \leq c}\right)\right| \leq \sum_{n=1}^{\infty} \mathbb{E}\left(\frac{1}{c} |X_n|^2 1_{|X_n| \leq c}\right) \leq \frac{1}{c} \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty.$$

## 22.5 Maximal Inequalities

**Theorem 22.46 (Kolmogorov's Inequality).** *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, $S_n := X_1 + \cdots + X_n$, and $S_n^* = \max_{j \le n} |S_j|$. Then for any $\alpha > 0$ we have*

$$P(S_N^* \ge \alpha) \le \frac{1}{\alpha^2} \mathbb{E}\left[S_N^2 : S_N^* \ge \alpha\right]. \tag{22.23}$$

**Proof. First proof.** As $\{S_n\}_{n=1}^\infty$ is a martingale relative to the filtration, $\mathcal{B}_n = \sigma(S_1, \dots, S_n)$, the inequality in Eq. (22.23) is a special case of Proposition 20.43 with $X_n = S_n^2$, also see Example 20.49.

***Second direct proof.** Let $\tau = \inf\{j : |S_j| \ge \alpha\}$ with the infimum of the empty set being taken to be equal to $\infty$. Observe that

$$\{\tau = j\} = \{|S_1| < \alpha, \dots, |S_{j-1}| < \alpha, |S_j| \ge \alpha\} \in \sigma(X_1, \dots, X_j).$$

Now

$$\begin{aligned}
\mathbb{E}\left[S_N^2 : |S_N^*| > \alpha\right] &= \mathbb{E}\left[S_N^2 : \tau \le N\right] = \sum_{j=1}^N \mathbb{E}\left[S_N^2 : \tau = j\right] \\
&= \sum_{j=1}^N \mathbb{E}\left[(S_j + S_N - S_j)^2 : \tau = j\right] \\
&= \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 + 2S_j(S_N - S_j) : \tau = j\right] \\
&\overset{(*)}{=} \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 : \tau = j\right] \\
&\ge \sum_{j=1}^N \mathbb{E}\left[S_j^2 : \tau = j\right] \ge \alpha^2 \sum_{j=1}^N P[\tau = j] = \alpha^2 P(|S_N^*| > \alpha).
\end{aligned}$$

The equality, $(*)$, is a consequence of the observations: 1) $1_{\tau=j} S_j$ is $\sigma(X_1, \dots, X_j)$ – measurable, 2) $(S_n - S_j)$ is $\sigma(X_{j+1}, \dots, X_n)$ – measurable and hence $1_{\tau=j} S_j$ and $(S_n - S_j)$ are independent, and so 3)

$$\begin{aligned}
\mathbb{E}\left[S_j(S_N - S_j) : \tau = j\right] &= \mathbb{E}\left[S_j 1_{\tau=j}(S_N - S_j)\right] \\
&= \mathbb{E}\left[S_j 1_{\tau=j}\right] \cdot \mathbb{E}\left[S_N - S_j\right] = \mathbb{E}\left[S_j 1_{\tau=j}\right] \cdot 0 = 0.
\end{aligned}$$

■

*Remark 22.47 (Another proof of Theorem 22.11).* Suppose that $\{Y_j\}_{j=1}^\infty$ are independent random variables such that $\sum_{j=1}^\infty \mathrm{Var}(Y_j) < \infty$ and let $S_n := \sum_{j=1}^n X_j$ where $X_j := Y_j - \mathbb{E}Y_j$. According to Kolmogorov's inequality, Theorem 22.46, for all $M < N$,

$$\begin{aligned}
P\left(\max_{M \le j \le N} |S_j - S_M| \ge \alpha\right) &\le \frac{1}{\alpha^2} \mathbb{E}\left[(S_N - S_M)^2\right] = \frac{1}{\alpha^2} \sum_{j=M+1}^N \mathbb{E}\left[X_j^2\right] \\
&= \frac{1}{\alpha^2} \sum_{j=M+1}^N \mathrm{Var}(X_j).
\end{aligned}$$

Letting $N \to \infty$ in this inequality shows, with $Q_M := \sup_{j \ge M} |S_j - S_M|$,

$$P(Q_M \ge \alpha) \le \frac{1}{\alpha^2} \sum_{j=M+1}^\infty \mathrm{Var}(X_j).$$

Since

$$\delta_M := \sup_{j,k \ge M} |S_j - S_k| \le \sup_{j,k \ge M} \left[|S_j - S_M| + |S_M - S_k|\right] \le 2Q_M$$

we may further conclude,

$$P(\delta_M \ge 2\alpha) \le \frac{1}{\alpha^2} \sum_{j=M+1}^\infty \mathrm{Var}(X_j) \to 0 \text{ as } M \to \infty,$$

i.e. $\delta_M \overset{P}{\to} 0$ as $M \to \infty$. Since $\delta_M$ is decreasing in $M$, it follows that $\lim_{M \to \infty} \delta_M =: \delta$ exists and because $\delta_M \overset{P}{\to} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n \to \infty} |S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^\infty$ is almost surely Cauchy and hence almost surely convergent. This gives a second proof of Kolmogorov's convergence criteria in Theorem 22.11.

**Corollary 22.48 ($L^2$ – SSLN).** *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, and $\sigma^2 = \mathbb{E}X_n^2 < \infty$. Letting $S_n = \sum_{k=1}^n X_k$ and $p > 1/2$, we have*

$$\frac{1}{n^p} S_n \to 0 \ a.s.$$

*If $\{Y_n\}$ is a sequence of independent random variables $\mathbb{E}Y_n = \mu$ and $\sigma^2 =$* $\mathrm{Var}(X_n) < \infty$, *then for any $\beta \in (0, 1/2)$,*

$$\frac{1}{n} \sum_{k=1}^n Y_k - \mu = O\left(\frac{1}{n^\beta}\right).$$

**Proof.** (The proof of this Corollary may be skipped as it has already been proved, see Corollary 22.15.) From Theorem 22.46, we have for every $\varepsilon > 0$ that

$$P\left(\frac{S_N^*}{N^p} \geq \varepsilon\right) = P\left(S_N^* \geq \varepsilon N^p\right) \leq \frac{1}{\varepsilon^2 N^{2p}} \mathbb{E}\left[S_N^2\right] = \frac{1}{\varepsilon^2 N^{2p}} CN = \frac{C}{\varepsilon^2 N^{(2p-1)}}.$$

Hence if we suppose that $N_n = n^\alpha$ with $\alpha(2p-1) > 1$, then we have

$$\sum_{n=1}^\infty P\left(\frac{S_{N_n}^*}{N_n^p} \geq \varepsilon\right) \leq \sum_{n=1}^\infty \frac{C}{\varepsilon^2 n^{\alpha(2p-1)}} < \infty$$

and so by the first Borel – Cantelli lemma we have

$$P\left(\left\{\frac{S_{N_n}^*}{N_n^p} \geq \varepsilon \text{ for } n \text{ i.o.}\right\}\right) = 0.$$

From this it follows that $\lim_{n\to\infty} \frac{S_{N_n}^*}{N_n^p} = 0$ a.s.

To finish the proof, for $m \in \mathbb{N}$, we may choose $n = n(m)$ such that

$$n^\alpha = N_n \leq m < N_{n+1} = (n+1)^\alpha.$$

Since

$$\frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \leq \frac{S_m^*}{m^p} \leq \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p}$$

and

$$N_{n+1}/N_n \to 1 \text{ as } n \to \infty,$$

it follows that

$$\begin{aligned} 0 = \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)}^p} &= \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \leq \lim_{m\to\infty} \frac{S_m^*}{m^p} \\ &\leq \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p} = \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)+1}^p} = 0 \text{ a.s.} \end{aligned}$$

That is $\lim_{m\to\infty} \frac{S_m^*}{m^p} = 0$ a.s. ■

We are going to give three more maximal inequalities before ending this section. In all case we will start with $\{X_n\}_{n=1}^\infty$ a sequence of (possibly with values in a separable Banach space, $Y$) random variables and we will let $S_n := \sum_{k\leq n} X_k$ and $S_n^* := \max_{k\leq n} \|S_k\|$. If $\tau$ is any $\mathcal{B}_n^X$ – stopping time and $f$ is a non-negative function on $Y$, then

$$\mathbb{E}\left[f\left(S_n - S_\tau\right) : \tau \leq n\right] = \sum_{k=1}^n \mathbb{E}\left[f\left(S_n - S_k\right) : \tau = k\right] = \sum_{k=1}^n \mathbb{E}\left[f\left(S_n - S_k\right)\right] \cdot P\left(\tau = k\right). \tag{22.24}$$

**Theorem 22.49 (Skorohod's Inequality).** *Suppose that $\{X_n\}_{n=1}^\infty$ are independent real or Banach valued random variables. Then for all $\alpha > 0$ we have*

$$P\left(\|S_N\| \geq \alpha\right) \geq \left(1 - c_N(\alpha)\right) P\left(S_N^* \geq 2\alpha\right) \text{ and} \tag{22.25}$$

$$P\left(\|S_N\| > \alpha\right) \geq \left(1 - c_N(\alpha)\right) P\left(S_N^* > 2\alpha\right) \tag{22.26}$$

*where*

$$c_N(\alpha) := \max_{1\leq k\leq N} P\left(\|S_N - S_k\| > \alpha\right). \tag{22.27}$$

**Proof.** We only prove Eq. (22.25) since the proof of Eq. (22.26) is similar and in fact can be deduced from Eq. (22.25) by a simple limiting argument. If $\tau = \inf\{n : \|S_n\| \geq 2\alpha\}$, then $\{\tau \leq N\} = \{S_N^* \geq 2\alpha\}$ and on this set,

$$\begin{aligned} \|S_N\| &= \|S_\tau + S_N - S_\tau\| \geq \|S_\tau\| - \|S_N - S_\tau\| \\ &\geq 2\alpha - \|S_N - S_\tau\|. \end{aligned}$$

From this it follows that

$$\{\tau \leq N \ \& \ \|S_N - S_\tau\| \leq \alpha\} \subset \{\|S_N\| \geq \alpha\}$$

and therefore,

$$\begin{aligned} P\left(\|S_N\| \geq \alpha\right) &\geq P\left(\tau \leq N \ \& \ \|S_N - S_\tau\| \leq \alpha\right) \\ &= \sum_{k=1}^N P\left(\tau = k\right) \cdot P\left(\|S_N - S_k\| \leq \alpha\right) \\ &\geq \min_{1\leq k\leq N} P\left(\|S_N - S_k\| \leq \alpha\right) \cdot \sum_{k=1}^N P\left(\tau = k\right) \\ &= \min_{1\leq k\leq N} \left[1 - P\left(\|S_N - S_k\| > \alpha\right)\right] \cdot \sum_{k=1}^N P\left(\tau = k\right) \\ &= \left[1 - c_N(\alpha)\right] \cdot P\left(S_N^* \geq 2\alpha\right). \end{aligned}$$

■

As an application of Theorem 22.49 we have the following convergence result.

**Theorem 22.50 (Lévy's Theorem).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables then $\sum_{n=1}^\infty X_n$ converges in probability iff $\sum_{n=1}^\infty X_n$ converges a.s.*

**Proof.** Let $S_n := \sum_{k=1}^n X_k$. Since almost sure convergence implies convergence in probability, it suffices to show; if $S_n$ is convergent in probability then $S_n$ is almost surely convergent. Given $M \in \mathbb{M}$, let $Q_M := \sup_{n\geq M} |S_n - S_M|$ and for $M < N$, let $Q_{M,N} := \sup_{M\leq n\leq N} |S_n - S_M|$. Given $\varepsilon \in (0,1)$, by assumption, there exists $M = M(\varepsilon) \in \mathbb{N}$ such that $\max_{M\leq j\leq N} P\left(|S_N - S_j| > \varepsilon\right) < \varepsilon$

for all $N \geq M$. An application of Skorohod's inequality (Theorem 22.49), then shows

$$P(Q_{M,N} \geq 2\varepsilon) \leq \frac{P(|S_N - S_M| > \varepsilon)}{(1 - \max_{M \leq j \leq N} P(|S_N - S_j| > \varepsilon))} \leq \frac{\varepsilon}{1-\varepsilon}.$$

Since $Q_{M,N} \uparrow Q_M$ as $N \to \infty$, we may conclude

$$P(Q_M \geq 2\varepsilon) \leq \frac{\varepsilon}{1-\varepsilon}.$$

Since,

$$\delta_M := \sup_{m,n \geq M} |S_n - S_m| \leq \sup_{m,n \geq M} [|S_n - S_M| + |S_M - S_m|] = 2Q_M$$

we may further conclude, $P(\delta_M > 4\varepsilon) \leq \frac{\varepsilon}{1-\varepsilon}$ and since $\varepsilon > 0$ is arbitrary, it follows that $\delta_M \xrightarrow{P} 0$ as $M \to \infty$. Moreover, since $\delta_M$ is decreasing in $M$, it follows that $\lim_{M\to\infty} \delta_M =: \delta$ exists and because $\delta_M \xrightarrow{P} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n\to\infty} |S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^{\infty}$ is almost surely Cauchy and hence almost surely convergent. ∎

*Remark 22.51 (Yet another proof of Theorem 22.11).* Suppose that $\{Y_j\}_{j=1}^{\infty}$ are independent random variables such that $\sum_{j=1}^{\infty} \mathrm{Var}(Y_j) < \infty$. By Proposition 22.10, the sum, $\sum_{j=1}^{\infty} (Y_j - \mathbb{E}Y_j)$, is $L^2(P)$ convergent and hence convergent in probability. An application of Lévy's Theorem 22.50 then shows $\sum_{j=1}^{\infty} (Y_j - \mathbb{E}Y_j)$ is almost surely convergent which gives another proof of Kolmogorov's convergence criteria in Theorem 22.11.

The next maximal inequality will be useful later in proving the "functional central limit theorem." It is actually a simple corollary of Skorohod's inequality (Theorem 22.49) along with Chebyshev's inequality.

**Corollary 22.52 (Ottaviani's maximal ineqaulity).** *Suppose that $\{X_n\}_{n=1}^{\infty}$ are independent real or Banach valued square integrable random variables. Then for all $\alpha > 0$ we have*

$$P(\|S_N\| \geq \alpha) \geq \left(1 - \frac{1}{\alpha^2} \max_{1 \leq k \leq N} \mathbb{E}\|S_N - S_k\|^2\right) P(S_N^* \geq 2\alpha)$$

*and in particular if $\alpha^2 > \max_{1 \leq k \leq N} \mathbb{E}\|S_N - S_k\|^2$, then*

$$P(S_N^* \geq 2\alpha) \leq \left(1 - \frac{1}{\alpha^2} \max_{1 \leq k \leq N} \mathbb{E}\|S_N - S_k\|^2\right)^{-1} P(\|S_N\| \geq \alpha).$$

*If we further assume that $\{X_n\}$ are real (or Hilbert valued) mean zero random variables, then*

$$P(S_N^* \geq 2\alpha) \leq \left(1 - \frac{1}{\alpha^2} \mathbb{E}\|S_N - X_1\|^2\right)^{-1} P(\|S_N\| \geq \alpha). \tag{22.28}$$

**Proof.** The first and second inequalities follow by Chebyshev's inequality and Skorohod's Theorem 22.49. When the $\{X_n\}$ are real or Hilbert valued mean zero square integrable random variables, we have

$$\max_{1 \leq k \leq N} \mathbb{E}\|S_N - S_k\|^2 = \max_{1 \leq k \leq N} \sum_{j=k+1}^{N} \mathbb{E}\|X_j\|^2 = \sum_{j=2}^{N} \mathbb{E}\|X_j\|^2 = \|S_N - X_1\|^2.$$

∎

**Corollary 22.53.** *Suppose $\lambda > 1$ and $\{X_n\}_{n=1}^{\infty}$ are independent real square integrable random variables with $\mathbb{E}X_n = 0$ and $\mathrm{Var}(X_n) = 1$ for all $n$. Then*

$$P\left(S_n^* \geq 2\lambda\sqrt{n}\right) \leq \left(1 - \frac{1}{\lambda^2}\right)^{-1} \cdot P\left(|S_n| \geq \lambda\sqrt{n}\right)$$

*and if we further assume that $\{X_n\}_{n=1}^{\infty}$ are i.i.d., then*

$$\lim_{n\to\infty} P\left(S_n^* \geq 2\lambda\sqrt{n}\right) \leq \sqrt{\frac{2}{\pi}} \left(1 - \frac{1}{\lambda^2}\right)^{-1} \frac{1}{\lambda} e^{-\lambda^2/2}.$$

**Proof.** The first inequality follows from Eq. (22.28) of Corollary 22.52 with $\alpha = \lambda\sqrt{n}$. For the second inequality we use the central limit theorem to conclude that

$$P\left(|S_n| \geq \lambda\sqrt{n}\right) = P\left(\frac{|S_n|}{\sqrt{n}} \geq \lambda\right) \to P(|Z| \geq \lambda)$$

where $Z$ is a standard normal random variable. We then estimate $P(|Z| \geq \lambda)$ using the Gaussian tail estimates in Lemma 9.71. ∎

We can significantly improve on Corollary 22.52 if we further assume that $X_n$ is symmetric for $n$ in which case the following reflection principle holds.

**Theorem 22.54.** *Suppose that $\{X_n\}_{n=1}^{\infty}$ are independent real or Banach valued random variables such that $X_n \overset{d}{=} -X_n$ for all $n$ and $\tau$ is any $\{\mathcal{B}_n^X = \sigma(X_1, \dots, X_n)\}_{n=1}^{\infty}$ stopping time. If we set*

$$S_n^\tau := 1_{n\le\tau} S_n - 1_{n>\tau}(S_n - S_\tau)$$
$$= 1_{n\le\tau} S_n - 1_{n>\tau} \sum_{\tau<k\le n} X_k,$$

*then* $(\{S_n\}_{n=1}^\infty, \tau)$ *and* $(\{S_n^\tau\}_{n=1}^\infty, \tau)$ *have the same distribution. (Notice* $S_n^\tau = S_n$ *for* $n \le \tau$ *and* $S_n^\tau$ *is* $S_n$ *"reflected about* $S_\tau$*" for* $n > \tau$*.)*

**Proof.** Let $N \in \mathbb{N}$ be given and $f : S^N \to \mathbb{R}$ be a bounded measurable function. Then, for all $k \le N$ we have,

$$\mathbb{E}[f(S_1^\tau, \dots, S_N^\tau) : \tau = k]$$
$$= \mathbb{E}[f(S_1, \dots, S_k, (S_k - X_{k+1}), \dots, (S_k - X_{k+1} - \dots - X_N)) : \tau = k]$$
$$= \mathbb{E}[f(S_1, \dots, S_k, (S_k + X_{k+1}), \dots, (S_k + X_{k+1} + \dots + X_N)) : \tau = k]$$
$$= \mathbb{E}[f(S_1, \dots, S_N) : \tau = k],$$
$$=$$

wherein we have used $(X_{k+1}, \dots, X_N) \overset{d}{=} -(X_{k+1}, \dots, X_N)$, $(X_{k+1}, \dots, X_N)$ is independent of $\mathcal{B}_k^X$, and $\{\tau = k\}$ and $(S_1, \dots, S_k)$ are $\mathcal{B}_k^X$ – measurable. This completes the proof since on $\{\tau = \infty\}$, $\{S_n\}_{n=1}^\infty = \{S_n^\tau\}_{n=1}^\infty$. ■

In order to exploit this principle we will need to combine it with the following simple **geometric reflection property** for Banach spaces; if $r > 0$ and $x, y \in Y$ ($Y$ is a normed space) such that $\|x\| \ge r$ while $\|x - y\| < r$, then $\|x + y\| > r$. This is easy to believe (draw the picture for $Y = \mathbb{R}^2$) and it is also easy to prove;

$$\|x + y\| = \|2x - (x - y)\|$$
$$\ge \|2x\| - \|x - y\|$$
$$\ge 2r - \|x - y\| > 2r - r = r.$$

**Proposition 22.55 (Reflection Principle).** *Suppose that* $\{X_n\}_{n=1}^\infty$ *are independent real or Banach valued random variables such that* $X_n \overset{d}{=} -X_n$ *for all* $n$*. Then*

$$P(S_N^* \ge r) \le P(\|S_N\| \ge r) + P(\|S_N\| > r) \le 2P(\|S_N\| \ge r). \tag{22.29}$$

**Proof.** Let $\tau := \inf\{n : \|S_n\| \ge r\}$ (a $\{\mathcal{B}_n^X\}_{n=1}^\infty$ – stopping time), then

$$P(S_N^* \ge r) = P(S_N^* \ge r, \|S_N\| \ge r) + P(S_N^* \ge r, \|S_N\| < r)$$
$$= P(\|S_N\| \ge r) + P(\tau \le N, \|S_N\| < r).$$

Moreover by the reflection principle (Theorem 22.54),

$$P(\tau \le N, \|S_N\| < r) = P(\tau \le N, \|S_N^\tau\| < r)$$
$$= P(\tau \le N, \|S_\tau - (S_N - S_\tau)\| < r).$$

If $\|S_\tau\| \ge r$ and $\|S_\tau - (S_N - S_\tau)\| < r$, then by the geometric reflection property, $\|S_N\| = \|S_\tau + (S_N - S_\tau)\| > r$ and therefore

$$P(\tau \le N, \|S_\tau - (S_N - S_\tau)\| < r) \ge P(\tau \le N, \|S_N\| > r) = P(\|S_N\| > r).$$

Combining this inequality with the first displayed inequality in the proof easily gives the result.

■

**Exercise 22.6 (Simple Random Walk Reflection principle ).** Let $\{X_n\}_{n=1}^\infty$ be i.i.d Bernoulli random variables with $P(X_n = \pm 1) = \frac{1}{2}$ for all $n$ and let $S_n := \sum_{k\le n} X_k$ be the standard simple random walk on $\mathbb{Z}$. Show for every $r \in \mathbb{N}$ that

$$P\left(\max_{k\le n} S_k \ge r\right) = P(S_n \ge r) + P(S_n > r).$$

## 22.6 Bone Yards

### 22.6.1 Kronecker's Lemma

*Remark 22.56.* Here is a continuous version of Lemma 22.13. If $a(s) \in (0, \infty)$ and $x(s) \in \mathbb{R}$ are continuous functions such that $a(s) \uparrow \infty$ as $s \to \infty$ and $\int_1^\infty \frac{x(s)}{a(s)} ds$ exists, then

$$\lim_{n\to\infty} \frac{1}{a(n)} \int_1^n x(s)\, ds = 0.$$

To prove this let $X(s) := \int_0^s x(u)\, du$ and

$$r(s) := \int_s^\infty \frac{X'(u)}{a(u)} du = \int_s^\infty \frac{x(u)}{a(u)} du.$$

Then by assumption, $r(s) \to 0$ as $s \to 0$ and $X'(s) = -a(s) r'(s)$. Integrating this equation shows

$$X(s) - X(s_0) = -\int_{s_0}^s a(u) r'(u)\, du = -a(u) r(u)|_{u=s_0}^s + \int_{s_0}^s r(u) a'(u)\, du.$$

Dividing this equation by $a(s)$ and then letting $s \to \infty$ gives

$$\limsup_{s\to\infty} \frac{|X(s)|}{a(s)} = \limsup_{s\to\infty} \left[\frac{a(s_0) r(s_0) - a(s) r(s)}{a(s)} + \frac{1}{a(s)} \int_{s_0}^s r(u) a'(u)\, du\right]$$
$$\le \limsup_{s\to\infty} \left[-r(s) + \frac{1}{a(s)} \int_{s_0}^s |r(u)|\, a'(u)\, du\right]$$
$$\le \limsup_{s\to\infty} \left[\frac{a(s) - a(s_0)}{a(s)} \sup_{u\ge s_0} |r(u)|\right] = \sup_{u\ge s_0} |r(u)| \to 0 \text{ as } s_0 \to \infty.$$

**Corollary 22.57.** *Let $\{X_n\}$ be a sequence of independent square integrable random variables and $b_n$ be a sequence such that $b_n \uparrow \infty$. If*

$$\sum_{k=1}^{\infty} \frac{\operatorname{Var}(X_k)}{b_k^2} < \infty$$

*then*

$$\frac{S_n - \mathbb{E}S_n}{b_n} \to 0 \text{ a.s. and in } L^2(P).$$

**Proof.** By Kolmogorov's convergence criteria, Theorem 22.11,

$$\sum_{k=1}^{\infty} \frac{X_k - \mathbb{E}X_k}{b_k} \text{ is convergent a.s. and in } L^2(P).$$

Therefore an application of Kronecker's Lemma 22.13 implies

$$0 = \lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} (X_k - \mathbb{E}X_k) = \lim_{n\to\infty} \frac{S_n - \mathbb{E}S_n}{b_n} \text{ a.s.}$$

Similarly by Kronecker's Lemma 22.13 we know that

$$0 = \lim_{n\to\infty} \frac{1}{b_n^2} \sum_{k=1}^{n} \operatorname{Var}(X_k) = \lim_{n\to\infty} \mathbb{E}\left(\frac{S_n - \mathbb{E}S_n}{b_n}\right)^2$$

which gives the $L^2(P)$ – convergence statement as well. ■

### 22.6.2 Older variants on the proof of Kolmogorov's three series Theorem

[The reader should skip this section.

**Lemma 22.58.** *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent random variables such that there exists $c < \infty$ such that $|Y_n| \le c < \infty$ a.s. and further assume $\mathbb{E}Y_n = 0$. If $\sum_{n=1}^{\infty} Y_n$ is almost surely convergent in $\mathbb{R}$ then $\sum_{n=1}^{\infty} \mathbb{E}Y_n^2 < \infty$. More precisely the following estimate holds,*

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 \le \frac{(\lambda + c)^2}{P(\sup_n |S_n| \le \lambda)} \text{ for all } \lambda > 0, \tag{22.30}$$

*where as usual, $S_n := \sum_{j=1}^{n} Y_j$.*

*Remark 22.59.* It follows from Eq. (22.30) that if $P(\sup_n |S_n| < \infty) > 0$, then $\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 < \infty$ and hence by Kolmogorov's convergence criteria (Theorem 22.11), $\sum_{j=1}^{\infty} Y_j = \lim_{n\to\infty} S_n$ exists a.s. and in particular, $P(\sup_n |S_n| < \infty) = 1$. This also follows from the fact that $\sup_n |S_n| < \infty$ is a tail event and hence $P(\sup_n |S_n| < \infty)$ is either 0 or 1 and as $P(\sup_n |S_n| < \infty) > 0$ we must have $P(\sup_n |S_n| < \infty) = 1$.

**Proof.** We will begin by proving that for every $N \in \mathbb{N}$ and $\lambda > 0$ that

$$\mathbb{E}\left[S_N^2\right] \le \frac{(\lambda + c)^2}{P\left(\sup_{n\le N} |S_n| \le \lambda\right)} \le \frac{(\lambda + c)^2}{P(\sup_n |S_n| \le \lambda)}. \tag{22.31}$$

To prove Eq. (22.31), let $S_0 := 0$, $\mathcal{B}_n := \sigma(S_0, \dots, S_n)$, and $\tau$ be the stopping time,

$$\tau = \tau_\lambda := \inf\{n \ge 1 : |S_n| > \lambda\},$$

where $\inf \emptyset = \infty$. Recall that $\{S_n\}_{n=0}^{\infty}$ and $\left\{M_n := S_n^2 - \mathbb{E}S_n^2\right\}_{n=0}^{\infty}$ are martingales[5], see Examples 20.3 and 20.19. Simple estimates along with Theorem 20.40 then shows

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &= \mathbb{E}\left[S_N^2 : \tau \le N\right] + \mathbb{E}\left[S_N^2 : \tau > N\right] \\
&\le \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_\tau} S_N^2 : \tau \le N\right] + \lambda^2 P(\tau > N) \\
&= \mathbb{E}\left[M_\tau + \mathbb{E}\left[S_N^2\right] : \tau \le N\right] + \lambda^2 P(\tau > N) \\
&= \mathbb{E}\left[M_\tau : \tau \le N\right] + \mathbb{E}\left[S_N^2\right] \cdot P(\tau \le N) + \lambda^2 P(\tau > N),
\end{aligned}$$

or equivalently that

$$\mathbb{E}\left[S_N^2\right] \cdot P(\tau > N) \le \mathbb{E}\left[M_\tau : \tau \le N\right] + \lambda^2 P(\tau > N).$$

As $M_n \le S_n^2$ and so $M_\tau \le S_\tau^2 \le (\lambda + c)^2$ on $\{\tau < \infty\} \supset \{\tau \le N\}$, we learn that

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &\le \frac{(\lambda + c)^2 P(\tau \le N) + \lambda^2 P(\tau > N)}{P(\tau > N)} \\
&\le \frac{(\lambda + c)^2}{P(\tau > N)} = \frac{(\lambda + c)^2}{P\left(\sup_{n\le N} |S_n| \le \lambda\right)} \\
&\le \frac{(\lambda + c)^2}{P(\sup_n |S_n| \le \lambda)}
\end{aligned}$$

[5] Here is the proof for $\{M_n\}$ again;

$$\begin{aligned}
\mathbb{E}_{\mathcal{B}_n}\left[S_{n+1}^2 - S_n^2\right] &= \mathbb{E}_{\mathcal{B}_n}\left[(S_{n+1} + S_n)(S_{n+1} - S_n)\right] \\
&= \mathbb{E}_{\mathcal{B}_n}\left[(S_{n+1} - S_n)(S_{n+1} - S_n)\right] \\
&= \mathbb{E}Y_{n+1}^2 = \mathbb{E}S_{n+1}^2 - \mathbb{E}S_n^2.
\end{aligned}$$

which proves Eq. (22.31).

Since $S_n$ is convergent a.s., it follows that $P(\sup_n |S_n| < \infty) = 1$ and therefore,

$$\lim_{\lambda\uparrow\infty} P\left(\sup_n |S_n| \le \lambda\right) = 1.$$

Hence for $\lambda$ sufficiently large, $P(\sup_n |S_n| \le \lambda) > 0$ and we learn from Eq. (22.31) that

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 = \lim_{N\to\infty} \mathbb{E}\left[S_N^2\right] \le \frac{(\lambda+c)^2}{P(\sup_n |S_n| \le \lambda)} < \infty.$$

■

*Remark 22.60.* For those skipping the martingale section here is another way to estimate $\mathbb{E}\left[S_N^2 : \tau \le N\right]$;

$$\begin{aligned}
\mathbb{E}\left[S_N^2 : \tau \le N\right] &= \sum_{j=1}^{N} \mathbb{E}\left[S_N^2 : \tau = j\right] = \sum_{j=1}^{N} \mathbb{E}\left[\left|S_j + S_N - S_j\right|^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 + 2S_j(S_N - S_j) + (S_N - S_j)^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 : \tau = j\right] + \sum_{j=1}^{N} \mathbb{E}\left[(S_N - S_j)^2\right] P[\tau = j] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[(S_{j-1} + Y_j)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] \sum_{j=1}^{N} P[\tau = j] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[(\lambda + c)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] P[\tau \le N] \\
&= \left[(\lambda + c)^2 + \mathbb{E}\left[S_N^2\right]\right] P[\tau \le N].
\end{aligned}$$

**Lemma 22.61.** *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent random variables such that there exists $c < \infty$ such that $|Y_n| \le c$ a.s. for all $n$. If $\sum_{n=1}^{\infty} Y_n$ converges in $\mathbb{R}$ a.s. then $\sum_{n=1}^{\infty} \mathbb{E}Y_n$ converges as well.*

**Proof.** Let $(\Omega_0, \mathcal{B}_0, P_0)$ be the probability space that $\{Y_n\}_{n=1}^{\infty}$ is defined on and let

$$\Omega := \Omega_0 \times \Omega_0,\ \mathcal{B} := \mathcal{B}_0 \otimes \mathcal{B}_0, \text{ and } P := P_0 \otimes P_0.$$

Further let $Y_n'(\omega_1, \omega_2) := Y_n(\omega_1)$ and $Y_n''(\omega_1, \omega_2) := Y_n(\omega_2)$ and

$$Z_n(\omega_1, \omega_2) := Y_n'(\omega_1, \omega_2) - Y_n''(\omega_1, \omega_2) = Y_n(\omega_1) - Y_n(\omega_2).$$

Then $|Z_n| \le 2c$ a.s., $\mathbb{E}Z_n = 0$, and

$$\sum_{n=1}^{\infty} Z_n(\omega_1, \omega_2) = \sum_{n=1}^{\infty} Y_n(\omega_1) - \sum_{n=1}^{\infty} Y_n(\omega_2) \text{ exists}$$

for $P$ a.e. $(\omega_1, \omega_2)$. Hence it follows from Lemma 22.58 that

$$\begin{aligned}
\infty > \sum_{n=1}^{\infty} \mathbb{E}Z_n^2 &= \sum_{n=1}^{\infty} \operatorname{Var}(Z_n) = \sum_{n=1}^{\infty} \operatorname{Var}(Y_n' - Y_n'') \\
&= \sum_{n=1}^{\infty} \left[\operatorname{Var}(Y_n') + \operatorname{Var}(Y_n'')\right] = 2\sum_{n=1}^{\infty} \operatorname{Var}(Y_n).
\end{aligned}$$

Thus by Kolmogorov's convergence theorem, it follows that $\sum_{n=1}^{\infty}(Y_n - \mathbb{E}Y_n)$ is convergent. Since $\sum_{n=1}^{\infty} Y_n$ is a.s. convergent, we may conclude that $\sum_{n=1}^{\infty} \mathbb{E}Y_n$ is also convergent. ■

We are now ready to complete the proof of Theorem 22.42.

**Proof of Theorem 22.42.** Our goal is to show if $\{X_n\}_{n=1}^{\infty}$ are independent random variables such that $\sum_{n=1}^{\infty} X_n$, is almost surely convergent then for all $c > 0$ the following three series converge;

1. $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$, and
3. $\sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ converges.

Since $\sum_{n=1}^{\infty} X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and hence for every $c > 0$, $P(\{|X_n| \ge c \text{ i.o.}\}) = 0$. According the Borel zero one law (Lemma 12.44) this implies for every $c > 0$ that $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$. Given this, we now know that $\{X_n\}$ and $\left\{X_n^c := X_n 1_{|X_n| \le c}\right\}$ are tail equivalent for all $c > 0$ and in particular $\sum_{n=1}^{\infty} X_n^c$ is almost surely convergent for all $c > 0$. So according to Lemma 22.61 (with $Y_n = X_n^c$),

$$\sum_{n=1}^{\infty} \mathbb{E}X_n^c = \sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right) \text{ converges.}$$

Letting $Y_n := X_n^c - \mathbb{E}X_n^c$, we may now conclude that $\sum_{n=1}^{\infty} Y_n$ is almost surely convergent. Since $\{Y_n\}$ is uniformly bounded and $\mathbb{E}Y_n = 0$ for all $n$, an application of Lemma 22.58 allows us to conclude

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) = \sum_{n=1}^{\infty} \mathbb{E}Y_n^2 < \infty.$$

■

# 23

# Weak Convergence Results

In this chapter we will discuss a couple of different ways to decide weather two probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ are "close" to one another. This will arise later as follows. Suppose $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random variables and $Y$ is another random variable (possibly defined on a different probability space). We would like to understand when, for large $n$, $Y_n$ and $Y$ have nearly the "same" distribution, i.e. when is $\mu_n := \mathrm{Law}(Y_n)$ close to $\mu = \mathrm{Law}(Y)$ for large $n$.

We will often be the case that $Y_n = X_1 + \dots + X_n$ where $\{X_i\}_{i=1}^{n}$ are independent random variables. For this reason it will be useful to record the procedure for computing the law of $Y_n$ in terms of the laws of the $\{X_i\}_{i=1}^{n}$. So before going to the main theme of this chapter let us pause to introduce the relevant notion of the convolution of probability measures on $\mathbb{R}^n$.

## 23.1 Convolutions

**Definition 23.1.** *Let $\mu$ and $\nu$ be two probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. The **convolution** of $\mu$ and $\nu$, denoted $\mu * \nu$, is the measure, $P \circ (X+Y)^{-1}$ where $\{X, Y\}$ are two independent random vectors such that $P \circ X^{-1} = \mu$ and $P \circ Y^{-1} = \nu$.*

Of course we may give a more direct definition of the convolution of $\mu$ and $\nu$ by observing for $A \in \mathcal{B}_{\mathbb{R}^n}$ that

$$\begin{aligned}(\mu * \nu)(A) &= P(X+Y \in A) \\ &= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y)\, 1_A(x+y) \quad (23.1)\\ &= \int_{\mathbb{R}^n} \nu(A-x)\, d\mu(x) \quad (23.2)\\ &= \int_{\mathbb{R}^n} \mu(A-x)\, d\nu(x). \quad (23.3)\end{aligned}$$

This may also be expressed as,

$$(\mu * \nu)(A) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x+y)\, d\mu(x)\, d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x+y)\, d(\mu \otimes \nu)(x,y). \quad (23.4)$$

**Exercise 23.1.** Let $\mu, \nu$, and $\gamma$ be three probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. Show;

1. $\mu * \nu = \nu * \mu$.
2. $\mu * (\nu * \gamma) = (\mu * \nu) * \gamma$. (So it is now safe to write $\mu * \nu * \gamma$ for either side of this equation.)
3. $(\mu * \delta_x)(A) = \mu(A - x)$ for all $x \in \mathbb{R}^n$ where $\delta_x(A) := 1_A(x)$ for all $A \in \mathcal{B}_{\mathbb{R}^n}$ and in particular $\mu * \delta_0 = \mu$.

As a consequence of item 2. of this exercise, if $\{Y_i\}_{i=1}^{n}$ are independent random vectors in $\mathbb{R}^n$ with $\mu_i = \mathrm{Law}(Y_i)$, then

$$\mathrm{Law}(Y_1 + \dots + Y_n) = \mu_1 * \mu_2 * \dots * \mu_n. \quad (23.5)$$

*Remark 23.2.* Suppose that $d\mu(x) = u(x)\,dx$ where $u(x) \geq 0$ and $\int_{\mathbb{R}^n} u(x)\,dx = 1$. Then using the translation invariance of Lebesgue measure and Tonelli's theorem, we have

$$\mu * \nu(f) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x+y)\, u(x)\, dx d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x)\, u(x-y)\, dx d\nu(y)$$

from which it follows that

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\, d\nu(y)\right] dx.$$

If we further assume that $d\nu(x) = v(x)\,dx$, then we have

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\, v(y)\, dy\right] dx.$$

To simplify notation we write,

$$u * v(x) = \int_{\mathbb{R}^n} u(x-y)\, v(y)\, dy = \int_{\mathbb{R}^n} v(x-y)\, u(y)\, dy.$$

*Example 23.3.* Suppose that $n = 1$, $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $d\nu(x) = 1_{[-1,0]}(x)\,dx$ so that $\nu(A) = \mu(-A)$. In this case

$$d(\mu * \nu)(x) = \left(1_{[0,1]} * 1_{[-1,0]}\right)(x)\, dx$$

where

$$\begin{aligned}\left(1_{[0,1]} * 1_{[-1,0]}\right)(x) &= \int_{\mathbb{R}} 1_{[-1,0]}(x-y)\, 1_{[0,1]}(y)\, dy \\ &= \int_{\mathbb{R}} 1_{[0,1]}(y-x)\, 1_{[0,1]}(y)\, dy \\ &= \int_{\mathbb{R}} 1_{[0,1]+x}(y)\, 1_{[0,1]}(y)\, dy \\ &= m([0,1]\cap(x+[0,1])) = (1-|x|)_+ .\end{aligned}$$

## 23.2 Total Variation Distance

**Definition 23.4.** *Let $\mu$ and $\nu$ be two probability measure on a measurable space, $(\Omega,\mathcal{B})$. The total variation distance, $d_{TV}(\mu,\nu)$, is defined as*

$$d_{TV}(\mu,\nu) := \sup_{A\in\mathcal{B}} |\mu(A)-\nu(A)|, \tag{23.6}$$

*i.e. $d_{TV}(\mu,\nu)$ is simply the supremum norm of $\mu-\nu$ as a function on $\mathcal{B}$.*

**Notation 23.5** *Suppose that $X$ and $Y$ are random variables, let*

$$d_{TV}(X,Y) := d_{TV}(\mu_X,\mu_Y) = \sup_{A\in\mathcal{B}_{\mathbb{R}}} |P(X\in A)-P(Y\in A)|,$$

*where $\mu_X = P\circ X^{-1}$ and $\mu_Y = P\circ Y^{-1}$.*

*Example 23.6.* For $x\in\mathbb{R}^n$, let $\delta_x(A) := 1_A(x)$ for all $A\in\mathcal{B}_{\mathbb{R}^n}$. Then one easily shows that $d_{TV}(\delta_x,\delta_y) = 1_{x\neq y}$. Thus if $x\neq y$, in this metric $\delta_x$ and $\delta_y$ are one unit apart no matter how close $x$ and $y$ are in $\mathbb{R}^n$. (This is not always a desirable feature and because of this will introduce shortly another notion of closeness for measures.) More generally if $\mu$ and $\nu$ are any two singular probability measures (i.e. there exists $A\in\mathcal{B}$ such that $\mu(A) = 1 = \nu(A^c)$, then $d_{TV}(\mu,\nu) = 1$.

**Exercise 23.2.** Let $\mathcal{P}_1$ denote the set of probability measures on $(\Omega,\mathcal{B})$. Show $d_{TV}$ is a complete metric on $\mathcal{P}_1$.

**Exercise 23.3.** Suppose that $\mu,\nu,$ and $\gamma$ are probability measures on $(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n})$. Show $d_{TV}(\mu*\nu,\mu*\gamma) \le d_{TV}(\nu,\gamma)$. Use this fact along with Exercise 23.2 to show,

$$d_{TV}(\mu_1*\mu_2*\cdots*\mu_n,\nu_1*\nu_2*\cdots*\nu_n) \le \sum_{i=1}^n d_{TV}(\mu_i,\nu_i)$$

for all choices probability measures, $\mu_i$ and $\nu_i$ on $(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n})$.

*Remark 23.7.* The function, $\lambda:\mathcal{B}\to\mathbb{R}$ defined by, $\lambda(A) := \mu(A)-\nu(A)$ for all $A\in\mathcal{B}$, is an example of a "signed measure." For signed measures, one usually defines

$$\|\lambda\|_{TV} := \sup\left\{\sum_{i=1}^n |\lambda(A_i)| : n\in\mathbb{N} \text{ and partitions, } \{A_i\}_{i=1}^n\subset\mathcal{B} \text{ of } \Omega\right\}.$$

You are asked to show in Exercise 23.4 below, that when $\lambda = \mu-\nu$, $d_{TV}(\mu,\nu) = \frac{1}{2}\|\mu-\nu\|_{TV}$.

**Lemma 23.8 (Scheffé's Lemma).** *Suppose that $m$ is another positive measure on $(\Omega,\mathcal{B})$ such that there exists measurable functions, $f,g:\Omega\to[0,\infty)$, such that $d\mu = fdm$ and $d\nu = gdm$.*[1] *Then*

$$d_{TV}(\mu,\nu) = \frac{1}{2}\int_\Omega |f-g|\, dm.$$

*Let us now further suppose that $\{\mu_n\}_{n=1}^\infty\cup\{\nu\}$ are probability measures of the form, $d\mu_n = f_n dm$ and $d\nu = gdm$ with $g,f_n:\Omega\to[0,\infty)$. If $f_n\to g$, $m$ - a.e. with $d\nu = gdm$ still being a probability measure, then $d_{TV}(\mu_n,\nu)\to 0$ as $n\to\infty$.*

**Proof.** Let $\lambda = \mu-\nu$ and $h := f-g:\Omega\to\mathbb{R}$ so that $d\lambda = hdm$ and

$$d_{TV}(\mu,\nu) = \sup_{A\in\mathcal{B}} |\mu(A)-\nu(A)| = \sup_{A\in\mathcal{B}} |\lambda(A)|.$$

Since

$$\lambda(\Omega) = \mu(\Omega)-\nu(\Omega) = 1-1 = 0,$$

if $A\in\mathcal{B}$ we have

$$\lambda(A)+\lambda(A^c) = \lambda(\Omega) = 0.$$

In particular this shows $|\lambda(A)| = |\lambda(A^c)|$ and therefore,

$$\begin{aligned}|\lambda(A)| &= \frac{1}{2}\left[|\lambda(A)|+|\lambda(A^c)|\right] = \frac{1}{2}\left[\left|\int_A hdm\right| + \left|\int_{A^c} hdm\right|\right] \\ &\le \frac{1}{2}\left[\int_A |h|\, dm + \int_{A^c} |h|\, dm\right] = \frac{1}{2}\int_\Omega |h|\, dm.\end{aligned} \tag{23.7}$$

This shows

$$d_{TV}(\mu,\nu) = \sup_{A\in\mathcal{B}} |\lambda(A)| \le \frac{1}{2}\int_\Omega |h|\, dm.$$

To prove the converse inequality, simply take $A = \{h>0\}$ (note $A^c = \{h\le 0\}$) in Eq. (23.7) to find

[1] Fact: it is always possible to do this by taking $m = \mu+\nu$ for example.

$$|\lambda(A)| = \frac{1}{2}\left[\int_A h dm - \int_{A^c} h dm\right]$$
$$= \frac{1}{2}\left[\int_A |h| dm + \int_{A^c} |h| dm\right] = \frac{1}{2}\int_\Omega |h| dm.$$

For the second assertion, observe that $|f_n - g| \to 0$ $m$ − a.e., $|f_n - g| \leq G_n := f_n + g \in L^1(m)$, $G_n \to G := 2g$ a.e. and $\int_\Omega G_n dm = 2 \to 2 = \int_\Omega G dm$ and $n \to \infty$. Therefore, by the dominated convergence Theorem 9.27,

$$\lim_{n\to\infty} d_{TV}(\mu_n, \nu) = \frac{1}{2}\lim_{n\to\infty}\int_\Omega |f_n - g| dm = 0.$$

■

For a concrete application of Scheffé's Lemma 23.8, see Proposition 23.53 below.

**Corollary 23.9.** *Let $\|h\|_\infty := \sup_{\omega\in\Omega}|h(\omega)|$ when $h : \Omega \to \mathbb{R}$ is a bounded random variable. Continuing the notation in Scheffé's lemma above, we have*

$$d_{TV}(\mu, \nu) = \frac{1}{2}\sup\left\{\left|\int_\Omega h d\mu - \int_\Omega h d\nu\right| : \|h\|_\infty \leq 1\right\}. \tag{23.8}$$

*Consequently,*

$$\left|\int_\Omega h d\mu - \int_\Omega h d\nu\right| \leq 2 d_{TV}(\mu, \nu) \cdot \|h\|_\infty \tag{23.9}$$

*and in particular, for all bounded and measurable functions, $h : \Omega \to \mathbb{R}$,*

$$\lim_{n\to\infty} d_{TV}(\mu_n, \nu) = 0 \implies \lim_{n\to\infty}\int_\Omega h d\mu_n = \int_\Omega h d\nu. \tag{23.10}$$

***Proof.*** *We begin by observing that*

$$\left|\int_\Omega h d\mu - \int_\Omega h d\nu\right| = \left|\int_\Omega h(f - g) dm\right| \leq \int_\Omega |h||f - g| dm$$
$$\leq \|h\|_\infty \int_\Omega |f - g| dm = 2 d_{TV}(\mu, \nu)\|h\|_\infty.$$

*Moreover, from the proof of Scheffé's Lemma 23.8, we have*

$$d_{TV}(\mu, \nu) = \frac{1}{2}\left|\int_\Omega h d\mu - \int_\Omega h d\nu\right|$$

*when $h := 1_{f>g} - 1_{f\leq g}$. These two equations prove Eqs. (23.8) and (23.9) and the latter implies Eq. (23.10).* ■

**Exercise 23.4.** Under the hypothesis of Scheffé's Lemma 23.8, show

$$\|\mu - \nu\|_{TV} = \int_\Omega |f - g| dm = 2 d_{TV}(\mu, \nu).$$

**Exercise 23.5.** Suppose that $\Omega$ is a (at most) countable set, $\mathcal{B} := 2^\Omega$, and $\{\mu_n\}_{n=0}^\infty$ are probability measures on $(\Omega, \mathcal{B})$. Show

$$d_{TV}(\mu_n, \mu_0) = \frac{1}{2}\sum_{\omega\in\Omega}|\mu_n(\{\omega\}) - \mu_0(\{\omega\})|$$

and $\lim_{n\to\infty} d_{TV}(\mu_n, \mu_0) = 0$ iff $\lim_{n\to\infty}\mu_n(\{\omega\}) = \mu_0(\{\omega\})$ for all $\omega \in \Omega$.

**Exercise 23.6.** Let $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = 1 - p$ and $\nu_\lambda(\{n\}) := e^{-\lambda}\frac{\lambda^n}{n!}$ for all $n \in \mathbb{N}_0$.

1. Find $d_{TV}(\mu_p, \mu_q)$ for all $0 \leq p, q \leq 1$.
2. Show $d_{TV}(\mu_p, \nu_p) = p(1 - e^{-p})$ for all $0 \leq p \leq 1$. From this estimate and the estimate,
$$1 - e^{-p} = \int_0^p e^{-x} dx \leq \int_0^p 1 dx = p, \tag{23.11}$$
it follows that $d_{TV}(\mu_p, \nu_p) \leq p^2$ for all $0 \leq p \leq 1$.
3. Show
$$d_{TV}(\nu_\lambda, \nu_\gamma) \leq |\lambda - \gamma| \text{ for all } \lambda, \gamma \in \mathbb{R}_+. \tag{23.12}$$

**Hints: (Andy Parrish's method – a former 280 student.)**

a) Observe that for any $n \in \mathbb{N}$ we have $\nu_\lambda$ and $\nu_\gamma$ are equal to the $n$ – fold convolutions of $\nu_{\lambda/n}$ and $\nu_{\gamma/n}$ and use this to conclude
$$d_{TV}(\nu_\lambda, \nu_\gamma) \leq n d_{TV}\left(\nu_{\lambda/n}, \nu_{\gamma/n}\right). \tag{23.13}$$
b) Using item 2. of this exercise, show
$$\left|d_{TV}\left(\nu_{\lambda/n}, \nu_{\gamma/n}\right) - d_{TV}\left(\mu_{\lambda/n}, \mu_{\gamma/n}\right)\right| \leq C n^{-2}.$$
c) Finally make use of your results in item 1. part b. in order to let $n \to \infty$ in Eq. (23.13).

The next theorem should be compared with Exercise 9.6 which may be stated as follows. If $\{Z_i\}_{i=1}^n$ are i.i.d. Bernoulli random variables with $P(Z_i = 1) = p = O(1/n)$ and $S = Z_1 + \cdots + Z_n$, then $P(S = k) \cong P(\text{Poisson}(pn) = k)$ which is valid for $k \ll n$.

**Theorem 23.10 (Law of rare events).** *Let $\{Z_i\}_{i=1}^n$ be independent Bernoulli random variables with $P(Z_i=1)=p_i\in(0,1)$ and $P(Z_i=0)=1-p_i$, $S:=Z_1+\dots+Z_n$, $a:=p_1+\dots+p_n$, and $X\overset{d}{=}\operatorname{Poi}(a)$. Then for any[2] $A\in\mathcal{B}_{\mathbb{R}}$ we have*

$$|P(S\in A)-P(X\in A)|\le\sum_{i=1}^n p_i^2, \tag{23.14}$$

*or in short,*

$$d_{TV}\left(\sum_{i=1}^n Z_i, X\right)\le\sum_{i=1}^n p_i^2.$$

*(Of course this estimate has no content unless $\sum_{i=1}^n p_i^2<1$.)*

**Proof.** Let $\{X_i\}_{i=1}^n$ be independent random variables with $X_i\overset{d}{=}Pois(p_i)$ for each $i$. It then follows from Exercises 23.3 and 23.6 that,

$$d_{TV}\left(\sum_{i=1}^n Z_i,\sum_{i=1}^n X_i\right)\le\sum_{i=1}^n d_{TV}(Z_i,X_i)=\sum_{i=1}^n p_i\left(1-e^{-p_i}\right)\le\sum_{i=1}^n p_i^2.$$

■

The reader should compare the proof of this theorem with the proof of the central limit theorem in Theorem 12.39. For another less quantitative Poisson limit theorem, see Theorem 25.23.

For the next result we will suppose that $(Y,\mathcal{M},\mu)$ is a finite measure space with the following properties;

1. $\{y\}\in\mathcal{M}$ and $\mu(\{y\})=0$ for all $y\in Y$,
2. to any $A\in\mathcal{M}$ and $\varepsilon>0$, there exists a finite partition $\{A_n\}_{n=1}^{N=N(\varepsilon)}\subset\mathcal{M}$ of $A$ such that $\mu(A_n)\le\varepsilon$ for all $n$. (This assumption actually follows from assumption the no-atom assumption 1. above, see Lemma 14.55.

In what follows below we will write $F(A)=o(\mu(A))$ provided there exits an increasing function, $\delta:\mathbb{R}_+\to\mathbb{R}_+$, such that $\delta(x)\to0$ as $x\to0$ and $|F(A)|\le\mu(A)\,\delta(\mu(A))$ for all $A\in\mathcal{M}$.

**Proposition 23.11 (Why Poisson).** *Suppose $(Y,\mathcal{M},\mu)$ is finite measure space with the properties given above and $\{N(A):A\in\mathcal{M}\}$ is a collection of $\mathbb{N}_0$ – valued random variables with the following properties;*

1. *If $\{A_j\}_{j=1}^n\subset\mathcal{M}$ are disjoint, then $\{N(A_i)\}_{i=1}^n$ are independent random variables and*
$$N\left(\sum_{i=1}^n A_i\right)=\sum_{i=1}^n N(A_i)\ \ a.s.$$
2. *$P(N(A)\ge2)=o(\mu(A))$.*
3. *$|P(N(A)\ge1)-\mu(A)|=o(\mu(A))$.*

*Then $N(A)\overset{d}{=}\operatorname{Poi}(\mu(A))$ for all $A\in\mathcal{M}$ and in particular $\mathbb{E}N(A)=\mu(A)$ for all $A\in\mathcal{M}$.*

**Proof.** Let $A\in\mathcal{M}$ and $\varepsilon>0$ be given. Choose a partition $\{A_i^\varepsilon\}_{i=1}^N\subset\mathcal{M}$ of $A$ such that $\mu(A_i^\varepsilon)\le\varepsilon$ for all $i$. Let $Z_i:=1_{N(A_i^\varepsilon)\ge1}$ and $S:=\sum_{i=1}^N Z_i$. Using

$$N(A)=\sum_{i=1}^N N(A_i^\varepsilon)$$

and Lemma 23.13, we have

$$|P(N(A)=k)-P(S=k)|\le P(N(A)\ne S)\le\sum_{i=1}^N P(Z_i\ne N(A_i^\varepsilon)).$$

Since $\{Z_i\ne N(A_i^\varepsilon)\}=\{N(A_i^\varepsilon)\ge2\}$ and $P(N(A_i^\varepsilon)\ge2)=o(\mu(A_i^\varepsilon))$, it follows that

$$\begin{aligned}|P(N(A)=k)-P(S=k)|&\le\sum_{i=1}^N\mu(A_i^\varepsilon)\,\delta(\mu(A_i^\varepsilon))\\&\le\sum_{i=1}^N\mu(A_i^\varepsilon)\,\delta(\varepsilon)=\delta(\varepsilon)\,\mu(A).\end{aligned}\tag{23.15}$$

On the other hand, $\{Z_i\}_{i=1}^N$ are independent Bernoulli random variables with

$$P(Z_i=1)=P(N(A_i^\varepsilon)\ge1),$$

and $a_\varepsilon=\sum_{i=1}^N P(N(A_i^\varepsilon)\ge1)$. Then by the Law of rare events Theorem 23.10,

$$\begin{aligned}\left|P(S=k)-\frac{a_\varepsilon^k}{k!}e^{-a_\varepsilon}\right|&\le\sum_{i=1}^N\left[P(N(A_i^\varepsilon)\ge1)\right]^2\le\sum_{i=1}^N\left[\mu(A_i^\varepsilon)+o(\mu(A_i^\varepsilon))\right]^2\\&\le\sum_{i=1}^N\mu(A_i^\varepsilon)^2\left(1+\delta'(\varepsilon)\right)^2=\left(1+\delta'(\varepsilon)\right)^2\varepsilon\mu(A).\end{aligned}\tag{23.16}$$

Combining Eqs. (23.15) and (23.16) shows

$$\left|P(N(A)=k)-\frac{a_\varepsilon^k}{k!}e^{-a_\varepsilon}\right|\le\left[\delta(\varepsilon)+\left(1+\delta'(\varepsilon)\right)^2\varepsilon\right]\mu(A) \tag{23.17}$$

[2] Actually, since $S$ and $X$ are $\mathbb{N}_0$ – valued, we may as well assume that $A\subset\mathbb{N}_0$.

where $a_\varepsilon$ satisfies

$$\begin{aligned}|a_\varepsilon - \mu(A)| &= \left|\sum_{i=1}^N \left[P\left(N\left(A_i^\varepsilon\right) \geq 1\right) - \mu\left(A_i^\varepsilon\right)\right]\right| \\ &\leq \sum_{i=1}^N \left|\left[P\left(N\left(A_i^\varepsilon\right) \geq 1\right) - \mu\left(A_i^\varepsilon\right)\right]\right| \leq \sum_{i=1}^N o\left(\mu\left(A_i^\varepsilon\right)\right) \\ &\leq \sum_{i=1}^N \mu\left(A_i^\varepsilon\right)\left|\delta'\left(\mu\left(A_i^\varepsilon\right)\right)\right| \leq \mu(A)\,\delta'(\varepsilon).\end{aligned}$$

Hence we may let $\varepsilon \downarrow 0$ in Eq. (23.17) to find

$$P(N(A) = k) = \frac{(\mu(A))^k}{k!} e^{-\mu(A)}.$$

■

See [36, p. 13-16.] for another variant of this theorem in the case that $\Omega = \mathbb{R}_+$. See Theorem 13.12 and Exercises 13.6 – 13.8 for concrete constructions of Poisson processes.

*Remark 23.12.* Here is the short version of the above Proposition. We are assuming $\{N(A_i)\}_{i=1}^n$ are independent if $\{A_i\}_{i=1}^n$ are pairwise disjoint and for sets $A$ such that $\mu(A)$ is small,

$$N(A) \overset{d}{=} Bern(\mu(A)) + o(\mu(A)). \tag{23.18}$$

This last statement is short hand for the assumptions that

$$P(N(A) = 0) = 1 - \mu(A) + o(\mu(A)) \text{ and } P(N(A) = 1) = \mu(A) + o(\mu(A)),$$

and as a consequence,

$$P(N(A) \geq 2) = 1 - (P(N(A) = 0) + P(N(A) = 1)) = o(\mu(A)).$$

Assumption (23.18) implies

$$\begin{aligned}\mathbb{E}\left[z^{N(A)}\right] &= (1 - \mu(A) + o(\mu(A))) + (\mu(A) + o(\mu(A)))\,z + o(\mu(A)) \\ &= 1 + \mu(A)(z-1) + o(\mu(A)) \\ &= \exp(1 + \mu(A)(z-1) + o(\mu(A))). \end{aligned} \tag{23.19}$$

Now for an arbitrary set $A$ with $\mu = \mu(A) < \infty$, we assume for any $n \in \mathbb{N}$ that there exists a partition $\{A_i\}_{i=1}^n$ of $A$ so that $\mu(A_i) = \mu/n$ for each $n$. It then follows from Eq. (23.19) and the independence of the $\{N(A_i)\}_{i=1}^n$ that

$$\begin{aligned}\mathbb{E}\left[z^{N(A)}\right] &= \mathbb{E}\left[z^{\sum_{i=1}^n N(A_i)}\right] = \prod_{i=1}^n \mathbb{E}\left[z^{N(A_i)}\right] = \prod_{i=1}^n \exp(1 + \mu/n\,(z-1) + o(\mu/n)) \\ &= \exp\left(\mu(z-1) + no\left(\frac{1}{n}\right)\right).\end{aligned}$$

Letting $n \to \infty$ in this equation shows $\mathbb{E}\left[z^{N(A)}\right] = \exp(\mu(z-1))$ from which it follows that that $N(A) \overset{d}{=} \mathrm{Poi}(\mu(A))$.

## 23.3 A Coupling Estimate

**Lemma 23.13 (Coupling Estimates).** *Suppose $X$ and $Y$ are any random variables on a probability space, $(\Omega, \mathcal{B}, P)$ and $A \in \mathcal{B}_{\mathbb{R}}$. Then*

$$|P(X \in A) - P(Y \in A)| \leq P(\{X \in A\} \triangle \{Y \in A\}) \leq P(X \neq Y). \tag{23.20}$$

**Proof.** The proof is simply;

$$\begin{aligned}|P(X \in A) - P(Y \in A)| &= |\mathbb{E}[1_A(X) - 1_A(Y)]| \\ &\leq \mathbb{E}|1_A(X) - 1_A(Y)| = \mathbb{E}1_{\{X\in A\}\triangle\{Y\in A\}} \\ &\leq \mathbb{E}1_{X\neq Y} = P(X \neq Y).\end{aligned}$$

■

Pushing the above proof a little more we have, if $\{A_i\}$ is a partition of $\Omega$, then

$$\begin{aligned}\sum_i |P(X \in A_i) - P(Y \in A_i)| &= \sum_i |\mathbb{E}[1_{A_i}(X) - 1_{A_i}(Y)]| \\ &\leq \sum_i \mathbb{E}|1_{A_i}(X) - 1_{A_i}(Y)| \\ &\leq \mathbb{E}[1_{X\neq Y} : X \in A_i \text{ or } Y \in A_i] \\ &\leq \mathbb{E}[1_{X\neq Y} : X \in A_i] + \mathbb{E}[1_{X\neq Y} : Y \in A_i] \\ &= 2P(X \neq Y).\end{aligned}$$

This shows

$$\|X_*P - Y_*P\|_{TV} \leq 2P(X \neq Y).$$

This is really not more general than Eq. (23.20) since the Hahn decomposition theorem we know that in fact the signed measure, $\mu := X_*P - Y_*P$, has total variation given by

$$\|\mu\|_{TV} = \mu(\Omega_+) - \mu(\Omega_-)$$

where $\Omega = \Omega_+ \sum \Omega_-$ with $\Omega_+$ being a positive set and $\Omega_-$ being a negative set. Moreover, since $\mu(\Omega) = 0$ we must in fact have $\mu(\Omega_+) = -\mu(\Omega_-)$ so that

$$\begin{aligned}\|X_* P - Y_* P\|_{TV} &= \|\mu\|_{TV} = 2\mu(\Omega_+) = 2\left|P(X \in \Omega_+) - P(Y \in \Omega_+)\right| \\ &\leq 2P(X \neq Y).\end{aligned}$$

Here is perhaps a better way to view the above lemma. Suppose that we are given two probability measures, $\mu, \nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ (or any other measurable space, $(S, \mathcal{B}_S)$). We would like to estimate $\|\mu - \nu\|_{TV}$. The lemma states that if $X, Y$ are random variables (vectors) on some probability space such that $\mathrm{Law}_P(X) = \mu$ and $\mathrm{Law}_P(Y) = \nu$, then

$$\|\mu - \nu\|_{TV} \leq 2P(X \neq Y).$$

Suppose that we let $\rho := \mathrm{Law}_P(X, Y)$ on $(S \times S, \mathcal{B}_S \otimes \mathcal{B}_S)$, $\pi_i : S \times S \to S$ be the projection maps for $i = 1, 2$, then $(\pi_1)_* \rho = \mu$, $(\pi_2)_* \rho = v$, and

$$\|\mu - \nu\|_{TV} \leq 2P(X \neq Y) = 2\rho(\pi_1 \neq \pi_2) = 2\rho\left(S^2 \setminus \Delta\right)$$

where $\Delta = \{(s, s) : s \in S\}$ is the diagonal in $S^2$. Thus finding a coupling amounts to fining a probability measure, $\rho$, on $\left(S^2, \mathcal{B}_S \otimes \mathcal{B}_S\right)$ whose marginals are $\mu$ and $\nu$ respectively. Then we will have the coupling estimate,

$$\|\mu - \nu\|_{TV} \leq 2\rho\left(S^2 \setminus \Delta\right).$$

**Lemma 23.14 (Optimality of coupling).** *Suppose that $S$ is a finite (or countable) set and $\mu$ and $\nu$ are two probabilities on $S$. Then there exists a coupling measure $\rho$ on $S \times S$ such that*

$$\|\mu - \nu\|_{TV} = 2\rho\left(S^2 \setminus \Delta\right).$$

**Proof.** I will assume that $S$ is a finite set for simplicity although it has little effect on the proof. Let $U$ be a uniform random variable and choose disjoint sub-intervals $\{J_x\}_{x \in S}$ of $(0, 1]$ such that $|J_x| = \mu(x) \wedge \nu(x)$ for all $x \in S$. Let

$$S_{\pm} = \begin{cases} x \in S \text{ if } \mu(x) > \nu(x) \\ x \in S \text{ if } \mu(x) < \nu(x) \end{cases}$$

and observe that

$$\begin{aligned}\sum_{x \in S} |J_x| + \sum_{x \in S_+} [\mu(x) - \nu(x)] &= \sum_{x \in S} \mu(x) \wedge \nu(x) + \sum_{x \in S_+} [\mu(x) - \nu(x)] \\ &= \sum_{x \in S} \mu(x) = 1\end{aligned}$$

and similarly,

$$\sum_{x \in S} |J_x| + \sum_{x \in S_-} [\nu(x) - \mu(x)] = \sum_{x \in S} \nu(x) = 1.$$

Assuming we have lined up the $\{J_x\}$ in $(0, \sum_{x \in S} |J_x|]$ we may choose two partitions $\{K_x\}_{x \in S_+}$ and $\{K_x\}_{x \in S_-}$ of $(\sum_{x \in S} |J_x|, 1]$ such that $|K_x| = |\mu(x) - \nu(x)|$ for $x \in S_+ \cup S_-$. If we then let

$$Y := \sum_{x \in S_+} x 1_{J_x \cup K_x} + \sum_{x \in S_-} x 1_{J_x} \text{ and } Z := \sum_{x \in S_-} x 1_{J_x \cup K_x} + \sum_{x \in S_+} x 1_{J_x}$$

then $P(Y = s) = \mu(s)$ and $P(Z = s) = \nu(s)$ for all $s \in S$. Moreover (as you should verify),

$$\{Y \neq Z\} = \cup_{x \in S_+} K_x = \cup_{x \in S_-} K_x$$

so that

$$P(Y \neq Z) = \sum_{x \in S_+} |K_x| = \frac{1}{2} \sum_{x \in S} |K_x| = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

Taking $\rho := \mathrm{Law}(Y, Z)$ as a measure on $S \times S$ then completes the proof. ■

As an example of how to use Lemma 23.13 let us give a coupling proof of Theorem 23.10.

**Proof.** (A coupling proof of Theorem 23.10.) We are going to construct a coupling for $S_* P$ and $X_* P$. Finding such a coupling amounts to representing $X$ and $S$ on the same probability space. We are going to do this by building all random variables in site out of $\{U_i\}_{i=1}^n$, where the $\{U_i\}_{i=1}^n$ are i.i.d. random variables distributed uniformly on $[0, 1]$.

If we define,

$$Z_i := 1_{(1-p_i, 1]}(U_i) = 1_{1 - p_i < U_i \leq 1},$$

then $\{Z_i\}_{i=1}^n$ are independent Bernoulli random variables with $P(Z_i = 1) = p_i$. We are now also going to construct[3] out of the $\{U_i\}_{i=1}^n$, a sequence of independent Poisson random variables, $\{X_i\}_{i=1}^n$ with $X_i = \mathrm{Poi}(p_i)$. To do this define

[3] At this point we could appeal to Lemma 23.14 in order to find define $X_i$ and $Z_i$ as functions of $U_i$ such that 1) $\{Z_i\}_{i=1}^n$ are independent Bernoulli random variables with $P(Z_i = 1) = p_i$, 2) $\{X_i\}_{i=1}^n$ are independent Poisson random variables with $X_i \overset{d}{=} \mathit{Pois}(p_i)$, and 3)

$$P(X_i \neq Z_i) = d_{TV}(\mathit{Bern}(p_i), \mathit{Pois}(p_i)) = p_i\left(1 - e^{-p_i}\right) \leq p_i^2.$$

$$\alpha_i(k) := P(\mathrm{Poi}(p_i) \le k) = e^{-p_i} \sum_{j=0}^{k} \frac{p_i^j}{j!}$$

with the convention that $\alpha_i(-1) = 0$. Notice that

$$e^{-p_i} \le \alpha_i(k) \le \alpha_i(k+1) \le 1 \text{ for all } k \in \mathbb{N}_0$$

and for $p_i$ small, we have

$$\alpha_i(0) = e^{-p_i} \cong 1 - p_i$$
$$\alpha_i(1) = e^{-p_i}(1 + p_i) \cong 1 - p_i^2.$$

If we define (see Figure 23.1) by

$$X_i := \sum_{k=0}^{\infty} k 1_{\alpha_i(k-1) < U_i \le \alpha_i(k)},$$

then $X_i = \mathrm{Poi}(p_i)$ since

$$P(X_i = k) = P(\alpha_i(k-1) < U_i \le \alpha_i(k))$$
$$= \alpha_i(k) - \alpha_i(k-1) = e^{-p_i} \frac{p_i^k}{k!}.$$

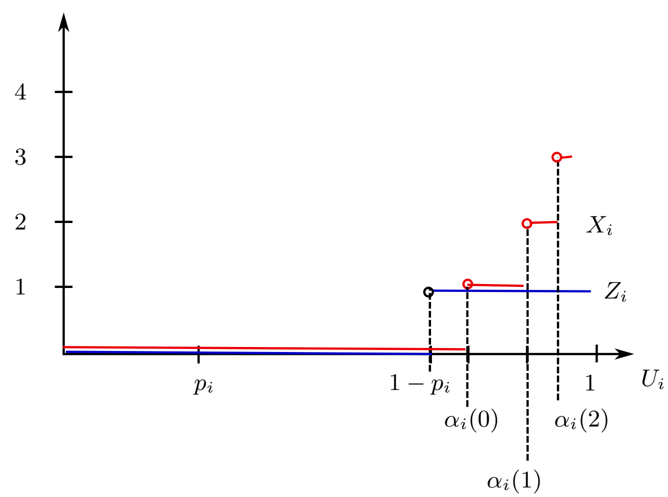



**Fig. 23.1.** Plots of $X_i$ and $Z_i$ as functions of $U_i$.

It is also clear that $\{X_i\}_{i=1}^n$ are independent and hence by Lemma 13.1, it follows that $X := \sum_{i=1}^n X_i \overset{d}{=} \mathrm{Poi}(a)$.

An application of Lemma 23.13 now shows

$$|P(S \in A) - P(X \in A)| \le P(S \ne X)$$

and since $\{S \ne X\} \subset \cup_{i=1}^n \{X_i \ne Z_i\}$, we may conclude

$$|P(S \in A) - P(X \in A)| \le \sum_{i=1}^{n} P(X_i \ne Z_i).$$

As is easily seen from Figure 23.1,

$$P(X_i \ne Z_i) = [\alpha_i(0) - (1 - p_i)] + 1 - \alpha_i(1)$$
$$= \left[e^{-p_i} - (1 - p_i)\right] + 1 - e^{-p_i}(1 + p_i)$$
$$= p_i\left(1 - e^{-p_i}\right) \le p_i^2$$

where we have used the estimate in Eq. (23.11) for the last inequality. ■

## 23.4 Weak Convergence

Recall that to each right continuous increasing function, $F : \mathbb{R} \to \mathbb{R}$ there is a unique measure, $\mu_F$, on $\mathcal{B}_\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a \le b < \infty$. To simplify notation in this section we will now write $F(A)$ for $\mu_F(A)$ for all $A \in \mathcal{B}_\mathbb{R}$ and in particular $F((a,b]) := F(b) - F(a)$ for all $-\infty < a \le b < \infty$.

*Example 23.15.* Suppose that $P\left(X_n = \frac{i}{n}\right) = \frac{1}{n}$ for $i \in \{1, 2, \dots, n\}$ so that $X_n$ is a discrete "approximation" to the uniform distribution, i.e. to $U$ where $P(U \in A) = m(A \cap [0,1])$ for all $A \in \mathcal{B}_\mathbb{R}$. If we let $A_n = \left\{\frac{i}{n} : i = 1, 2, \dots, n\right\}$, then $P(X_n \in A_n) = 1$ while $P(U \in A_n) = 0$. Therefore, it follows that $d_{TV}(X_n, U) = 1$ for all $n$.[4]

Nevertheless we would like $X_n$ to be close to $U$ in distribution. Let us observe that if we let $F_n(y) := P(X_n \le y)$ and $F(y) := P(U \le y)$, then

$$F_n(y) = P(X_n \le y) = \frac{1}{n} \# \left\{ i \in \{1, 2, \dots, n\} : \frac{i}{n} \le y \right\}$$

and

$$F(y) := P(U \le y) = (y \wedge 1) \vee 0.$$

From these formula, it easily follows that $F(y) = \lim_{n\to\infty} F_n(y)$ for all $y \in \mathbb{R}$, see Figure 23.2. This suggest that we should say that $X_n$ converges in distribution to $X$ iff $P(X_n \le y) \to P(X \le y)$ for all $y \in \mathbb{R}$. However, the next simple example shows this definition is also too restrictive.

[4] More generally, if $\mu$ and $\nu$ are two probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$ while $\nu$ concentrates on a countable set, then $d_{TF}(\mu, \nu) = 1$.

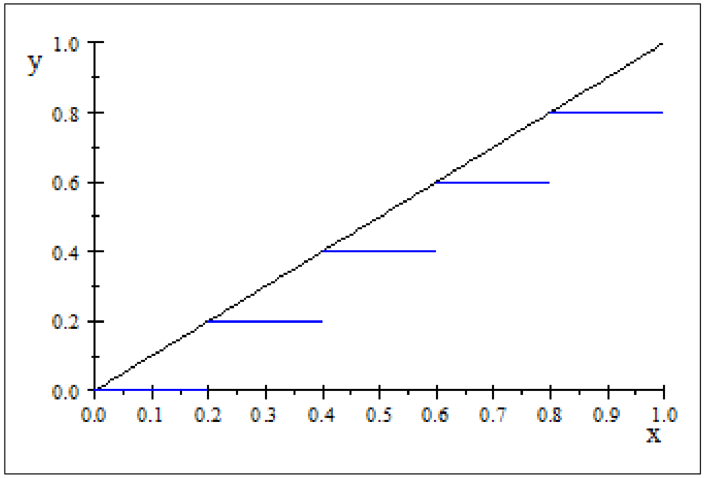



**Fig. 23.2.** The plot of $F_5$ in blue and $F$ is black on $[0,1]$.

*Example 23.16.* Suppose that $P(X_n = 1/n) = 1$ for all $n$ and $P(X_0 = 0) = 1$. Then it is reasonable to insist that $X_n$ converges of $X_0$ in distribution. However, $F_n(y) = 1_{y \geq 1/n} \to 1_{y \geq 0} = F_0(y)$ for all $y \in \mathbb{R}$ **except** for $y = 0$. Observe that $y$ is the only point of discontinuity of $F_0$.

**Notation 23.17** *Let $(X, d)$ be a metric space, $f : X \to \mathbb{R}$ be a function. The set of $x \in X$ where $f$ is continuous (discontinuous) at $x$ will be denoted by $\mathcal{C}(f)$ ($\mathcal{D}(f)$).*

*Remark 23.18.* If $F : \mathbb{R} \to [0,1]$ is a non-decreasing function, then $\mathcal{D} := \mathcal{D}(F)$ is at most countable. To see this, suppose that $\varepsilon > 0$ is given and let

$$\mathcal{D}_\varepsilon := \{y \in \mathbb{R} : F(y+) - F(y-) \geq \varepsilon\}.$$

If $y < y'$ with $y, y' \in \mathcal{D}_\varepsilon$, then $F(y+) < F(y'-)$ and $(F(y-), F(y+))$ and $(F(y'-), F(y'+))$ are disjoint intervals of length greater that $\varepsilon$. Hence it follows that

$$1 = m([0,1]) \geq \sum_{y \in \mathcal{D}_\varepsilon} m((F(y-), F(y+))) \geq \varepsilon \cdot \#(\mathcal{D}_\varepsilon)$$

and hence that $\#(\mathcal{D}_\varepsilon) \leq \varepsilon^{-1} < \infty$. Therefore $\mathcal{D} := \cup_{k=1}^\infty \mathcal{D}_{1/k}$ is at most countable.

**Definition 23.19.** *Let $\{F, F_n : n = 1, 2, \dots\}$ be a collection of right continuous non-increasing functions from $\mathbb{R}$ to $[0,1]$. Then*

1. *$F_n$ converges to $F$ **vaguely** and write, $F_n \xrightarrow{v} F$, iff $F_n((a,b]) \to F((a,b])$ for all $a, b \in \mathcal{C}(F)$.*
2. *$F_n$ converges to $F$ **weakly** and write, $F_n \xrightarrow{w} F$, iff $F_n(x) \to F(x)$ for all $x \in \mathcal{C}(F)$.*
3. *We say $F$ is **proper**, if $F$ is a distribution function of a probability measure, i.e. if $F(\infty) = 1$ and $F(-\infty) = 0$.*

*Example 23.20.* If $X_n$ and $U$ are as in Example 23.15 and $F_n(y) := P(X_n \leq y)$ and $F(y) := P(Y \leq y)$, then $F_n \xrightarrow{v} F$ and $F_n \xrightarrow{w} F$.

*Example 23.21.* Suppose that $Z$ is a random variable and $F(x) := P(Z \leq x)$. Let $X_n = n + Z$ and

$$\begin{aligned} F_n(x) &:= P(X_n \leq x) = P(n + Z \leq x) \\ &= P(Z \leq x - n) = F(x - n) \to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus it follows that $F_n \xrightarrow{w} 0 = F_\infty$. Notice that limit, $F_\infty$, is no longer a distribution function, i.e. $F_\infty$ is not proper.

**Lemma 23.22.** *Let $\{F, F_n : n = 1, 2, \dots\}$ be a collection of **proper** distribution functions. Then $F_n \xrightarrow{v} F$ iff $F_n \xrightarrow{w} F$.*

**Proof.** If $F_n \xrightarrow{w} F$, then $F_n((a,b]) = F_n(b) - F_n(a) \to F(b) - F(a) = F((a,b])$ for all $a, b \in \mathcal{C}(F)$ and therefore $F_n \xrightarrow{v} F$. So now suppose $F_n \xrightarrow{v} F$ and let $a < x$ with $a, x \in \mathcal{C}(F)$. Then

$$F(x) = F(a) + \lim_{n\to\infty} [F_n(x) - F_n(a)] \leq F(a) + \liminf_{n\to\infty} F_n(x).$$

Letting $a \downarrow -\infty$, using the fact that $F$ is proper, implies

$$F(x) \leq \liminf_{n\to\infty} F_n(x).$$

Likewise,

$$F(x) - F(a) = \lim_{n\to\infty} [F_n(x) - F_n(a)] \geq \limsup_{n\to\infty} [F_n(x) - 1] = \limsup_{n\to\infty} F_n(x) - 1$$

which upon letting $a \uparrow \infty$, (so $F(a) \uparrow 1$) allows us to conclude,

$$F(x) \geq \limsup_{n\to\infty} F_n(x).$$

■

**Definition 23.23.** *In the case where $F_n$ and $F$ are proper and $F_n \xrightarrow{w} F$, we will write $F_n \implies F$. Moreover if $\{X_n\}_{n=0}^\infty$ is a sequence of random variables, we say $X_n$ **converges weakly** or to **converges in distribution** to $X_0$ iff $F_n(y) := P(X_n \leq y) \implies F(y) := P(X \leq y)$ and we abbreviate this convergence by writing $X_n \implies X$.*

*Example 23.24 (Central Limit Theorem).* The central limit theorem (see Theorems 9.74, Corollary 12.41, and Theorem 24.29) states; if $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^2(P)$ random variables with $\mu := \mathbb{E}X_1$ and $\sigma^2 = \operatorname{Var}(X_1)$, then

$$\frac{S_n - n\mu}{\sqrt{n}} \implies N(0,\sigma) \overset{d}{=} \sigma N(0,1).$$

Written out explicitly we find

$$\begin{aligned}\lim_{n\to\infty} P\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) &= P(a < N(0,1) \le b) \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx\end{aligned}$$

or equivalently put

$$\lim_{n\to\infty} P\left(n\mu + \sigma\sqrt{n}a < S_n \le n\mu + \sigma\sqrt{n}b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

More intuitively, we have

$$S_n \overset{d}{\cong} n\mu + \sqrt{n}\sigma N(0,1) \overset{d}{=} N\left(n\mu, n\sigma^2\right).$$

*Example 23.25.* Suppose that $P(X_n = n) = 1$ for all $n$, then $F_n(y) = 1_{y\ge n} \to 0 = F(y)$ as $n \to \infty$. Notice that $F$ is not a distribution function because all of the mass went off to $+\infty$. Similarly, if we suppose, $P(X_n = \pm n) = \frac{1}{2}$ for all $n$, then $F_n = \frac{1}{2}1_{[-n,n)} + 1_{[n,\infty)} \to \frac{1}{2} = F(y)$ as $n \to \infty$. Again, $F$ is not a distribution function on $\mathbb{R}$ since half the mass went to $-\infty$ while the other half went to $+\infty$.

*Example 23.26.* Suppose $X$ is a non-zero random variables such that $X \overset{d}{=} -X$, then $X_n := (-1)^n X \overset{d}{=} X$ for all $n$ and therefore, $X_n \implies X$ as $n \to \infty$. On the other hand, $X_n$ does not converge to $X$ almost surely or in probability.

**Lemma 23.27.** *Suppose $X$ is a random variable, $\{c_n\}_{n=1}^\infty \subset \mathbb{R}$, and $X_n = X + c_n$. If $c := \lim_{n\to\infty} c_n$ exists, then $X_n \implies X + c$.*

**Proof.** Let $F(x) := P(X \le x)$ and

$$F_n(x) := P(X_n \le x) = P(X + c_n \le x) = F(x - c_n).$$

Clearly, if $c_n \to c$ as $n \to \infty$, then for all $x \in \mathcal{C}(F(\cdot - c))$ we have $F_n(x) \to F(x-c)$. Since $F(x-c) = P(X + c \le x)$, we see that $X_n \implies X + c$. Observe that $F_n(x) \to F(x-c)$ only for $x \in \mathcal{C}(F(\cdot - c))$ but this is sufficient to assert $X_n \implies X + c$. ∎

**Lemma 23.28.** *Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of random variables on a common probability space and $c \in \mathbb{R}$. Then $X_n \implies c$ iff $X_n \overset{P}{\to} c$.*

**Proof.** Recall that $X_n \overset{P}{\to} c$ iff for all $\varepsilon > 0$, $P(|X_n - c| > \varepsilon) \to 0$. Since

$$\{|X_n - c| > \varepsilon\} = \{X_n > c + \varepsilon\} \cup \{X_n < c - \varepsilon\}$$

it follows $X_n \overset{P}{\to} c$ iff $P(X_n > x) \to 0$ for all $x > c$ and $P(X_n < x) \to 0$ for all $x < c$. These conditions are also equivalent to $P(X_n \le x) \to 1$ for all $x > c$ and $P(X_n \le x) \le P(X_n < x') \to 0$ for all $x < c$ (where $x < x' < c$). So $X_n \overset{P}{\to} c$ iff

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 \text{ if } x < c \\ 1 \text{ if } x > c \end{cases} = F(x)$$

where $F(x) = P(c \le x) = 1_{x\ge c}$. Since $\mathcal{C}(F) = \mathbb{R} \setminus \{c\}$, we have shown $X_n \overset{P}{\to} c$ iff $X_n \implies c$.

**Alternative proof using Theorem** 23.32 below. For the implication that $X_n \overset{P}{\to} c$ implies $X_n \implies c$, see Corollary 23.35. Conversely if $X_n \implies c$ and $\varepsilon > 0$ let $f_\varepsilon \in C(\mathbb{R}, [0,1])$ such that $f_\varepsilon(c) = 0$ and $f_\varepsilon(x) = 1$ for $|x - c| \ge \varepsilon$. (A simple piecewise linear function will do here.) Then

$$P(|X_n - c| \ge \varepsilon) \le \mathbb{E}[f_\varepsilon(X_n)] \to \mathbb{E}[f(c)] = 0 \text{ as } n \to \infty.$$

∎

**Notation 23.29** *Given a proper distribution function, $F : \mathbb{R} \to [0,1]$, let $Y = F^{\leftarrow} : (0,1) \to \mathbb{R}$ be the function defined by*

$$Y(x) = F^{\leftarrow}(x) = \sup\{y \in \mathbb{R} : F(y) < x\}.$$

*Similarly, let*

$$Y^+(x) := \inf\{y \in \mathbb{R} : F(y) > x\}.$$

We will need the following simple observations about $Y$ and $Y^+$ which are easily understood from Figure 23.3.

1. $Y(x) \le Y^+(x)$ and $Y(x) < Y^+(x)$ iff $x$ is the height of a "flat spot" of $F$.
2. The set,
$$E := \left\{x \in (0,1) : Y(x) < Y^+(x)\right\}, \tag{23.21}$$
of flat spot heights is at most countable. This is because, $\left\{(Y(x), Y^+(x))\right\}_{x\in E}$ is a collection of pairwise disjoint intervals which is necessarily countable. (Each such interval contains a rational number.)

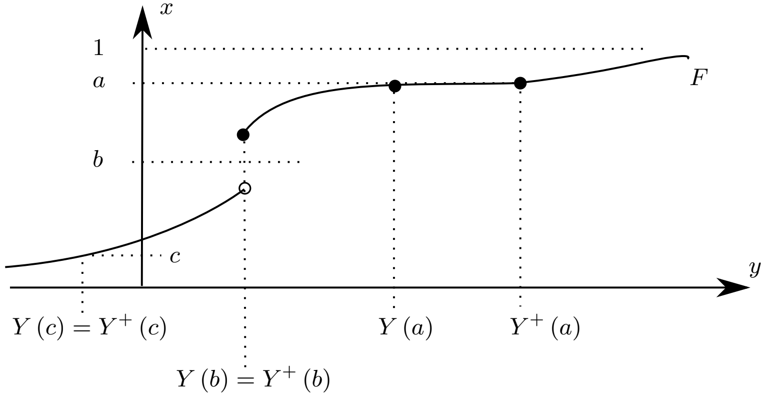



**Fig. 23.3.** The functions $Y$ and $Y^+$ associated to $F$.

3. The following inequality holds,
$$F\left(Y\left(x\right)-\right) \leq x \leq F\left(Y\left(x\right)\right) \text{ for all } x \in \left(0,1\right). \tag{23.22}$$
Indeed, if $y > Y\left(x\right)$, then $F\left(y\right) \geq x$ and by right continuity of $F$ it follows that $F\left(Y\left(x\right)\right) \geq x$. Similarly, if $y < Y\left(x\right)$, then $F\left(y\right) < x$ and hence $F\left(Y\left(x\right)-\right) \leq x$.
4. $\{x \in (0,1) : Y(x) \leq y_0\} = (0, F(y_0)] \cap (0,1)$. To prove this assertion first suppose that $Y(x) \leq y_0$, then according to Eq. (23.22) we have $x \leq F(Y(x)) \leq F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely, if $x \in (0,1)$ and $x \leq F(y_0)$, then $Y(x) \leq y_0$ by definition of $Y$.
5. As a consequence of item 4. we see that $Y$ is $\mathcal{B}_{(0,1)}/\mathcal{B}_{\mathbb{R}}$ – measurable and $m \circ Y^{-1} = F$, where $m$ is Lebesgue measure on $\left((0,1), \mathcal{B}_{(0,1)}\right)$.

**Theorem 23.30 (Baby Skorohod Theorem).** *Suppose that* $\{F_n\}_{n=0}^\infty$ *is a collection of distribution functions such that* $F_n \implies F_0$. *Then there exists a probability space,* $(\Omega, \mathcal{B}, P)$ *and random variables,* $\{Y_n\}_{n=1}^\infty$ *such that* $P(Y_n \leq y) = F_n(y)$ *for all* $n \in \mathbb{N} \cup \{\infty\}$ *and* $\lim_{n\to\infty} Y_n = Y$ *a.s..*

**Proof.** We will take $\Omega := (0,1)$, $\mathcal{B} = \mathcal{B}_{(0,1)}$, and $P = m$ – Lebesgue measure on $\Omega$ and let $Y_n := F_n^{\leftarrow}$ and $Y := F_0^{\leftarrow}$ as in Notation 23.29. Because of the above comments, $P(Y_n \leq y) = F_n(y)$ and $P(Y \leq y) = F_0(y)$ for all $y \in \mathbb{R}$. So in order to finish the proof it suffices to show, $Y_n(x) \to Y(x)$ for all $x \notin E$, where $E$ is the countable null set defined as in Eq. (23.21).

We now suppose $x \notin E$. If $y \in \mathcal{C}(F_0)$ with $y < Y(x)$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) < x$ and in particular, $F_n(y) < x$ for almost all $n$. This implies that $Y_n(x) \geq y$ for a.a. $n$ and hence that $\liminf_{n\to\infty} Y_n(x) \geq y$. Letting $y \uparrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies
$$\liminf_{n\to\infty} Y_n(x) \geq Y(x).$$

Similarly, for $x \notin E$ and $y \in \mathcal{C}(F_0)$ with $Y(x) = Y^+(x) < y$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) > x$ and in particular, $F_n(y) > x$ for almost all $n$. This implies that $Y_n(x) \leq y$ for a.a. $n$ and hence that $\limsup_{n\to\infty} Y_n(x) \leq y$. Letting $y \downarrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies
$$\limsup_{n\to\infty} Y_n(x) \leq Y(x).$$

Hence we have shown, for $x \notin E$, that
$$\limsup_{n\to\infty} Y_n(x) \leq Y(x) \leq \liminf_{n\to\infty} Y_n(x)$$
which shows
$$\lim_{n\to\infty} F_n^{\leftarrow}(x) = \lim_{n\to\infty} Y_n(x) = Y(x) = F^{\leftarrow}(x) \text{ for all } x \notin E. \tag{23.23}$$
■

In preparation for the full version of Skorohod's Theorem 23.86 it will be useful to record a special case of Theorem 23.30 which has both a stronger hypothesis and a stronger conclusion.

**Theorem 23.31 (Prenatal Skorohod Theorem).** *Suppose* $S = \{1, 2, \dots, m\} \subset \mathbb{R}$ *and* $\{\mu_n\}_{n=1}^\infty$ *is a sequence of probabilities on* $S$ *such that* $\mu_n \implies \mu$ *for some probability* $\mu$ *on* $S$. *Let* $P := \mu \otimes m$ *on* $\Omega := S \times (0,1]$, $Y(i,\theta) = i$ *for all* $(i,\theta) \in \Omega$. *Then there exists* $Y_n : \Omega \to S$ *such that* $\mathrm{Law}_P(Y_n) = \mu_n$ *for all* $n$ *and* $Y_n(i,\theta) = i$ *if* $\theta \leq \mu_n(i)/\mu(i)$ *where we take* $0/0 = 1$ *in this expression. In particular,* $\lim_{n\to\infty} Y_n(i,\theta) = Y(i,\theta)$ *a.s.*

**Proof.** The main point is to show for any probability measure, $\nu$, on $S$ there exists $Y_\nu : \Omega \to S$ such that $Y_\nu(i,\theta) = i$ when $\theta \leq \nu(i)/\mu(i)$ and $\mathrm{Law}_P(Y_\nu) = \nu$. If we can do this then we need only take $Y_n = Y_{\mu_n}$ for all $n$ to complete the proof.

In the proof to follow we will use the simple observation that for any $a \in (0,1)$ and $\alpha_i \geq 0$ with $\sum_{i=1}^m \alpha_i = 1$, then there exists a partition, $\{J_i\}_{i=1}^m$ of $(a,1]$ such that $m(J_i) = \alpha_i m((a,1]) = \alpha_i(1-a)$ – simply take $J_i = (a_{i-1}, a_i]$ where $a_0 = a$ and $a_i = \left(\sum_{j\leq i} \alpha_i\right) a$ for $1 \leq i \leq m$.

Let $\nu$ be any probability on $S$ and let
$$A_i := \{i\} \times \left(0, \frac{\nu(i)}{\mu(i)} \wedge 1\right]$$
and

$$C = \Omega \setminus \left(\sum_{i=1}^m A_i\right) = \sum_{i=1}^m \{i\} \times \left(\frac{\nu(i)}{\mu(i)} \wedge 1, 1\right]$$

and observe that

$$P(A_i) = \mu(i) \cdot \left(\frac{\nu(i)}{\mu(i)} \wedge 1\right) = \nu(i) \wedge \mu(i).$$

Using the observation in the previous paragraph we may write $\{k\} \times \left(\frac{\nu(k)}{\mu(k)} \wedge 1, 1\right] = \sum_{i=1}^m C_{k,i}$ with

$$P(C_{k,i}) = \alpha_i \cdot P\left(\{k\} \times \left(\frac{\nu(k)}{\mu(k)} \wedge 1, 1\right]\right).$$

The sets $C_i := \sum_{k=1}^m C_{k,i}$ then form a partition of $C$ such that $P(C_i) = \alpha_i P(C)$ for all $i$.

We now define

$$Y_\nu(i, \theta) := \sum_{i=1}^m i 1_{A_i \cup C_i}$$

so that $Y_\nu = i$ on $A_i$ and in particular $Y_\nu(i, \theta) = i$ when $\theta \le v(i)/\mu(i)$.

To finish the proof we need only choose the $\{\alpha_i\}_{i=1}^m$ so that $P(Y_\nu = i) = \nu(i)$ for all $i$, i.e. we must require,

$$\begin{aligned} \nu(i) = P(Y_\nu = i) &= P(A_i \cup C_i) = P(A_i) + \alpha_i P(C) \\ &= \nu(i) \wedge \mu(i) + \alpha_i P(C) \end{aligned} \tag{23.24}$$

and therefore we must define

$$\alpha_i = (\nu(i) - \nu(i) \wedge \mu(i)) / P(C) \ge 0.$$

To see this is an admissible choice (i.e. $\sum_{i=1}^m \alpha_i = 1$) notice that

$$\begin{aligned} P(C) &= \sum_i [\mu(i) - \nu(i) \wedge \mu(i)] \\ &= \sum_{\nu(i) < \mu(i)} (\mu(i) - \nu(i)) = \sum_{\mu(i) \le \nu(i)} (\nu(i) - \mu(i)), \end{aligned} \tag{23.25}$$

wherein we have used the fact that

$$\sum_{i \in S} (\mu(i) - \nu(i)) = 1 - 1 = 0.$$

Making use of these identities we find,

$$\sum_{i \in S} \alpha_i = \frac{1}{P(C)} \sum_{\mu(i) \le \nu(i)} (\nu(i) - \mu(i)) = 1.$$

■

The next theorem summarizes a number of useful equivalent characterizations of weak convergence. (The reader should compare Theorem 23.32 with Corollary 23.9.) In this theorem we will write $BC(\mathbb{R})$ for the bounded continuous functions, $f : \mathbb{R} \to \mathbb{R}$ (or $f : \mathbb{R} \to \mathbb{C}$) and $C_c(\mathbb{R})$ for those $f \in C(\mathbb{R})$ which have compact support, i.e. $f(x) \equiv 0$ if $|x|$ is sufficiently large.

**Theorem 23.32.** *Suppose that $\{\mu_n\}_{n=0}^\infty$ is a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and for each $n$, let $F_n(y) := \mu_n((-\infty, y])$ be the (proper) distribution function associated to $\mu_n$. Then the following are equivalent.*

1. *For all $f \in BC(\mathbb{R})$,*
$$\int_\mathbb{R} f d\mu_n \to \int_\mathbb{R} f d\mu_0 \text{ as } n \to \infty. \tag{23.26}$$
2. *Eq. (23.26) holds for all $f \in BC(\mathbb{R})$ which are uniformly continuous.*
3. *Eq. (23.26) holds for all $f \in C_c(\mathbb{R})$.*
4. *$F_n \implies F$.*
5. *There exists a probability space $(\Omega, \mathcal{B}, P)$ and random variables, $Y_n$, on this space such that $P \circ Y_n^{-1} = \mu_n$ for all $n$ and $Y_n \to Y_0$ a.s.*

**Proof.** Clearly 1. $\implies$ 2. $\implies$ 3. and 5. $\implies$ 1. by the dominated convergence theorem. Indeed, we have

$$\int_\mathbb{R} f d\mu_n = \mathbb{E}[f(Y_n)] \overset{\text{D.C.T.}}{\to} \mathbb{E}[f(Y)] = \int_\mathbb{R} f d\mu_0$$

for all $f \in BC(\mathbb{R})$. The implication that 4. $\implies$ 5. is Skorohod's Theorem 23.30 above. Therefore it suffices to prove 3. $\implies$ 4.

(3. $\implies$ 4.) Let $-\infty < a < b < \infty$ with $a, b \in \mathcal{C}(F_0)$ and for $\varepsilon > 0$, let $f_\varepsilon(x) \ge 1_{(a,b]}$ and $g_\varepsilon(x) \le 1_{(a,b]}$ be the functions in $C_c(\mathbb{R})$ pictured in Figure 23.4. Then

$$\limsup_{n\to\infty} \mu_n((a,b]) \le \limsup_{n\to\infty} \int_\mathbb{R} f_\varepsilon d\mu_n = \int_\mathbb{R} f_\varepsilon d\mu_0 \tag{23.27}$$

and

$$\liminf_{n\to\infty} \mu_n((a,b]) \ge \liminf_{n\to\infty} \int_\mathbb{R} g_\varepsilon d\mu_n = \int_\mathbb{R} g_\varepsilon d\mu_0. \tag{23.28}$$

Since $f_\varepsilon \to 1_{[a,b]}$ and $g_\varepsilon \to 1_{(a,b)}$ as $\varepsilon \downarrow 0$, we may use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in Eqs. (23.27) and (23.28) to conclude,

$$\limsup_{n\to\infty} \mu_n((a,b]) \le \mu_0([a,b]) = \mu_0((a,b])$$

and

$$\liminf_{n\to\infty} \mu_n((a,b]) \ge \mu_0((a,b)) = \mu_0((a,b]),$$

where the second equality in each of the equations holds because $a$ and $b$ are points of continuity of $F_0$. Hence we have shown that $\lim_{n\to\infty} \mu_n((a,b])$ exists and is equal to $\mu_0((a,b])$.

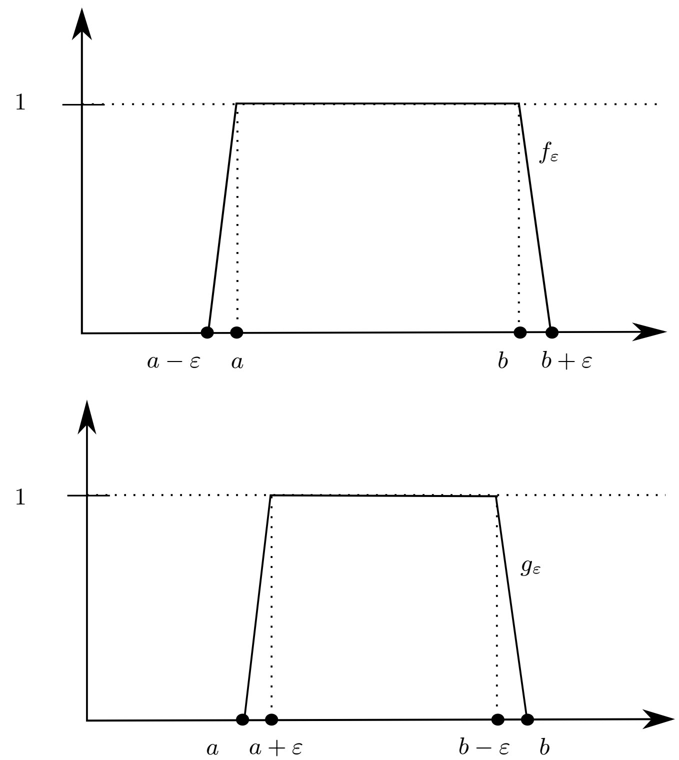



**Fig. 23.4.** The picture definition of the trapezoidal functions, $f_\varepsilon$ and $g_\varepsilon$.

■

*Example 23.33.* Suppose that $\{\mu_n\}_{n=1}^\infty$ and $\mu$ are measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\lim_{n\to\infty} d_{TV}(\mu_n, \mu) = 0$, then $\mu_n \implies \mu$. To prove this simply observe that for $f \in BC(\mathbb{R})$ we have by Corollary 23.9 that

$$|\mu(f) - \mu_n(f)| \le 2\|f\|_u d_{TV}(\mu_n, \mu) \to 0 \text{ as } n \to \infty.$$

Alternatively, simply note that

$$|F_n(x) - F(x)| = |\mu_n((-\infty,x]) - \mu((-\infty,x])| \le d_{TV}(\mu_n,\mu) \to 0 \text{ as } n\to\infty$$

for all $x \in \mathbb{R}$ and in particular for all $x \in \mathcal{C}(F)$.

**Proposition 23.34.** *Suppose that $\{\mu_n\}_{n=0}^\infty$ are measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu_n(\mathbb{Z}) = 1$ for all $n$. Then $\mu_n \implies \mu$ iff $\lim_{n\to\infty} d_{TV}(\mu_n, \mu) = 0$.*

**Proof.** In light of Example 23.33 we need only show the forward implication. For $m \in \mathbb{Z}$, let $f_m \in C(\mathbb{R}, [0,1])$ such that $f_m(m) = 1$ and $f_m(x) = 0$ for $|x-m| \ge 1$. Then $\mu_n \implies \mu$ implies

$$\mu_n(\{m\}) = \int_\mathbb{R} f_m d\mu_n \to \int_\mathbb{R} f_m d\mu = \mu(\{m\})$$

for all $m \in \mathbb{Z}$. Now apply the results of Exercise 23.5. ■

**Corollary 23.35.** *Suppose that $\{X_n\}_{n=0}^\infty$ is a sequence of random variables, such that $X_n \xrightarrow{P} X_0$, then $X_n \implies X_0$. (Recall that Example 23.26 shows the converse is in general false.)*

**Proof.** Let $g \in BC(\mathbb{R})$, then by Corollary 14.12, $g(X_n) \xrightarrow{P} g(X_0)$ and since $g$ is bounded, we may apply the dominated convergence theorem (see Corollary 14.9) to conclude that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X_0)]$. ■

We end this section with a few more equivalent characterizations of weak convergence. The combination of Theorem 23.32 and 23.36 is often called the Portmanteau[5] Theorem. A review of the notions of closure, interior, and boundary of a set $A$ which are used in the next theorem may be bound in Subsection 23.10.1 below.

**Theorem 23.36 (The Baby Portmanteau Theorem).** *Suppose $\{F_n\}_{n=0}^\infty$ are proper distribution functions. (Recall that we are denoting $\mu_{F_n}(A)$ simply by $F_n(A)$ for all $A \in \mathcal{B}_\mathbb{R}$.) Then the following are equivalent.*

1. $F_n \implies F_0$.
2. $\liminf_{n\to\infty} F_n(U) \ge F_0(U)$ *for open subsets,* $U \subset \mathbb{R}$.
3. $\limsup_{n\to\infty} F_n(C) \le F_0(C)$ *for all closed subsets,* $C \subset \mathbb{R}$.
4. $\lim_{n\to\infty} F_n(A) = F_0(A)$ *for all* $A \in \mathcal{B}_\mathbb{R}$ *such that* $F_0(\mathrm{bd}(A)) = 0$.

[5] Portmanteua: 1) A new word formed by joining two others and combining their meanings, or 2) A large travelling bag made of stiff leather.

**Proof.** (1. $\implies$ 2.) By Skorohod's Theorem 23.30 we may choose random variables, $Y_n$, such that $P(Y_n \le y) = F_n(y)$ for all $y \in \mathbb{R}$ and $n \in \mathbb{N}$ and $Y_n \to Y_0$ a.s. as $n \to \infty$. Since $U$ is open, it follows that

$$1_U(Y) \le \liminf_{n\to\infty} 1_U(Y_n) \text{ a.s.}$$

and so by Fatou's lemma,

$$\begin{aligned} F(U) &= P(Y \in U) = \mathbb{E}[1_U(Y)] \\ &\le \liminf_{n\to\infty} \mathbb{E}[1_U(Y_n)] = \liminf_{n\to\infty} P(Y_n \in U) = \liminf_{n\to\infty} F_n(U). \end{aligned}$$

(2. $\iff$ 3.) This follows from the observations: 1) $C \subset \mathbb{R}$ is closed iff $U := C^c$ is open, 2) $F(U) = 1 - F(C)$, and 3) $\liminf_{n\to\infty}(-F_n(C)) = -\limsup_{n\to\infty} F_n(C)$.

(2. and 3. $\iff$ 4.) If $F_0(\mathrm{bd}(A)) = 0$, then $A^o \subset A \subset \bar{A}$ with $F_0(\bar{A} \setminus A^o) = F_0(\mathrm{bd}(A)) = 0$. Therefore

$$F_0(A) = F_0(A^o) \le \liminf_{n\to\infty} F_n(A^o) \le \limsup_{n\to\infty} F_n(\bar{A}) \le F_0(\bar{A}) = F_0(A).$$

(4. $\implies$ 1.) Let $a, b \in \mathcal{C}(F_0)$ and take $A := (a, b]$. Then $F_0(\mathrm{bd}(A)) = F_0(\{a, b\}) = 0$ and therefore, $\lim_{n\to\infty} F_n((a,b]) = F_0((a,b])$, i.e. $F_n \implies F_0$. ■

**Exercise 23.7.** Suppose that $F$ is a continuous proper distribution function. Show,

1. $F : \mathbb{R} \to [0,1]$ is uniformly continuous.
2. If $\{F_n\}_{n=1}^\infty$ is a sequence of distribution functions converging weakly to $F$, then $F_n$ converges to $F$ uniformly on $\mathbb{R}$, i.e.

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F(x) - F_n(x)| = 0.$$

In particular, it follows that

$$\begin{aligned} &\sup_{a<b} |\mu_F((a,b]) - \mu_{F_n}((a,b])| \\ &\quad = \sup_{a<b} |F(b) - F(a) - (F_n(b) - F_n(a))| \\ &\quad \le \sup_b |F(b) - F_n(b)| + \sup_a |F_n(a) - F_n(a)| \to 0 \text{ as } n \to \infty. \end{aligned}$$

**Hints for part 2.** Given $\varepsilon > 0$, show that there exists, $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_n = \infty$, such that $|F(\alpha_{i+1}) - F(\alpha_i)| \le \varepsilon$ for all $i$. Now show, for $x \in [\alpha_i, \alpha_{i+1})$, that

$$\begin{aligned} &|F(x) - F_n(x)| \\ &\quad \le (F(\alpha_{i+1}) - F(\alpha_i)) + |F(\alpha_i) - F_n(\alpha_i)| + (F_n(\alpha_{i+1}) - F_n(\alpha_i)). \end{aligned}$$

Most of the results above generalize to the case where $\mathbb{R}$ is replaced by a complete separable metric space as described in Section 23.10 below. The definition of weak convergence in this generality is as follows.

**Definition 23.37 (Weak convergence).** *Let $(S, \rho)$ be a metric space. A sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ is said to converge weakly to a probability $\mu$ if $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ for every $f \in BC(S)$.*[6] *We will write this convergence as $\mu_n \implies \mu$ or $\mu_n \xrightarrow{w} \mu$ as $n \to \infty$.*

As a warm up to these general results and compactness results to come, let us consider in more detail the case where $S = \mathbb{R}^d$.

**Proposition 23.38.** *Suppose that $\{\mu_n\}_{n=1}^\infty \cup \{\mu\}$ are probability measures on $(S := \mathbb{R}^d, \mathcal{B} = \mathcal{B}_{\mathbb{R}^d})$ such that $\mu(f) = \lim_{n\to\infty} \mu_n(f)$ for all $f \in C_c^\infty(S)$ then $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ for all $f \in C_c(S)$.*

**Proof.** Let $\rho \in C_c^\infty(S)$ such that $0 \le \rho \le 1_{C_1}$ and $\int_S \rho(z)\,dz = 1$. For $f \in C_c(S)$ and $\varepsilon > 0$, let

$$f_\varepsilon(x) := \int_S f(x + \varepsilon z)\rho(z)\,dz. \tag{23.29}$$

It then follows that

$$\begin{aligned} M_\varepsilon := \max_x |f(x) - f_\varepsilon(x)| &= \max_x \left| \int_S [f(x) - f(x + \varepsilon z)]\rho(z)\,dz \right| \\ &\le \max_x \int_S |f(x) - f(x + \varepsilon z)|\rho(z)\,dz \\ &\le \max_x \max_{|z|\le\varepsilon} |f(x) - f(x+z)| \end{aligned}$$

where the latter expression goes to zero as $\varepsilon \downarrow 0$ by the uniform continuity of $f$. Thus we have shown that $f_\varepsilon \to f$ uniformly in $x$ as $\varepsilon \downarrow 0$. Making the change of variables $y = x + \varepsilon z$ in Eq. (23.29) shows

$$f_\varepsilon(x) := \frac{1}{\varepsilon^d} \int_S f(y)\rho\left(\frac{y - x}{\varepsilon}\right) dy$$

from which it follows that $f_\varepsilon$ is smooth. Using this information we find,

$$\begin{aligned} &\limsup_{n\to\infty} |\mu(f) - \mu_n(f)| \\ &\quad \le \limsup_{n\to\infty} [|\mu(f) - \mu(f_\varepsilon)| + |\mu(f_\varepsilon) - \mu_n(f_\varepsilon)| + |\mu_n(f_\varepsilon) - \mu_n(f)|] \\ &\quad \le 2M_\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0. \end{aligned}$$

■

[6] This is actually "**weak-* convergence**" when viewing $\mu_n \in BC(S)^*$.

**Theorem 23.39.** *Suppose that* $\{\mu_n\}_{n=1}^{\infty} \cup \{\mu\}$ *are probability measures on* $\left(S := \mathbb{R}^d, \mathcal{B} = \mathcal{B}_{\mathbb{R}^d}\right)$ *(or some other locally compact Hausdorff space) such that* $\mu(f) = \lim_{n\to\infty} \mu_n(f)$ *for all* $f \in C_c(S)$*, then;*

1. *For all* $\varepsilon > 0$ *there exists a compact set* $K_\varepsilon \subset S$ *such that* $\mu(K_\varepsilon) \ge 1 - \varepsilon$ *and* $\mu_n(K_\varepsilon) \ge 1 - \varepsilon$ *for all* $n \in \mathbb{N}$.
2. *If* $f \in BC(S)$*, then* $\lim_{n\to\infty} \mu_n(f) = \mu(f)$.

**Proof.** For all $R > 0$ let $C_R := \{x \in S : |x| \le R\}$ and then choose $\varphi_R \in C_c(S)$ such that $\varphi_R = 1$ on $C_{R/2}$ and $0 \le \varphi_R \le 1_{C_R}$.

1. With this notation it follows that
$$\mu_n(C_R) \ge \mu_n(\varphi_R) \to \mu(\varphi_R) \ge \mu\left(C_{R/2}\right).$$
Choose $R$ so large that $\mu_n(C_R) \ge \mu\left(C_{R/2}\right) \ge 1 - \varepsilon/2$. Then for $n \ge N_\varepsilon$ we will have $\mu_n(C_R) \ge 1 - \varepsilon$ for all $n \ge N_\varepsilon$. By increasing $R$ more if necessary we may also assume that $\mu_n(C_R) \ge 1 - \varepsilon$ for all $n < N_\varepsilon$. Taking $K_\varepsilon := C_R$ for this $R$ completes the proof of item 1.
2. Let $f \in BC(S)$ and for $R > 0$ let $f_R := \varphi_R \cdot f \in C_c(S)$. Then
$$\begin{aligned}\limsup_{n\to\infty} &|\mu(f) - \mu_n(f)| \\ &\le \limsup_{n\to\infty} \left[|\mu(f) - \mu(f_R)| + |\mu(f) - \mu_n(f_R)| + |\mu_n(f_R) - \mu_n(f)|\right] \\ &= |\mu(f) - \mu(f_R)| + \limsup_{n\to\infty} |\mu_n(f_R) - \mu_n(f)|. \qquad (23.30)\end{aligned}$$
By the dominated convergence theorem, $\lim_{R\to\infty} |\mu(f) - \mu(f_R)| = 0$. For the second term if $M = \max_{x\in S} |f(x)|$ we will have
$$\begin{aligned}\sup_n |\mu_n(f_R) - \mu_n(f)| &\le \sup_n \mu_n(|f_R - f|) \\ &\le M \cdot \sup_n \mu_n(\varphi_R \ne 1) \le M \cdot \sup_n \mu_n\left(S \setminus C_{R/2}\right).\end{aligned}$$
However, by item 1. it follows that $\lim_{R\to\infty} \sup_n \mu_n\left(S \setminus C_{R/2}\right) = 0$. Therefore letting $R \to \infty$ in Eq. (23.30) show that $\limsup_{n\to\infty} |\mu(f) - \mu_n(f)| = 0$.

■

## 23.5 "Derived" Weak Convergence

**Lemma 23.40.** *Let* $(X, d)$ *be a metric space,* $f : X \to \mathbb{R}$ *be a function, and* $\mathcal{D}(f)$ *be the set of* $x \in X$ *where* $f$ *is discontinuous at* $x$*. Then* $\mathcal{D}(f)$ *is a Borel measurable subset of* $X$.

**Proof.** For $x \in X$ and $\delta > 0$, let $B_x(\delta) = \{y \in X : d(x, y) < \delta\}$. Given $\delta > 0$, let $f^\delta : X \to \mathbb{R} \cup \{\infty\}$ be defined by,
$$f^\delta(x) := \sup_{y \in B_x(\delta)} f(y).$$
We will begin by showing $f^\delta$ is **lower semi-continuous**, i.e. $\left\{f^\delta \le a\right\}$ is closed (or equivalently $\left\{f^\delta > a\right\}$ is open) for all $a \in \mathbb{R}$. Indeed, if $f^\delta(x) > a$, then there exists $y \in B_x(\delta)$ such that $f(y) > a$. Since this $y$ is in $B_{x'}(\delta)$ whenever $d(x, x') < \delta - d(x, y)$ (because then, $d(x', y) \le d(x, y) + d(x, x') < \delta$) it follows that $f^\delta(x') > a$ for all $x' \in B_x(\delta - d(x, y))$. This shows $\left\{f^\delta > a\right\}$ is open in $X$.

We similarly define $f_\delta : X \to \mathbb{R} \cup \{-\infty\}$ by
$$f_\delta(x) := \inf_{y \in B_x(\delta)} f(y).$$
Since $f_\delta = -(-f)^\delta$, it follows that
$$\{f_\delta \ge a\} = \left\{(-f)^\delta \le -a\right\}$$
is closed for all $a \in \mathbb{R}$, i.e. $f_\delta$ is **upper semi-continuous**. Moreover, $f_\delta \le f \le f^\delta$ for all $\delta > 0$ and $f^\delta \downarrow f^0$ and $f_\delta \uparrow f_0$ as $\delta \downarrow 0$, where $f_0 \le f \le f^0$ and $f_0 : X \to \mathbb{R} \cup \{-\infty\}$ and $f^0 : X \to \mathbb{R} \cup \{\infty\}$ are measurable functions. The proof is now complete since it is easy to see that
$$\mathcal{D}(f) = \left\{f^0 > f_0\right\} = \left\{f^0 - f_0 \ne 0\right\} \in \mathcal{B}_X.$$

■

*Remark 23.41.* Suppose that $x_n \to x$ with $x \in \mathcal{C}(f) := \mathcal{D}(f)^c$. Then $f(x_n) \to f(x)$ as $n \to \infty$.

**Theorem 23.42 (Continuous Mapping Theorem).** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a Borel measurable function. If* $X_n \implies X_0$ *and* $P(X_0 \in \mathcal{D}(f)) = 0$*, then* $f(X_n) \implies f(X_0)$*. If in addition,* $f$ *is bounded,* $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X_0)$*. (This result generalizes easily to the case where* $f : S \to T$ *is a Borel measurable function between metric spaces and* $X_n, X_0$ *are not* $S$ *– valued random functions.)*

**Proof.** Let $\{Y_n\}_{n=0}^{\infty}$ be random variables on some probability space as in Theorem 23.30. For $g \in BC(\mathbb{R})$ we observe that $\mathcal{D}(g \circ f) \subset \mathcal{D}(f)$ and therefore,
$$P(Y_0 \in \mathcal{D}(g \circ f)) \le P(Y_0 \in \mathcal{D}(f)) = P(X_0 \in \mathcal{D}(f)) = 0.$$
Hence it follows that $g \circ f \circ Y_n \to g \circ f \circ Y_0$ a.s. So an application of the dominated convergence theorem (see Corollary 14.9) implies

$$\mathbb{E}\left[g\left(f\left(X_n\right)\right)\right]=\mathbb{E}\left[g\left(f\left(Y_n\right)\right)\right]\rightarrow\mathbb{E}\left[g\left(f\left(Y_0\right)\right)\right]=\mathbb{E}\left[g\left(f\left(X_0\right)\right)\right]. \tag{23.31}$$

This proves the first assertion. For the second assertion we take $g(x)=(x\wedge M)\vee(-M)$ in Eq. (23.31) where $M$ is a bound on $|f|$. ■

**Theorem 23.43 (Slutzky's Theorem).** *Suppose that $X_n \implies X\in\mathbb{R}^m$ and $Y_n \xrightarrow{P} c\in\mathbb{R}^n$ where $c\in\mathbb{R}^n$ is constant. Assuming all random vectors are on the same probability space we will have $(X_n,Y_n)\implies(X,c)$ – see Definition 23.37. In particular if $m=n$, by taking $f(x,y)=g(x+y)$ and $f(x,y)=h(x\cdot y)$ with $g\in BC(\mathbb{R}^n)$ and $h\in BC(\mathbb{R})$, we learn $X_n+Y_n \implies X+c$ and $X_n\cdot Y_n \implies X\cdot c$ respectively. (The first part of this theorem generalizes to metric spaces as well.)*

**Proof.** According to Theorem 23.39 it suffices to show for

$$\lim_{n\to\infty}\mathbb{E}\left[f\left(X_n,Y_n\right)\right]=\mathbb{E}\left[f\left(X,c\right)\right] \tag{23.32}$$

for all $f\in BC\left(\mathbb{R}^{m\times n}\right)$ which are uniformly continuous or even only $f\in C_c\left(\mathbb{R}^{m\times n}\right)$. For a uniformly continuous function we have for every $\varepsilon>0$ a $\delta:=\delta(\varepsilon)>0$ such that

$$|f(x,y)-f(x',y')|\le\varepsilon \text{ if } \|(x,y)-(x',y')\|\le\delta.$$

Then

$$\begin{aligned}\left|\mathbb{E}\left[f\left(X_n,Y_n\right)-f\left(X_n,c\right)\right]\right| &\le \mathbb{E}\left[\left|f\left(X_n,Y_n\right)-f\left(X_n,c\right)\right| : \|Y_n-c\|\le\delta\right]\\ &\quad+\mathbb{E}\left[\left|f\left(X_n,Y_n\right)-f\left(X_n,c\right)\right| : \|Y_n-c\|>\delta\right]\\ &\le \varepsilon+2MP\left(\|Y_n-c\|>\delta\right)\to\varepsilon \text{ as } n\to\infty,\end{aligned}$$

where $M=\sup|f|$. Since, $X_n \implies X$, we know $\mathbb{E}\left[f\left(X_n,c\right)\right]\to\mathbb{E}\left[f\left(X,c\right)\right]$ and hence we have shown,

$$\begin{aligned}&\limsup_{n\to\infty}\left|\mathbb{E}\left[f\left(X_n,Y_n\right)-f\left(X,c\right)\right]\right|\\ &\quad\le\limsup_{n\to\infty}\left|\mathbb{E}\left[f\left(X_n,Y_n\right)-f\left(X_n,c\right)\right]\right|+\limsup_{n\to\infty}\left|\mathbb{E}\left[f\left(X_n,c\right)-f\left(X,c\right)\right]\right|\le\varepsilon.\end{aligned}$$

As $\varepsilon>0$ was arbitrary this proves Eq. (23.32). ■

**Theorem 23.44 ($\delta-$method).** *Suppose that $\{X_n\}_{n=1}^\infty$ are random variables, $b\in\mathbb{R}$, $a_n\in\mathbb{R}\setminus\{0\}$ with $\lim_{n\to\infty}a_n=0$, and*

$$Y_n:=\frac{X_n-b}{a_n}\implies Z.$$

*If $g:\mathbb{R}\to\mathbb{R}$ be a measurable function which is differentiable at $b$, then*

$$\frac{g\left(X_n\right)-g\left(b\right)}{a_n}\implies g'\left(b\right)Z. \tag{23.33}$$

*Put more informally, if*

$$X_n\overset{d}{\simeq}b+a_nZ \text{ then } g\left(X_n\right)\overset{d}{\simeq}g\left(b\right)+g'\left(b\right)a_nZ. \tag{23.34}$$

**Proof.** Informally we have $X_n=a_nY_n+b\overset{d}{\cong}a_nZ+b$ and therefore

$$\frac{g\left(X_n\right)-g\left(b\right)}{a_n}\overset{d}{\cong}\frac{g\left(a_nZ+b\right)-g\left(b\right)}{a_nZ}Z\to g'\left(b\right)Z \text{ as } n\to\infty.$$

We now make the proof rigorous.

By Skorohod's Theorem 23.30 we may assume that $\{Y_n\}_{n=1}^\infty$ and $Z$ are on the same probability space and that $Y_n\to Z$ a.s. and we may take $X_n:=a_nY_n+b$. By the definition of the derivative of $g$ at $b$, we have

$$g\left(b+\Delta\right)-g\left(b\right)=g'\left(b\right)\Delta+\varepsilon\left(\Delta\right)\Delta$$

where $\varepsilon(\Delta)\to0$ as $\Delta\to0$. Taking $\Delta=a_nY_n$ in this equation shows

$$\begin{aligned}\frac{g\left(X_n\right)-g\left(b\right)}{a_n}&=\frac{g\left(a_nY_n+b\right)-g\left(b\right)}{a_n}\\ &=\frac{g'\left(b\right)a_nY_n+\varepsilon\left(a_nY_n\right)a_nY_n}{a_n}\to g'\left(b\right)Z \text{ a.s.}\end{aligned}$$

which implies Eq. (23.33) because of Corollary 23.35. ■

*Example 23.45.* Suppose that $\{U_n\}_{n=1}^\infty$ are i.i.d. random variables which are uniformly distributed on $[0,1]$ and let $Y_n:=\prod_{j=1}^n U_j^{\frac{1}{n}}$. Our goal is to find $a_n$ and $b_n$ such that $\frac{Y_n-b_n}{a_n}$ is weakly convergent to a non-constant random variable. To this end, let

$$X_n:=\ln Y_n=\frac{1}{n}\sum_{j=1}^n\ln U_j$$

Since

$$\begin{aligned}\mathbb{E}\left[\ln U_1\right]&=\int_0^1\ln x dx=-1,\\ \mathbb{E}\left[\ln U_1\right]^2&=\int_0^1\ln^2 x dx=2,\end{aligned}$$

$\mathrm{Var}(\ln U_1) = 1$ and so by the central limit theorem

$$\sqrt{n}\left[X_n - (-1)\right] = \frac{\sum_{j=1}^n \left[\ln U_j + 1\right]}{\sqrt{n}} \implies Z \overset{d}{=} N(0,1).$$

In other words, $\ln Y_n = X_n \overset{d}{\simeq} -1 + \frac{1}{\sqrt{n}} Z$ and so we expect,

$$Y_n \overset{d}{\simeq} e^{-1+\frac{1}{\sqrt{n}}Z} \simeq e^{-1}\left(1 + \frac{1}{\sqrt{n}} Z + \dots\right)$$

and thus we conjecture

$$\sqrt{n}\left[Y_n - e^{-1}\right] \implies e^{-1} Z \overset{d}{=} N\left(0, e^{-2}\right).$$

To verify this is correct recall that by the $\delta$ – method if $g'(-1)$ exits, then

$$g(X_n) \overset{d}{\simeq} g(-1) + g'(-1)\frac{1}{\sqrt{n}} Z.$$

Taking $g(x) = e^x$ then implies $Y_n \overset{d}{\simeq} e^{-1} + e^{-1}\frac{1}{\sqrt{n}} Z$ or more precisely,

$$\sqrt{n}\left[\prod_{j=1}^n U_j^{\frac{1}{n}} - e^{-1}\right] = \sqrt{n}\left(Y_n - e^{-1}\right) \implies e^{-1} Z = N\left(0, e^{-2}\right).$$

**Exercise 23.8.** Given a function, $f : X \to \mathbb{R}$ and a point $x \in X$, let

$$\liminf_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \inf_{y \in B'_x(\delta)} f(y) \text{ and} \tag{23.35}$$

$$\limsup_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \sup_{y \in B'_x(\delta)} f(y), \tag{23.36}$$

where

$$B'_x(\delta) := \{y \in X : 0 < d(x,y) < \delta\}.$$

Show $f$ is lower (upper) semi-continuous iff $\liminf_{y \to x} f(y) \geq f(x)$ $\left(\limsup_{y \to x} f(y) \leq f(x)\right)$ for all $x \in X$.

## 23.6 Convergence of Types

Given a sequence of random variables $\{X_n\}_{n=1}^\infty$ we often look for centerings $\{b_n\}_{n=1}^\infty \subset \mathbb{R}$ and scalings $\{a_n > 0\}_{n=1}^\infty$ such that there exists a non-constant random variable $Y$ such that

$$\frac{X_n - b_n}{a_n} \implies Y. \tag{23.37}$$

Assuming this can be done it is reasonable to ask how unique are the centering, scaling parameters, and the limiting distribution $Y$. To answer this question let us suppose there exists another collection of centerings $\{\beta_n\}_{n=1}^\infty \subset \mathbb{R}$ and scalings $\{\alpha_n > 0\}_{n=1}^\infty$ along with a non-constant random variable $Z$ such that Thus if

$$\frac{X_n - \beta_n}{\alpha_n} \implies Z. \tag{23.38}$$

Working informally we expect that

$$X_n \overset{d}{\cong} \alpha_n Z + \beta_n$$

and putting this expression back into Eq. (23.38) leads us to expect;

$$\frac{\alpha_n}{a_n} Z + \frac{\beta_n - b_n}{a_n} = \frac{\alpha_n Z + \beta_n - b_n}{a_n} \implies Y.$$

It is reasonable to expect that this can only happen if the limits

$$A = \lim_{n\to\infty} \frac{\alpha_n}{a_n} \in (0,\infty) \text{ and } B := \lim_{n\to\infty} \frac{\beta_n - b_n}{a_n} \tag{23.39}$$

exist and

$$Y \overset{d}{=} AZ + B. \tag{23.40}$$

Notice that $A > 0$ as both $Y$ and $Z$ are assumed to be non-constant. That these results are correct is the content of Theorem 23.49 below.

Let us now explain how to choose the $\{a_n\}$ and the $\{b_n\}$. Let $F_n(x) := P(X_n \leq x)$, then Eq. (23.37) states,

$$F_n(a_n y + b_n) = P(X_n \leq a_n y + b_n) = P\left(\frac{X_n - b_n}{a_n} \leq y\right) \implies P(Y \leq y).$$

Taking $y = 0$ and $y = 1$ in this equation leads us to expect,

$$\lim_{n\to\infty} F_n(b_n) = P(Y \leq 0) = \gamma_1 \in (0,1) \text{ and}$$
$$\lim_{n\to\infty} F_n(a_n + b_n) = P(Y \leq 1) = \gamma_2 \in (0,1).$$

In fact there is nothing so special about 0 and 1 in these equation for if $Y \overset{d}{=} AZ + B$ we will have $Z = A^{-1}(Y - B)$ and so

$$P(Y \leq 0) = P(AZ + B \leq 0) = P(Z \leq -B/A) \text{ and}$$
$$P(Y \leq 1) = P(AZ + B \leq 1) = P(Z \leq (1 - B)/A).$$

**Definition 23.46.** *Two random variables, $Y$ and $Z$, are said to be of the* ***same type*** *if there exists constants, $A > 0$ and $B \in \mathbb{R}$ such that Eq. (23.40) holds. Alternatively put, if $U(y) := P(Y \le y)$ and $V(z) := P(Z \le z)$, then $U$ and $V$ should satisfy,*

$$V(z) = P(Z \le z) = P(Y \le Az + B) = U(Az + B)$$

*for all $z \in \mathbb{R}$.*

*Remark 23.47.* Suppose that $Y \overset{d}{=} AZ + B$ and $Y$ and $Z$ are square integrable random variables. Then

$$\mathbb{E}Y = A \cdot \mathbb{E}Z + B \text{ and } \operatorname{Var}(Y) = A^2 \operatorname{Var}(Z)$$

from which it follows that $A^2 = \operatorname{Var}(Y) / \operatorname{Var}(Z)$ and $B = \mathbb{E}Y - A \cdot \mathbb{E}Z$. In particular, given $Y \in L^2(P)$ there is a unique $Z$ of the same type such that $\mathbb{E}Z = 0$ and $\operatorname{Var}(Z) = 1$. On these grounds it is often reasonable to try to choose $\{b_n\}$ and $\{a_n > 0\}$ so that $\bar{X}_n := a_n^{-1}(X_n - b_n)$ has mean zero and variance one.

We will need the following elementary observation for the proof of Theorem 23.49.

**Lemma 23.48.** *If $Y$ is non-constant (a.s.) random variable and $U(y) := P(Y \le y)$, then $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$ for all $\gamma_1$ sufficiently close to $0$ and $\gamma_2$ sufficiently close to $1$ – see Notation 23.29 for the meaning of $U^{\leftarrow}$.*

**Proof.** Observe that $Y$ is constant iff $U(y) = 1_{y \ge c}$ for some $c \in \mathbb{R}$, i.e. iff $U$ only takes on the values, $\{0,1\}$. So since $Y$ is not constant, there exists $y \in \mathbb{R}$ such that $0 < U(y) < 1$. Hence if $\gamma_2 > U(y)$ then $U^{\leftarrow}(\gamma_2) \ge y$ and if $\gamma_1 < U(y)$ then $U^{\leftarrow}(\gamma_1) \le y$. Moreover, if we suppose that $\gamma_1$ is not the height of a flat spot of $U$, then in fact, $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$. This inequality then remains valid as $\gamma_1$ decreases and $\gamma_2$ increases. ■

**Theorem 23.49 (Convergence of Types).** *Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables and $a_n, \alpha_n \in (0,\infty)$, $b_n, \beta_n \in \mathbb{R}$ are constants and $Y$ and $Z$ are non-constant random variables. Then*

1. *if both Eq. (23.37) and Eq. (23.38) hold then the limits, in Eq. (23.39) exists and $Y \overset{d}{=} AZ + B$ and in particular $Y$ and $Z$ are of the same type.*
2. *If the limits in Eq. (23.39) hold then either of the convergences in Eqs. (23.37) or (23.38) implies the others with $Z$ and $Y$ related by Eq. (23.40).*
3. *If there are some constants, $a_n > 0$ and $b_n \in \mathbb{R}$ and a non-constant random variable $Y$, such that Eq. (23.37) holds, then Eq. (23.38) holds using $\alpha_n$ and $\beta_n$ of the form,*

$$\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1) \text{ and } \beta_n := F_n^{\leftarrow}(\gamma_1) \tag{23.41}$$

*for some $0 < \gamma_1 < \gamma_2 < 1$. If the $F_n$ are invertible functions, Eq. (23.41) may be written as*

$$F_n(\beta_n) = \gamma_1 \text{ and } F_n(\alpha_n + \beta_n) = \gamma_2. \tag{23.42}$$

**Proof.** (2) Assume the limits in Eq. (23.39) hold. If Eq. (23.37) is satisfied, then by Slutsky's Theorem 15.23,

$$\begin{aligned} \frac{X_n - \beta_n}{\alpha_n} &= \frac{X_n - b_n + b_n - \beta_n}{a_n} \frac{a_n}{\alpha_n} \\ &= \frac{X_n - b_n}{a_n} \frac{a_n}{\alpha_n} - \frac{\beta_n - b_n}{a_n} \frac{a_n}{\alpha_n} \\ &\implies A^{-1}(Y - B) =: Z \end{aligned}$$

Similarly, if Eq. (23.38) is satisfied, then

$$\frac{X_n - b_n}{a_n} = \frac{X_n - \beta_n}{\alpha_n} \frac{\alpha_n}{a_n} + \frac{\beta_n - b_n}{a_n} \implies AZ + B =: Y.$$

(1) If $F_n(y) := P(X_n \le y)$, then

$$P\left(\frac{X_n - b_n}{a_n} \le y\right) = F_n(a_n y + b_n) \text{ and } P\left(\frac{X_n - \beta_n}{\alpha_n} \le y\right) = F_n(\alpha_n y + \beta_n).$$

By assumption we have

$$F_n(a_n y + b_n) \implies U(y) \text{ and } F_n(\alpha_n y + \beta_n) \implies V(y).$$

If $w := \sup\{y : F_n(a_n y + b_n) < x\}$, then $a_n w + b_n = F_n^{\leftarrow}(x)$ and hence

$$\sup\{y : F_n(a_n y + b_n) < x\} = \frac{F_n^{\leftarrow}(x) - b_n}{a_n}.$$

Similarly,

$$\sup\{y : F_n(\alpha_n y + \beta_n) < x\} = \frac{F_n^{\leftarrow}(x) - \beta_n}{\alpha_n}.$$

With these identities, it now follows from the proof of Skorohod's Theorem 23.30 (see Eq. (23.23)) that there exists an at most countable subset, $\Lambda$, of $(0,1)$ such that,

$$\begin{aligned} \frac{F_n^{\leftarrow}(x) - b_n}{a_n} &= \sup\{y : F_n(a_n y + b_n) < x\} \to U^{\leftarrow}(x) \text{ and} \\ \frac{F_n^{\leftarrow}(x) - \beta_n}{\alpha_n} &= \sup\{y : F_n(\alpha_n y + \beta_n) < x\} \to V^{\leftarrow}(x) \end{aligned}$$

for all $x \notin \Lambda$. Since $Y$ and $Z$ are not constants a.s., we can choose, by Lemma 23.48, $\gamma_1 < \gamma_2$ not in $\Lambda$ such that $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$ and $V^{\leftarrow}(\gamma_1) < V^{\leftarrow}(\gamma_2)$. In particular it follows that

$$\frac{F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)}{a_n} = \frac{F_n^{\leftarrow}(\gamma_2) - b_n}{a_n} - \frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1) > 0 \tag{23.43}$$

and similarly

$$\frac{F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)}{\alpha_n} \to V^{\leftarrow}(\gamma_2) - V^{\leftarrow}(\gamma_1) > 0.$$

Taking ratios of the last two displayed equations shows,

$$\frac{\alpha_n}{a_n} \to A := \frac{U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1)}{V^{\leftarrow}(\gamma_2) - V^{\leftarrow}(\gamma_1)} \in (0, \infty).$$

Moreover,

$$\frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to U^{\leftarrow}(\gamma_1) \text{ and} \tag{23.44}$$

$$\frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{a_n} = \frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{\alpha_n}\frac{\alpha_n}{a_n} \to AV^{\leftarrow}(\gamma_1)$$

and therefore,

$$\frac{\beta_n - b_n}{a_n} = \frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{a_n} - \frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to AV^{\leftarrow}(\gamma_1) - U^{\leftarrow}(\gamma_1) := B.$$

(3) Now suppose that we define $\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)$ and $\beta_n := F_n^{\leftarrow}(\gamma_1)$, then according to Eqs. (23.43) and (23.44)we have

$$\alpha_n / a_n \to U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1) \in (0, 1) \text{ and}$$
$$\frac{\beta_n - b_n}{a_n} \to U^{\leftarrow}(\gamma_1) \text{ as } n \to \infty.$$

Thus we may always center and scale the $\{X_n\}$ using $\alpha_n$ and $\beta_n$ of the form described in Eq. (23.41). ■

## 23.7 Weak Convergence Examples

*Example 23.50.* Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. $\exp(\lambda)$ – random variables, i.e. $X_n \geq 0$ a.s. and $P(X_n \geq x) = e^{-\lambda x}$ for all $x \geq 0$. In this case

$$F(x) := P(X_1 \leq x) = 1 - e^{-\lambda(x \vee 0)} = \left(1 - e^{-\lambda x}\right)_+.$$

Consider $M_n := \max(X_1, \ldots, X_n)$. We have, for $x \geq 0$ and $c_n \in (0, \infty)$ that

$$F_n(x) := P(M_n \leq x) = P\left(\cap_{j=1}^n \{X_j \leq x\}\right) = \prod_{j=1}^n P(X_j \leq x) = [F(x)]^n = \left(1 - e^{-\lambda x}\right)^n.$$

We now wish to find $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\frac{M_n - b_n}{a_n} \implies Y$.

1. To this end we note that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(M_n \leq a_n x + b_n) = F_n(a_n x + b_n) = [F(a_n x + b_n)]^n.$$

If we demand (c.f. Eq. (23.42) above)

$$P\left(\frac{M_n - b_n}{a_n} \leq 0\right) = F_n(b_n) = [F(b_n)]^n \to \gamma_1 \in (0, 1),$$

then $b_n \to \infty$ and we find

$$\ln \gamma_1 \sim n \ln F(b_n) = n \ln\left(1 - e^{-\lambda b_n}\right) \sim -ne^{-\lambda b_n}.$$

From this it follows that $b_n \sim \lambda^{-1} \ln n$. Given this, we now try to find $a_n$ by requiring,

$$P\left(\frac{M_n - b_n}{a_n} \leq 1\right) = F_n(a_n + b_n) = [F(a_n + b_n)]^n \to \gamma_2 \in (0, 1).$$

However, by what we have done above, this requires $a_n + b_n \sim \lambda^{-1} \ln n$. Hence we may as well take $a_n$ to be constant and for simplicity we take $a_n = 1$.

2. We now compute

$$\lim_{n\to\infty} P\left(M_n - \lambda^{-1} \ln n \leq x\right) = \lim_{n\to\infty}\left(1 - e^{-\lambda\left(x + \lambda^{-1}\ln n\right)}\right)^n = \lim_{n\to\infty}\left(1 - \frac{e^{-\lambda x}}{n}\right)^n = \exp\left(-e^{-\lambda x}\right).$$

The function $F(x) = \exp\left(-e^{-\lambda x}\right)$ is the CDF for a "Gumbel distribution," see Figure 23.5. Thus letting $Y$ be a random variable with this distribution (i.e. $P(Y \leq x) = \exp\left(-e^{-\lambda x}\right)$) we have shown $M_n - \frac{1}{\lambda}\ln n \implies Y$, i.e.

$$\max(X_1, \ldots, X_n) - \frac{1}{\lambda}\ln n \implies Y.$$

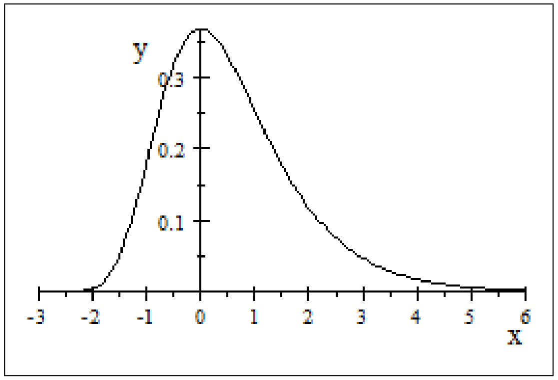



**Fig. 23.5.** Here is a plot of the density function for $Y$ when $\lambda = 1$.

*Example 23.51.* For $p \in (0,1)$, let $X_p$ denote the number of trials to get success in a sequence of independent trials with success probability $p$. Then $P(X_p > n) = (1-p)^n$ and therefore for $x > 0$,

$$P(pX_p > x) = P\left(X_p > \frac{x}{p}\right) = (1-p)^{\left[\frac{x}{p}\right]} = e^{\left[\frac{x}{p}\right]\ln(1-p)}$$
$$\sim e^{-p\left[\frac{x}{p}\right]} \to e^{-x} \text{ as } p \to 0.$$

Therefore $pX_p \implies T$ where $T \overset{d}{=} \exp(1)$, i.e. $P(T > x) = e^{-x}$ for $x \geq 0$ or alternatively, $P(T \leq y) = 1 - e^{-y \vee 0}$.

**Remarks on this example.** Let us see in a couple of ways where the appropriate centering and scaling of the $X_p$ come from in this example. For this let $q = 1-p$, then $P(X_p = n) = (1-p)^{n-1} p = q^{n-1}p$ for $n \in \mathbb{N}$. Also let

$$F_p(x) = P(X_p \leq x) = P(X_p \leq [x]) = 1 - q^{[x]}$$

where $[x] := \sum_{n=1}^{\infty} n \cdot 1_{[n,n+1)}$.

**Method 1.** Our goal is to choose $a_p > 0$ and $b_p \in \mathbb{R}$ such that $\lim_{p\downarrow 0} F_p(a_p x + b_p)$ exists. As above, we first demand (taking $x = 0$) that

$$\lim_{p\downarrow 0} F_p(b_p) = \gamma_1 \in (0,1).$$

Since, $\gamma_1 \sim F_p(b_p) \sim 1 - q^{b_p}$ we require, $q^{b_p} \sim 1 - \gamma_1$ and hence, $c \sim b_p \ln q = b_p \ln(1-p) \sim -b_p p$. This suggests that we take $b_p = 1/p$ say. Having done this, we would like to choose $a_p$ such that

$$F_0(x) := \lim_{p\downarrow 0} F_p(a_p x + b_p) \text{ exists.}$$

Since,

$$F_0(x) \sim F_p(a_p x + b_p) \sim 1 - q^{a_p x + b_p}$$

this requires that

$$(1-p)^{a_p x + b_p} = q^{a_p x + b_p} \sim 1 - F_0(x)$$

and hence that

$$\ln(1 - F_0(x)) = (a_p x + b_p)\ln q \sim (a_p x + b_p)(-p) = -p a_p x - 1.$$

From this (setting $x = 1$) we see that $p a_p \sim c > 0$. Hence we might take $a_p = 1/p$ as well. We then have

$$F_p(a_p x + b_p) = F_p\left(p^{-1}x + p^{-1}\right) = 1 - (1-p)^{\left[p^{-1}(x+1)\right]}$$

which is equal to 0 if $x \leq -1$, and for $x > -1$ we find

$$(1-p)^{\left[p^{-1}(x+1)\right]} = \exp\left(\left[p^{-1}(x+1)\right]\ln(1-p)\right) \to \exp(-(x+1)).$$

Hence we have shown,

$$\lim_{p\downarrow 0} F_p(a_p x + b_p) = [1 - \exp(-(x+1))]\, 1_{x \geq -1}$$

$$\frac{X_p - 1/p}{1/p} = pX_p - 1 \implies T - 1$$

or again that $pX_p \implies T$.

**Method 2.** (Center and scale using the first moment and the variance of $X_p$.) The generating function is given by

$$f(z) := \mathbb{E}\left[z^{X_p}\right] = \sum_{n=1}^{\infty} z^n q^{n-1} p = \frac{pz}{1-qz}.$$

Observe that $f(z)$ is well defined for $|z| < \frac{1}{q}$ and that $f(1) = 1$, reflecting the fact that $P(X_p \in \mathbb{N}) = 1$, i.e. a success must occur almost surely. Moreover, we have

$$f'(z) = \mathbb{E}\left[X_p z^{X_p - 1}\right],\ f''(z) = \mathbb{E}\left[X_p(X_p - 1) z^{X_p - 2}\right], \dots$$
$$f^{(k)}(z) = \mathbb{E}\left[X_p(X_p - 1)\dots(X_p - k + 1) z^{X_p - k}\right]$$

and in particular,

$$\mathbb{E}\left[X_p(X_p - 1)\dots(X_p - k + 1)\right] = f^{(k)}(1) = \left(\frac{d}{dz}\right)^k |_{z=1} \frac{pz}{1-qz}.$$

Since

$$\frac{d}{dz}\frac{pz}{1-qz} = \frac{p(1-qz)+qpz}{(1-qz)^2} = \frac{p}{(1-qz)^2}$$

and

$$\frac{d^2}{dz^2}\frac{pz}{1-qz} = 2\frac{pq}{(1-qz)^3}$$

it follows that

$$\mu_p := \mathbb{E}X_p = \frac{p}{(1-q)^2} = \frac{1}{p} \text{ and}$$
$$\mathbb{E}[X_p(X_p-1)] = 2\frac{pq}{(1-q)^3} = \frac{2q}{p^2}.$$

Therefore,

$$\sigma_p^2 = \mathrm{Var}(X_p) = \mathbb{E}X_p^2 - (\mathbb{E}X_p)^2 = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2$$
$$= \frac{2q+p-1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

Thus, if we had used $\mu_p$ and $\sigma_p$ to center and scale $X_p$ we would have considered,

$$\frac{X_p - \frac{1}{p}}{\frac{\sqrt{1-p}}{p}} = \frac{pX_p - 1}{\sqrt{1-p}} \implies T - 1$$

instead.

**Theorem 23.52 (This is already done in Theorem 9.74).** *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $P(X_n = \pm 1) = 1/2$ and let $S_n := X_1 + \cdots + X_n$ – the position of a drunk after $n$ steps. Observe that $|S_n|$ is an odd integer if $n$ is odd and an even integer if $n$ is even. Then $\frac{S_m}{\sqrt{m}} \implies N(0,1)$ as $m \to \infty$.*

**Proof.** (Sketch of the proof.) We start by observing that $S_{2n} = 2k$ iff

$$\#\{i \le 2n : X_i = 1\} = n + k \text{ while}$$
$$\#\{i \le 2n : X_i = -1\} = 2n - (n+k) = n - k$$

and therefore,

$$P(S_{2n} = 2k) = \binom{2n}{n+k}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n+k)! \cdot (n-k)!}\left(\frac{1}{2}\right)^{2n}.$$

Recall Stirling's formula states,

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \text{ as } n \to \infty$$

and therefore,

$$\begin{aligned}
&P(S_{2n} = 2k) \\
&\sim \frac{(2n)^{2n} e^{-2n}\sqrt{4\pi n}}{(n+k)^{n+k} e^{-(n+k)}\sqrt{2\pi(n+k)} \cdot (n-k)^{n-k} e^{-(n-k)}\sqrt{2\pi(n-k)}}\left(\frac{1}{2}\right)^{2n} \\
&= \sqrt{\frac{n}{\pi(n+k)(n-k)}}\left(1+\frac{k}{n}\right)^{-(n+k)} \cdot \left(1-\frac{k}{n}\right)^{-(n-k)} \\
&= \frac{1}{\sqrt{\pi n}}\sqrt{\frac{1}{\left(1+\frac{k}{n}\right)\left(1-\frac{k}{n}\right)}}\left(1-\frac{k^2}{n^2}\right)^{-n} \cdot \left(1+\frac{k}{n}\right)^{-k} \cdot \left(1-\frac{k}{n}\right)^{k} \\
&= \frac{1}{\sqrt{\pi n}}\left(1-\frac{k^2}{n^2}\right)^{-n} \cdot \left(1+\frac{k}{n}\right)^{-k-1/2} \cdot \left(1-\frac{k}{n}\right)^{k-1/2}.
\end{aligned}$$

So if we let $x := 2k/\sqrt{2n}$, i.e. $k = x\sqrt{n/2}$ and $k/n = \frac{x}{\sqrt{2n}}$, we have

$$\begin{aligned}
&P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right) \\
&\sim \frac{1}{\sqrt{\pi n}}\left(1-\frac{x^2}{2n}\right)^{-n} \cdot \left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}-1/2} \cdot \left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}-1/2} \\
&\sim \frac{1}{\sqrt{\pi n}}e^{x^2/2} \cdot e^{\frac{x}{\sqrt{2n}}\left(-x\sqrt{n/2}-1/2\right)} \cdot e^{-\frac{x}{\sqrt{2n}}\left(x\sqrt{n/2}-1/2\right)} \\
&\sim \frac{1}{\sqrt{\pi n}}e^{-x^2/2},
\end{aligned}$$

wherein we have repeatedly used

$$(1+a_n)^{b_n} = e^{b_n \ln(1+a_n)} \sim e^{b_n a_n} \text{ when } a_n \to 0.$$

We now compute

$$\begin{aligned}
P\left(a \le \frac{S_{2n}}{\sqrt{2n}} \le b\right) &= \sum_{a \le x \le b} P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right) \\
&= \frac{1}{\sqrt{2\pi}}\sum_{a \le x \le b} e^{-x^2/2}\frac{2}{\sqrt{2n}} \qquad (23.45)
\end{aligned}$$

where the sum is over $x$ of the form, $x = \frac{2k}{\sqrt{2n}}$ with $k \in \{0, \pm 1, \ldots, \pm n\}$. Since $\frac{2}{\sqrt{2n}}$ is the increment of $x$ as $k$ increases by 1, we see the latter expression in Eq. (23.45) is the Riemann sum approximation to

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

This proves $\frac{S_{2n}}{\sqrt{2n}} \implies N(0,1)$. Since

$$\frac{S_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n}+X_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n}}{\sqrt{2n}}\frac{1}{\sqrt{1+\frac{1}{2n}}} + \frac{X_{2n+1}}{\sqrt{2n+1}},$$

it follows directly (or see Slutsky's Theorem 23.36) that $\frac{S_{2n+1}}{\sqrt{2n+1}} \implies N(0,1)$ as well. ■

**Proposition 23.53.** *Suppose that* $\{U_n\}_{n=1}^\infty$ *are i.i.d. random variables which are uniformly distributed in* $(0,1)$. *Let* $U_{(k,n)}$ *denote the position of the* $k^{th}$ *– largest number from the list,* $\{U_1, U_2, \dots, U_n\}$. *Further let* $k(n)$ *be chosen so that* $\lim_{n\to\infty} k(n) = \infty$ *while* $\lim_{n\to\infty}\frac{k(n)}{n} = 0$ *and let*

$$X_n := \frac{U_{(k(n),n)} - k(n)/n}{\frac{\sqrt{k(n)}}{n}}.$$

*Then* $d_{TV}(X_n, N(0,1)) \to 0$ *as* $n\to\infty$.

**Proof.** (Sketch only. See Resnick, Proposition 8.2.1 for more details.) Observe that, for $x\in(0,1)$, that

$$P\left(U_{(k,n)} \le x\right) = P\left(\sum_{i=1}^n X_i \ge k\right) = \sum_{l=k}^n \binom{n}{l} x^l (1-x)^{n-l}.$$

From this it follows that $\rho_n(x) := 1_{(0,1)}(x)\frac{d}{dx}P\left(U_{(k,n)}\le x\right)$ is the probability density for $U_{(k,n)}$. It now turns out that $\rho_n(x)$ is a Beta distribution,

$$\rho_n(x) = \binom{n}{k} k\cdot x^{k-1}(1-x)^{n-k}.$$

Giving a direct computation of this result is not so illuminating. So let us go another route. To do this we are going to estimate, $P\left(U_{(k,n)}\in(x,x+\Delta]\right)$, for $\Delta\in(0,1)$. Observe that if $U_{(k,n)}\in(x,x+\Delta]$, then there must be at least one $U_i\in(x,x+\Delta]$, for otherwise, $U_{(k,n)}\le x+\Delta$ would imply $U_{(k,n)}\le x$ as well and hence $U_{(k,n)}\notin(x,x+\Delta]$. Let

$$\Omega_i := \{U_i\in(x,x+\Delta] \text{ and } U_j\notin(x,x+\Delta] \text{ for } j\ne i\}.$$

Since

$$P(U_i,U_j\in(x,x+\Delta] \text{ for some } i\ne j \text{ with } i,j\le n) \le \sum_{i<j\le n} P(U_i,U_j\in(x,x+\Delta]) \le \frac{n^2-n}{2}\Delta^2,$$

we see that

$$\begin{aligned} P\left(U_{(k,n)}\in(x,x+\Delta]\right) &= \sum_{i=1}^n P\left(U_{(k,n)}\in(x,x+\Delta],\Omega_i\right) + O\left(\Delta^2\right) \\ &= nP\left(U_{(k,n)}\in(x,x+\Delta],\Omega_1\right) + O\left(\Delta^2\right). \end{aligned}$$

Now on the set, $\Omega_1$; $U_{(k,n)}\in(x,x+\Delta]$ iff there are exactly $k-1$ of $U_2,\dots,U_n$ in $[0,x]$ and $n-k$ of these in $[x+\Delta,1]$. This leads to the conclusion that

$$P\left(U_{(k,n)}\in(x,x+\Delta]\right) = n\binom{n-1}{k-1}x^{k-1}(1-(x+\Delta))^{n-k}\Delta + O\left(\Delta^2\right)$$

and therefore,

$$\rho_n(x) = \lim_{\Delta\downarrow 0}\frac{P\left(U_{(k,n)}\in(x,x+\Delta]\right)}{\Delta} = \frac{n!}{(k-1)!\cdot(n-k)!}x^{k-1}(1-x)^{n-k}.$$

By Stirling's formula,

$$\begin{aligned} &\frac{n!}{(k-1)!\cdot(n-k)!} \\ &\sim \frac{n^n e^{-n}\sqrt{2\pi n}}{(k-1)^{(k-1)}e^{-(k-1)}\sqrt{2\pi(k-1)}\,(n-k)^{(n-k)}e^{-(n-k)}\sqrt{2\pi(n-k)}} \\ &= \frac{\sqrt{n}e^{-1}}{\sqrt{2\pi}}\frac{1}{\left(\frac{k-1}{n}\right)^{(k-1)}\sqrt{\frac{k-1}{n}}\left(\frac{n-k}{n}\right)^{(n-k)}\sqrt{\frac{n-k}{n}}} \\ &= \frac{\sqrt{n}e^{-1}}{\sqrt{2\pi}}\frac{1}{\left(\frac{k-1}{n}\right)^{(k-1/2)}\left(1-\frac{k}{n}\right)^{(n-k+1/2)}}. \end{aligned}$$

Since

$$\begin{aligned} \left(\frac{k-1}{n}\right)^{(k-1/2)} &= \left(\frac{k}{n}\right)^{(k-1/2)}\cdot\left(\frac{k-1}{k}\right)^{(k-1/2)} \\ &= \left(\frac{k}{n}\right)^{(k-1/2)}\cdot\left(1-\frac{1}{k}\right)^{(k-1/2)} \\ &\sim e^{-1}\left(\frac{k}{n}\right)^{(k-1/2)} \end{aligned}$$

we arrive at

$$\frac{n!}{(k-1)!\cdot(n-k)!}\sim\frac{\sqrt{n}}{\sqrt{2\pi}}\frac{1}{\left(\frac{k}{n}\right)^{(k-1/2)}\left(1-\frac{k}{n}\right)^{(n-k+1/2)}}.$$

By the change of variables formula, with

$$x=\frac{u-k(n)/n}{\frac{\sqrt{k(n)}}{n}}$$

on noting the $du=\frac{\sqrt{k(n)}}{n}dx$, $x=-\sqrt{k(n)}$ at $u=0$, and

$$\begin{aligned}x&=\frac{1-k(n)/n}{\frac{\sqrt{k(n)}}{n}}=\frac{n-k(n)}{\sqrt{k(n)}}\\&=\frac{n}{\sqrt{k(n)}}\left(1-\frac{k(n)}{n}\right)=\sqrt{n}\sqrt{\frac{n}{k(n)}}\left(1-\frac{k(n)}{n}\right)=:b_n,\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[F(X_n)\right]&=\int_0^1\rho_n(u)\,F\left(\frac{u-k(n)/n}{\frac{\sqrt{k(n)}}{n}}\right)du\\&=\int_{-\sqrt{k(n)}}^{b_n}\frac{\sqrt{k(n)}}{n}\rho_n\left(\frac{\sqrt{k(n)}}{n}x+k(n)/n\right)F(x)\,du.\end{aligned}$$

Using this information, it is then shown in Resnick that

$$\frac{\sqrt{k(n)}}{n}\rho_n\left(\frac{\sqrt{k(n)}}{n}x+k(n)/n\right)\to\frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

which upon an application of Scheffé's Lemma 23.8 completes the proof. ■

*Remark 23.54.* It is possible to understand the normalization constants in the definition of $X_n$ by computing the mean and the variance of $U_{(n,k)}$. After some computations (see Chapter **??**), one arrives at

$$\begin{aligned}\mathbb{E}U_{(k,n)}&=\int_0^1\frac{n!}{(k-1)!\cdot(n-k)!}x^{k-1}(1-x)^{n-k}\,xdx\\&=\frac{k}{n+1}\sim\frac{k}{n},\\\mathbb{E}U_{(k,n)}^2&=\int_0^1\frac{n!}{(k-1)!\cdot(n-k)!}x^{k-1}(1-x)^{n-k}\,x^2dx\\&=\frac{(k+1)\,k}{(n+2)(n+1)}\text{ and}\\\operatorname{Var}\left(U_{(k,n)}\right)&=\frac{(k+1)\,k}{(n+2)(n+1)}-\frac{k^2}{(n+1)^2}\\&=\frac{k}{n+1}\left[\frac{k+1}{n+2}-\frac{k}{n+1}\right]\\&=\frac{k}{n+1}\left[\frac{n-k+1}{(n+2)(n+1)}\right]\sim\frac{k}{n^2}.\end{aligned}$$

## 23.8 Compactness and tightness of measures on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$

**Notation 23.55** *If $(\Omega,\mathcal{B})$ is a measurable space let $\mathcal{P}_1(\Omega,\mathcal{B})$ denote the collection of probability measure on $(\Omega,\mathcal{B})$.*

**Definition 23.56.** *A subset $\Gamma\subset\mathcal{P}_1(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ is* ***weakly sequentially precompact*** *iff for every sequence, $\{\mu_n\}_{n=1}^\infty\subset\Gamma$, there exists a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ such that $\mu_{n_k}\implies\mu_0$ for some probability measure $\mu_0\in\mathcal{P}_1(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.*

**Definition 23.57.** *A collection of probability measures, $\Gamma$, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ is* ***tight*** *iff for every $\varepsilon>0$ there exists $M_\varepsilon<\infty$ such that*

$$\inf_{\mu\in\Gamma}\mu\left([-M_\varepsilon,M_\varepsilon]\right)\geq1-\varepsilon. \tag{23.46}$$

*We further say that a collection of random variables, $\{X_\lambda:\lambda\in\Lambda\}$ is* ***tight*** *iff the collection probability measures, $\left\{P\circ X_\lambda^{-1}:\lambda\in\Lambda\right\}$ is tight. Equivalently put, $\{X_\lambda:\lambda\in\Lambda\}$ is tight iff*

$$\lim_{M\to\infty}\sup_{\lambda\in\Lambda}P\left(|X_\lambda|\geq M\right)=0. \tag{23.47}$$

Observe that the definition of uniform integrability (see Definition 14.38) is considerably stronger than the notion of tightness. It is also worth observing that if $\alpha>0$ and $C:=\sup_{\lambda\in\Lambda}\mathbb{E}\left|X_\lambda\right|^\alpha<\infty$, then by Chebyschev's inequality,

$$\sup_\lambda P\left(|X_\lambda|\geq M\right)\leq\sup_\lambda\left[\frac{1}{M^\alpha}\mathbb{E}\left|X_\lambda\right|^\alpha\right]\leq\frac{C}{M^\alpha}\to0\text{ as }M\to\infty$$

and therefore $\{X_\lambda:\lambda\in\Lambda\}$ is tight.

**Proposition 23.58.** *If $\Gamma\subset\mathcal{P}_1(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ is weakly sequentially pre-compact, then $\Gamma$ is tight.*

**Proof.** For sake of contradiction, suppose there exists $\Gamma\subset\mathcal{P}_1(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ which is weakly sequentially pre-compact but is **not** tight. Since $\Gamma$ is not tight, there exists an $\varepsilon>0$ such that $\inf_{\mu\in\Gamma}\mu([-M,M])<1-\varepsilon$ for all $M\in(0,\infty)$. Hence it is possible to choose $\{\mu_n\}_{n=1}^\infty\subset\Gamma$ such that $\mu_n([-n,n])<1-\varepsilon$ for all $n\in\mathbb{N}$. Since $\Gamma$ is pre-compact there exists a subsequence, $\{\nu_k:=\mu_{n_k}\}_{k=1}^\infty$ and $\mu_0\in\mathcal{P}_1(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\nu_k\implies\mu_0$. Hence for every $M\in\mathbb{R}_+$ such that $M\in\mathcal{C}(\mu_0)$ we have

$$\mu_0((-M,M])=\lim_{k\to\infty}\nu_k((-M,M])\le\liminf_{k\to\infty}\nu_k([-n_k,n_k])\le1-\varepsilon.$$

We may now let $M\in\mathcal{C}(\mu_0)$ tend to infinity in order to conclude $\mu_0(\mathbb{R})\le1-\varepsilon<1$ which contradicts the assumption that $\mu_0$ is a probability measure. ∎

The goal of this section is to prove the converse of Proposition 23.58, see Theorem 23.64. In order to do this we will need find ways to find convergence subsequences of probability measures. The method presented here will make heavy use of cumulative distribution functions.

Suppose that $\Lambda\subset\mathbb{R}$ is a dense set and $F$ and $\tilde F$ are two right continuous functions. If $F=\tilde F$ on $\Lambda$, then $F=\tilde F$ on $\mathbb{R}$. Indeed, for $x\in\mathbb{R}$ we have

$$F(x)=\lim_{\Lambda\ni\lambda\downarrow x}F(\lambda)=\lim_{\Lambda\ni\lambda\downarrow x}\tilde F(\lambda)=\tilde F(x).$$

**Lemma 23.59.** *If $G:\Lambda\to\mathbb{R}$ is a non-decreasing function, then*

$$F(x):=G_+(x):=\inf\{G(\lambda):x<\lambda\in\Lambda\}\tag{23.48}$$

*is a non-decreasing right continuous function.*

**Proof.** To show $F$ is right continuous, let $x\in\mathbb{R}$ and $\lambda\in\Lambda$ such that $\lambda>x$. Then for any $y\in(x,\lambda)$,

$$F(x)\le F(y)=G_+(y)\le G(\lambda)$$

and therefore,

$$F(x)\le F(x+):=\lim_{y\downarrow x}F(y)\le G(\lambda).$$

Since $\lambda>x$ with $\lambda\in\Lambda$ is arbitrary, we may conclude, $F(x)\le F(x+)\le G_+(x)=F(x)$, i.e. $F(x+)=F(x)$. ∎

**Proposition 23.60.** *Suppose that $\{F_n\}_{n=1}^\infty$ is a sequence of distribution functions and $\Lambda\subset\mathbb{R}$ is a dense set such that $G(\lambda):=\lim_{n\to\infty}F_n(\lambda)\in[0,1]$ exists for all $\lambda\in\Lambda$. If, for all $x\in\mathbb{R}$, we define $F=G_+$ as in Eq. (23.48), then $F_n(x)\to F(x)$ for all $x\in\mathcal{C}(F)$ [see Notation 23.17]. (Note well; as we have already seen, it is possible that $F(\infty)<1$ and $F(-\infty)>0$ so that $F$ need not be a distribution function for a measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.)*

**Proof.** Suppose that $x,y\in\mathbb{R}$ with $x<y$ and and $s,t\in\Lambda$ are chosen so that $x<s<y<t$. Then passing to the limit in the inequality,

$$F_n(s)\le F_n(y)\le F_n(t)$$

implies

$$F(x)=G_+(x)\le G(s)\le\liminf_{n\to\infty}F_n(y)\le\limsup_{n\to\infty}F_n(y)\le G(t).$$

Taking the infimum over $t\in\Lambda\cap(y,\infty)$ and then letting $x\in\mathbb{R}$ tend up to $y$, we may conclude

$$F(y-)\le\liminf_{n\to\infty}F_n(y)\le\limsup_{n\to\infty}F_n(y)\le F(y)\text{ for all }y\in\mathbb{R}.$$

This completes the proof, since $F(y-)=F(y)$ for $y\in\mathcal{C}(F)$. ∎

The next theorem deals with weak convergence of measures on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$. So as not have to introduce any new machinery, the reader should identify $\bar{\mathbb{R}}$ with $[-1,1]\subset\mathbb{R}$ via the map,

$$[-1,1]\ni x\to\tan\left(\frac{\pi}{2}x\right)\in\bar{\mathbb{R}}.$$

Hence a probability measure on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$ may be identified with a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ which is supported on $[-1,1]$. Using this identification, we see that a $-\infty$ should only be considered a point of continuity of a distribution function, $F:\bar{\mathbb{R}}\to[0,1]$ iff and only if $F(-\infty)=0$. On the other hand, $\infty$ is always a point of continuity.

**Theorem 23.61 (Helly's Selection Theorem).** *Every sequence of probability measures, $\{\mu_n\}_{n=1}^\infty$, on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$ has a sub-sequence which is weakly convergent to a probability measure, $\mu_0$ on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$.*

**Proof.** Using the identification described above, rather than viewing $\mu_n$ as probability measures on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$, we may view them as probability measures on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ which are supported on $[-1,1]$, i.e. $\mu_n([-1,1])=1$. As usual, let

$$F_n(x):=\mu_n((-\infty,x])=\mu_n((-\infty,x]\cap[-1,1]).$$

Since $\{F_n(x)\}_{n=1}^\infty\subset[0,1]$ and $[0,1]$ is compact, for each $x\in\mathbb{R}$ we may find a convergence subsequence of $\{F_n(x)\}_{n=1}^\infty$. Hence by Cantor's diagonalization argument we may find a subsequence, $\{G_k:=F_{n_k}\}_{k=1}^\infty$ of the $\{F_n\}_{n=1}^\infty$ such that $G(x):=\lim_{k\to\infty}G_k(x)$ exists for all $x\in\Lambda:=\mathbb{Q}$.

Letting $F(x):=G(x+)$ as in Eq. (23.48), it follows from Lemma 23.59 and Proposition 23.60 that $G_k=F_{n_k}\implies F_0$. Moreover, since $G_k(x)=0$ for all $x\in\mathbb{Q}\cap(-\infty,-1)$ and $G_k(x)=1$ for all $x\in\mathbb{Q}\cap[1,\infty)$. Therefore, $F_0(x)=1$ for all $x\ge1$ and $F_0(x)=0$ for all $x<-1$ and the corresponding measure, $\mu_0$ is supported on $[-1,1]$. Hence $\mu_0$ may now be transferred back to a measure on $(\bar{\mathbb{R}},\mathcal{B}_{\bar{\mathbb{R}}})$. ∎

*Example 23.62.* Here are there simple examples showing that probabilities may indeed transfer to the points at $\pm\infty$; 1) $\delta_{-n} \implies \delta_{-\infty}$, 2) $\delta_n \implies \delta_\infty$ and 3) $\frac{1}{2}(\delta_n + \delta_{-n}) \implies \frac{1}{2}(\delta_\infty + \delta_{-\infty})$.

**Theorem 23.63 (Helly's Selection Theorem for $\mathbb{R}^n$).** *Every sequence of probability measures, $\{\mu_n\}_{n=1}^\infty$, on $\left(\bar{\mathbb{R}}^n, \mathcal{B}_{\bar{\mathbb{R}}^n}\right)$ has a sub-sequence which is weakly convergent to a probability measure, $\mu_0$ on $\left(\bar{\mathbb{R}}^n, \mathcal{B}_{\bar{\mathbb{R}}^n}\right)$.*

**Proof.** The proof is very similar to Theorem 23.61 provided we replace Lemma 23.59 by Lemma 6.82 and use a multi-variate version of Proposition 23.60. ∎

The next question we would like to address is when is the limiting measure, $\mu_0$ on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ concentrated on $\mathbb{R}$.

**Theorem 23.64.** *A subset, $\Gamma \subset \mathcal{P}_1(\mathbb{R}, \mathcal{B}_\mathbb{R}) \subset \mathcal{P}_1\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$, is tight, iff every subsequently limit measure, $\mu_0 \in \mathcal{P}_1\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$, is supported on $\mathbb{R}$. In particular if $\Gamma$ is tight, there is a weakly convergent subsequence of $\Gamma$ converging to a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. (This is greatly generalized in Prokhorov's Theorem 23.89 below.) [This theorem generalizes to $\mathbb{R}^n$ in a fairly obvious way.]*

**Proof.** Suppose that $\Gamma \ni \mu_{n_k} \implies \mu_0$ with $\mu_0$ being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ – note the bars here! As usual, let $F_0(x) := \mu_0([-\infty, x])$. If $\Gamma$ is tight and $\varepsilon > 0$ is given, we may find $M_\varepsilon < \infty$ such that $M_\varepsilon, -M_\varepsilon \in \mathcal{C}(F_0)$ and $\mu([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$ for all $\mu \in \Gamma$. Hence it follows that

$$\mu_0((-M_\varepsilon, M_\varepsilon]) = \lim_{k\to\infty} \mu_{n_k}((-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$$

and by letting $\varepsilon \downarrow 0$ we conclude that $\mu_0(\mathbb{R}) = \lim_{\varepsilon\downarrow 0} \mu_0([-M_\varepsilon, M_\varepsilon]) = 1$.

Conversely, suppose there is a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ such that $\mu_{n_k} \implies \mu_0$ with $\mu_0$ being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ such that $\mu_0(\mathbb{R}) < 1$. In this case $\varepsilon_0 := \mu_0(\{-\infty, \infty\}) > 0$ and hence for all $M < \infty$ we have

$$\mu_0([-M, M]) \leq \mu_0\left(\bar{\mathbb{R}}\right) - \mu_0(\{-\infty, \infty\}) = 1 - \varepsilon_0.$$

By choosing $M$ so that $-M$ and $M$ are points of continuity of $F_0$, it then follows that

$$\lim_{k\to\infty} \mu_{n_k}([-M, M]) = \mu_0([-M, M]) \leq 1 - \varepsilon_0.$$

Therefore,

$$\inf_{n\in\mathbb{N}} \mu_n(([-M, M])) \leq 1 - \varepsilon_0 \text{ for all } M < \infty$$

and $\{\mu_n\}_{n=1}^\infty$ is not tight. ∎

## 23.9 Extensions to $\mathbb{R}^n$

Some of this section came out of the analysis notes and should be combined with the previous section and sections later on weak convergence of probability measuresSupplement: Generalizations of Theorem **??** or **??** to $\mathbb{R}^n$. There is a lot of redundancy in this section with Section 6.12. The main goal should be Theorem 23.84.

**Definition 23.65.** *Let $\mu$ be a probability measure on $(\mathbb{R}^n, \mathcal{B} = \mathcal{B}_{\mathbb{R}^n})$ and $F : \mathbb{R}^n \to [0,1]$ be its* ***cumulative distribution function*** *(CDF for short) defined by $F(x) := \mu((-\infty, x])$ where*

$$S_x := (-\infty, x] := (-\infty, x_1] \times \cdots \times (-\infty, x_n].$$

Since

$$(a, b] = S_b \setminus \left[S_{(a_1, b_2, \ldots, b_n)} \cup S_{(b_1, a_2, \ldots, b_n)} \cup \cdots \cup S_{(b_1, b_2, \ldots, b_{n-1}, a_n)}\right],$$

it follows that $\{S_b\}_{b\in\mathbb{R}^n}$ is a $\pi$ – system generating the $\mathcal{B}$ and therefore $\mu$ is uniquely determined by $F$. One of the main goals of this chapter is to characterize the collection of CDF's which appears in Theorem 23.77 at the end of this section. Let us begin with the following basic analytic result.

**Theorem 23.66.** *Suppose $\alpha, \beta \in \mathbb{R}^n$ with $\alpha < \beta$ and $\mu : \mathcal{C} := \mathcal{A}_{(\alpha,\beta]} \to [0, \infty)$ is a finitely additive measure. Then $\mu$ extends to a measure on $\mathcal{B}_{(\alpha,\beta]} = \sigma(\mathcal{C})$ iff the function $\mu((a, b])$ is right continuous in each of the variables $a$ and $b$ for $\alpha < a < b < \beta$. [Kallenberg gives another interesting proof of this theorem in Theorem 3.25 on p. 59 which goes by constructing a random vector on $[0,1]$ with the correct distribution.]*

**Proof.** Following example **??** one easily shows

$$\mathcal{E} := \{(a, b] : \alpha \leq a < b \leq \beta\}$$

is an elementary family which, by definition, generates the algebra $\mathcal{C}$ and the $\sigma$ – algebra $\mathcal{B}_{(\alpha,\beta]}$. So according to Proposition **??** and Theorem **??** to finish the proof we must show; $\alpha \leq a < b \leq \beta$ and $(a, b] = \sum_{n=1}^\infty (a_n, b_n]$, then

$$\mu((a, b]) \leq \sum_{n=1}^\infty \mu((a_n, b_n]). \tag{23.49}$$

Let $a < \tilde{a} < b$ and $\varepsilon > 0$ be given. Use the right continuity assumption of $\mu$ to find $\tilde{b}_n > b_n$ such that

$$\mu\left((a_n, \tilde{b}_n \wedge b]\right) \leq \mu((a_n, b_n]) + \varepsilon 2^{-n} \text{ for } n \in \mathbb{N}.$$

We have

$$[\tilde{a}, b] \subset (a, b] \subset \cup_n \left(a_n, \tilde{b}_n\right).$$

So by compactness there exists $N < \infty$ such that

$$(\tilde{a}, b] \subset [\tilde{a}, b] \subset \cup_{n=1}^N \left(a_n, \tilde{b}_n\right) \cap (a, b] \subset \cup_{n=1}^N (a_n, \tilde{b}_n \wedge b]$$

By finite sub-additivity it now follows that

$$\begin{aligned}\mu((\tilde{a}, b]) &\leq \sum_{n=1}^N \mu\left((a_n, \tilde{b}_n \wedge b]\right) \\ &\leq \sum_{n=1}^\infty \mu\left((a_n, \tilde{b}_n \wedge b]\right) \\ &\leq \sum_{n=1}^\infty \left[\mu((a_n, b_n]) + \varepsilon 2^{-n}\right] \\ &\leq \sum_{n=1}^\infty \mu((a_n, b_n]) + \varepsilon.\end{aligned}$$

As $\varepsilon > 0$ was arbitrary we may conclude,

$$\mu((\tilde{a}, b]) \leq \sum_{n=1}^\infty \mu((a_n, b_n]).$$

Then using the right continuity in the first variable we may now let $\tilde{a} \downarrow a$ to conclude that Eq. (23.49) holds. ■

**Lemma 23.67.** *Let $D \subset \mathbb{R}$ be a dense set and suppose that $G : D^n \to \mathbb{R}$ is an increasing function in the sense that $G(x) \leq G(y)$ whenever $x, y \in D$ with $x \leq y$. Define $F : \mathbb{R}^n \to \mathbb{R}$ by*

$$F(x) := \inf_{y > x} G(y). \tag{23.50}$$

*Then,*

1. *$F$ is an increasing function,*
2. *$F(a) \leq G(y) \leq F(b)$ whenever $a < y \leq b$ with $y \in D^n$,*
3. *$F$ is right continuous function, and*
4. *for all $a, b \in \mathbb{R}^n$ with $a < b$ we have*

$$\inf_{x \in (a,b)} F(x) = F(a) = \inf_{y \in (a,b) \cap D^n} G(y) \tag{23.51}$$

$$\sup_{x \in (a,b)} F(x) = \sup_{y \in (a,b) \cap D^n} G(y) \text{ and} \tag{23.52}$$

**Proof.** The first two items easily follow from the increasing properties of $G$ and the definition of $F$. To see that $F$ is right continuous as $x \in \mathbb{R}^n$, let $\varepsilon > 0$ be given and choose $y \in D^n$ with $y > x$ such that $F(x) \leq G(y) \leq F(x) + \varepsilon$. Then for $x \leq x' < y$ we have that

$$F(x) \leq F(x') \leq G(y) \leq F(x) + \varepsilon$$

which proves the right continuity of $F$. The first equality in Eq. (23.51) follows from the right continuity of $F$ while the second follows from the definition of $F$ and the fact that $G$ is increasing. Since for every $x \in (a, b)$ and $y \in (a, b) \cap D^n$ with $y > x$ we have $F(x) \leq G(y)$ it follows that

$$F(x) \leq \sup_{y \in (a,b) \cap D^n} G(y) \implies \sup_{x \in (a,b)} F(x) \leq \sup_{y \in (a,b) \cap D^n} G(y).$$

Similarly, for any $y \in (a, b) \cap D^n$ and $x \in (y, b)$ we have

$$G(y) \leq F(x) \leq \sup_{x \in (a,b)} F(x) \implies \sup_{y \in (a,b) \cap D^n} G(y) \leq \sup_{x \in (a,b)} F(x).$$

■

**Lemma 23.68.** *Let $D \subset \mathbb{R}$ be a dense set and suppose that $F_k : \mathbb{R}^n \to \mathbb{R}$ is a sequence of right continuous increasing functions such that $G(x) := \lim_{k\to\infty} F_k(x)$ exists for all $x \in D^n$. Then*

$$\lim_{k\to\infty} F_k(x) = F(x) \text{ for all } x \in \mathcal{C}(F)$$

*where $F$ is given as in Eq. (23.50) and $\mathcal{C}(F)$ are the points of continuity of $F$.*

**Proof.** Let $x \in \mathcal{C}(F)$ and $\varepsilon > 0$ be given. Choose $a < b$ such that $x \in (a, b)$ and $F$ varies by at most $\varepsilon$ on $(a, b)$. By item 4. of Lemma 23.67 it that $G$ and $F$ differ by at most $2\varepsilon$ on $(a, b) \cap D^n$. Thus for any $\alpha, \beta \in (a, b) \cap D^n$ with $\alpha < x < \beta$ we have

$$F_k(\alpha) \leq F_k(x) \leq F_k(\beta)$$

from which it follows that

$$G(\alpha) = \lim_{k\to\infty} F_k(\alpha) \leq \liminf_{k\to\infty} F_k(x) \leq \limsup_{k\to\infty} F_k(x) \leq \lim_{k\to\infty} F_k(\beta) = G(\beta).$$

Hence we conclude that $\liminf_{k\to\infty} F_k(x)$ and $\limsup_{k\to\infty} F_k(x)$ are both within $2\varepsilon$ of $F(x)$ and as $\varepsilon > 0$ was arbitrary the proof is complete. ■

**Corollary 23.69.** *Let $F_k : \mathbb{R}^n \to [0, 1]$ be a sequence of right continuous increasing functions. Then there a right continuous increasing function $F : \mathbb{R}^n \to [0, 1]$ and a subsequence $\Lambda \subset \mathbb{N}$ such that $\lim_{\Lambda \ni k\to\infty} F_k(x) = F(x)$ exists for all $x \in \mathcal{C}(F)$.*

**Proof.** Let $D \subset \mathbb{R}$ be a countalbe dense set and then use Cantor's diagonalization argument to find a $\Lambda \subset \mathbb{N}$ such that

$$G(y) := \lim_{\Lambda \ni k \to \infty} F_k(y) \text{ exists for } y \in D^n.$$

Now apply Lemma 23.68 to finish the proof. ∎

### 23.9.1 Finitely additive measures for $\mathbb{R}^n$

Suppose $V$ is a vector space and $\mu : \mathcal{A} \to V$ is a finitely additive measure. Let $F(x) := \mu(S_x)$ for all $x \in \bar{\mathbb{R}}^n$. If $a < b$ in $\mathbb{R}^n$, then

$$\begin{aligned} 1_{(a,b]} &= \prod_{i=1}^n 1_{(a_i,b_i]} = \prod_{i=1}^n \left[1_{(-\infty,b_i]} - 1_{(-\infty,a_i]}\right] \\ &= \sum_{y \in \prod_{i=1}^n \{a_i,b_i\}} \operatorname{sgn}(y) 1_{(-\infty,y]} = \sum_{k \in \{0,1\}^n} (-1)^{|k|} 1_{(-\infty,y^k(a,b)]}, \end{aligned}$$

where

$$\begin{aligned} \operatorname{sgn}(y) &:= (-1)^{\#\{i:y_i=a_i\}}, \\ |k| &:= k_1 + \dots + k_n \end{aligned}$$

and $\left\{y^k(a,b)\right\}_{k\in\{0,1\}^n}$ are the $2^n$ – vertices of $[a,b]$ indexed as;

$$\left[y^k(a,b)\right]_i := \begin{cases} a_i \text{ if } k_i = 1 \\ b_i \text{ if } k_i = 0. \end{cases}$$

**Proposition 23.70.** *Suppose $F$ and $\mu$ are as above, then for every $a, b \in \bar{\mathbb{R}}^n$ with $a < b$,*

$$\mu((a,b]) = \sum_{y \in \prod_{i=1}^n \{a_i,b_i\}} \operatorname{sgn}(y) F(y) \tag{23.53}$$

$$= \sum_{k \in \{0,1\}^n} (-1)^{|k|} F\left(y^k(a,b)\right). \tag{23.54}$$

This is really a statement of the inclusion exclusion formula. For example, when $n = 2$ we have

$$(a,b] = S_b \setminus \left[S_{(a_1,b_2)} \cup S_{(b_1,a_2)}\right]$$

and by the inclusion exclusion formula (using $S_{(a_1,b_2)} \cap S_{(b_1,a_2)} = S_a$) we have

$$\begin{aligned} \mu((a,b]) &= \mu(S_b) - \mu\left(S_{(a_1,b_2)}\right) - \mu\left(S_{(b_1,a_2)}\right) + \mu(S_a) \\ &= F(b) - F(a_1,b_1) - F(b_1,a_2) + F(a). \end{aligned}$$

It will be convenient to rewrite Eq. (23.53) in another more useful way.

**Notation 23.71** *For $i \in \{1,2,\dots,n\}$ and $\alpha \in \mathbb{R}$ let*

$$(e_i^\alpha G)(x) := G(x)|_{x_i=\alpha} = G(x_1,\dots,x_{i-1},\alpha,x_{i+1},\dots,x_n)$$

*and for $\alpha < \beta$ in $\mathbb{R}$ let $\Delta_i^{(\alpha,\beta)} = e_i^\beta - e_i^\alpha$ so that*

$$\begin{aligned} \left(\Delta_i^{(\alpha,\beta)} G\right)(x) &= G(x)|_{x_i=\alpha}^{x_i=\beta} \\ &= G(x_1,\dots,x_{i-1},\beta,x_{i+1},\dots,x_n) \\ &\quad - G(x_1,\dots,x_{i-1},\alpha,x_{i+1},\dots,x_n). \end{aligned}$$

Since $e_i^\beta e_j^\alpha = e_j^\alpha e_i^\beta$ for all $i \neq j$, it follows that the $\Delta_i$ and $\Delta_j$ also commute. With this notation we have,

$$\mu((a,b]) = \mu\left((a_1,b_1] \times (\tilde{a},\tilde{b}]\right) = \Delta_1^{(a_1,b_1)} \mu\left((-\infty,x_1) \times (\tilde{a},\tilde{b}]\right).$$

Then working inductively we find

$$\begin{aligned} \mu((a,b]) &= \Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} \mu(S_x) \\ &= \Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} F \end{aligned} \tag{23.55}$$

This formula agrees with Eq. (23.53) since,

$$\begin{aligned} \Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} F &= \prod_{i=1}^n \left(e_i^{b_i} - e_i^{a_i}\right) F \\ &= \sum_{y \in \prod_{i=1}^n \{a_i,b_i\}} \operatorname{sgn}(y) \prod_{i=1}^n e_i^{y_i} F \\ &= \sum_{y \in \prod_{i=1}^n \{a_i,b_i\}} \operatorname{sgn}(y) F(y). \end{aligned}$$

If $F$ happens to be sufficiently differentiable the fundamental theorem of calculus along with Eq. (23.55) implies

$$\begin{aligned} \mu((a,b]) &= \left(\prod_{i=1}^n \int_{a_i}^{b_i} dx_i \partial_i\right) F \\ &= \int_{(a,b]} (\partial_1 \dots \partial_n F)(x)\, dx_1 \dots dx_n. \end{aligned}$$

From this we see that if $G$ is another function such that $\partial_1 \dots \partial_n F = \partial_1 \dots \partial_n G$ then

$$\Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} F = \Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} G.$$

As $\operatorname{Nul}(\partial_1 \dots \partial_n)$ is the linear combination of functions which depend on $x$ in all but one of its coordinates, the previous remarks motivate the following lemma.

**Lemma 23.72.** *If $F:\mathbb{R}^n\to V$ is a function such that $\Delta_1^{(a_1,b_1)}\dots\Delta_n^{(a_n,b_n)}F=0$ for all $a<b$ in $\mathbb{R}^n$ then $F=\sum_{i=1}^n\psi_i$ where $\psi_i:\mathbb{R}^n\to V$ is a function which does not depend on $x_i$, i.e. $\Delta_i\psi_i=0$ for each $i$.*

**Proof.** When $n=1$ the result is obvious. When $n=2$, we have $\Delta_1\Delta_2^{(a_2,b_2)}F=0$ implies $F(x,b_2)-F(x,a_2)=C(a_2,b_2)$. Fixing $a_2$ and then letting $b_2=y>a_2$ implies,

$$F(x,y)=F(x,a_2)+C(a_2,y)\ \text{ for all } y>a_2.$$

Thus we have shown for any $\alpha>-\infty$ that there exists functions $f_\alpha$ and $g_\alpha$ on $\mathbb{R}$ such that

$$F(x,y)=f_\alpha(x)+g_\alpha(y)\ \text{ for } y>\alpha.$$

Now suppose that $\alpha<\beta$ so that

$$f_\alpha(x)+g_\alpha(y)=F(x,y)=f_\beta(x)+g_\beta(y)\ \text{ for } y>\beta,$$

i.e.

$$f_\alpha(x)-f_\beta(x)=g_\beta(y)-g_\alpha(y)\ \text{ for } y>\beta.$$

From this we see that we $f_\alpha(x)-f_\beta(x)=g_\beta(y)-g_\alpha(y)=C$ for $y>\beta$, i.e.

$$f_\alpha(x)=f_\beta(x)+C \text{ and } g_\alpha(y)=g_\beta(y)-C \text{ for } y>\beta.$$

Thus given a fixed $\beta$ we may always choose $f_\alpha$ and $g_\alpha$ so that $f_\alpha=f_\beta$ and $g_\alpha=g_\beta$ on $y>\beta$. Thus we may consistently find $f$ and $g$ such that $F(x,y)=f(x)+g(y)$. The general case now follows by an induction argument which we omit. ∎

The upshot of Lemma 23.72 is that when $n\geq 2$, there is considerable ambiguity in the choice of $F$ that we may use so that Eq. (23.55) holds. Since the functions $\{\psi_i\}$ are completely arbitrary they can be chosen to be highly discontinuous and with no monotonicity properties whatsoever. Thus if we are trying to characterize distribution functions, $F$, Eq. (23.55) we have to be careful to have chosen a good representative of $F$ where we write $F\sim G$ iff $\Delta_1^{(a_1,b_1)}\dots\Delta_n^{(a_n,b_n)}(F-G)=0$ for all $a<b$.

Our next goal is to prove Theorem 23.74 which shows that Eq. (23.53) or equivalently Eq. (**??**) may be used to construct finitely additive measures on $\mathcal{A}$. We begin with a couple of general results which will be used in the proof.

**Proposition 23.73.** *Suppose that $\mathcal{A}\subset 2^X$ is an algebra and for each $t\in\bar{\mathbb{R}}$ let $\mathcal{B}\subset 2^{\mathbb{R}}$ denote the algebra generated by $\mathcal{E}:=\{(a,b]\cap\mathbb{R}:-\infty\leq a\leq b\leq\infty\}$. Then there is a unique additive measure $\mu$ on $\mathcal{C}$, the algebra generated by $\mathcal{A}\dot{\times}\mathcal{B}$ such that*

$$\mu(A\times(a,b])=\mu_b(A)-\mu_a(A)\ \forall\ (a,b]\in\mathcal{E} \text{ and } A\in\mathcal{A}.$$

**Proof.** By Proposition **??**, for each $A\in\mathcal{A}$, the function $(a,b]\to\mu(A\times(a,b])$ extends to a unique measure on $\mathcal{B}$ which we continue to denote by $\mu$. Now if $B\in\mathcal{B}$, then $B=\sum_k I_k$ with $I_k\in\mathcal{E}$, then

$$\mu(A\times B)=\sum_k\mu(A\times I_k)$$

from which we learn that $A\to\mu(A\times B)$ is still finitely additive. The proof is complete with an application of Theorem **??**. ∎

**Theorem 23.74.** *Suppose that $F:\bar{\mathbb{R}}^n\to V$ is a function. Then there exists a unique finitely additive measure $\mu:\mathcal{A}\to V$ such that*

$$\mu((a,b])=\Delta_1^{(a_1,b_1)}\dots\Delta_n^{(a_n,b_n)}F \tag{23.56}$$

*for all $a\leq b$ in $\bar{\mathbb{R}}^n$.*

**Proof.** We will prove the result by induction. The case $n=1$ has already been carried out in Proposition **??** above. For the induction step notice that $\mathcal{A}(\mathbb{R}^n)=\mathcal{A}\left(\mathbb{R}^{n-1}\right)\otimes\mathcal{A}(\mathbb{R})$. For $t\in\mathbb{R}$ and $A\in\mathcal{A}\left(\mathbb{R}^{n-1}\right)$, let

$$\mu_t(A)=\mu_{F(\cdot,t)}(A)$$

where $\mu_{F(\cdot,t)}$ is defined by the induction hypothesis. Then

$$\mu_F(A\times(a,b])=\mu_b(A)-\mu_a(A)$$

has, by Proposition 23.73, a unique extension to $\mathcal{A}\left(\mathbb{R}^{n-1}\right)\otimes\mathcal{A}(\mathbb{R})$ as a finitely additive measure. This measure has the desired properties. ∎

**Theorem 23.75.** *Suppose that $\alpha<\beta$ in $\mathbb{R}^n$ and $F:[\alpha,\beta]\to\mathbb{R}$ is right continuous and satisfies*

$$\Delta_1^{(a_1,b_1)}\dots\Delta_n^{(a_n,b_n)}F\geq 0$$

*for all $\alpha\leq a<b\leq\beta$. Then there exists a unique measure $\mu=\mu_F$ on $\mathcal{B}_{(\alpha,\beta]}$ such that*

$$\mu((a,b])=\Delta_1^{(a_1,b_1)}\dots\Delta_n^{(a_n,b_n)}F \tag{23.57}$$

*for all $a\leq b$ with $a,b\in(\alpha,\beta]$.*

**Proof.** As in Theorem 23.74, there exists a finitely additive $(\mu)$ measure on $\mathcal{A}_{(\alpha,\beta]}$ such that Eq. (23.57) holds. So according to Theorem 23.66 it suffices to show that $\mu$ satisfying Eq. (23.57) is right continuous in both $a$ and $b$. However, this follows easily from the assumed right continuity of $F$. Indeed using Eq. (23.54) we have,

$$\mu((a,b]) = \sum_{\varepsilon\in\{0,1\}^n} (-1)^{|\varepsilon|} F(y^\varepsilon(a,b))$$

where $y^\varepsilon(a,b)$ satisfies, $y^\varepsilon(a,b') \downarrow y^\varepsilon(a,b)$ as $b' \downarrow b$ and $y^\varepsilon(a',b) \downarrow y^\varepsilon(a,b)$ as $a' \downarrow a$ and therefore it follows that $\mu((a,b])$ is right continuous in both $a$ and $b$ provided $F$ is right continuous. ∎

**Corollary 23.76.** *Suppose that $F : \mathbb{R}^n \to \mathbb{R}$ is a right continuous and satisfies*

$$\Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} F \geq 0$$

*for all $a < b$ in $\mathbb{R}^n$. Then there exists a unique measure $\mu = \mu_F$ on $\mathcal{B}$ such that Eq. (23.57) holds for all $a < b$ in $\mathbb{R}^n$. Moreover this measure is finite on compact sets.*

**Proof.** Let $\beta_k := (k, k, \dots, k)$ for $k \in \mathbb{N}$. Then by Theorem 23.75 there exists measure $\mu_k$ on $\mathcal{B}_{(-\beta_k,\beta_k]}$ such that Eq. (23.57) holds for all $a < b$ with $a, b \in (-\beta_k, \beta_k]$. We may view each of these as measure on $\mathcal{B}$ by setting $\mu_k(A) := \mu_k(A \cap (-\beta_k, \beta_k])$ for all $A \in \mathcal{B}$. With this convention, we have a sequence of increasing measure $\{\mu_k\}_{k=1}^\infty$. The desired measure may now be constructed as $\mu := \lim_{k\to\infty} \mu_k$. ∎

**Theorem 23.77 (Characterizing CDF's).** *If $F : \mathbb{R}^n \to [0,1]$ is the CDF of a probability measure $(\mu)$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ [i.e. $F(x) = \mu(S_x)$] then,*

1. *$F$ is right continuous,*
2. *$F$ is increasing in the sense that $F(a) \leq F(b)$ whenever $a \leq b$,*
3. *$\lim_{x_i\to-\infty} F(a_1, \dots a_{i-1}, x_i, a_{i+1} \dots, a_n) = 0$ for all $i$ and $a \in \mathbb{R}^n$,*
4. *$\lim_{x\to\infty} F(x) = 1$ where we say $x \to \infty$ if $x_i \to \infty$ for all $i$, and*
5. *$\Delta_1^{(a_1,b_1)} \dots \Delta_n^{(a_n,b_n)} F \geq 0$ for all $a < b$.*

*Conversely if $F : \mathbb{R}^n \to [0,1]$ is a function with the above properties then $F$ is the CDF of a unique probability measure $(\mu)$ on $\mathcal{B}$.*

**Proof.** It is an elementary exercise to show that every CDF must posses properties 1. – 5. above. For the converse assertion, we let $\mu$ be the measure given in Corollary 23.76. To see that $\mu$ is the desired measure we compute,

$$\begin{aligned}\mu(S_b) &= \lim_{a\to-\infty} \mu((a,b]) \\ &= \lim_{a\to-\infty} \sum_{y\in\prod_{i=1}^n \{a_i,b_i\}} \mathrm{sgn}(y) F(y) = F(b)\end{aligned}$$

wherein we have used that each $y \in \prod_{i=1}^n \{a_i, b_i\}$ other than $y = b$ has at least one coordinate tending to $-\infty$ and therefore $F(y)$ is converging to 0 for these $y$. Also notice that

$$\mu(\mathbb{R}^n) = \lim_{b\to\infty} \mu(S_b) = \lim_{b\to\infty} F(b) = 1$$

so that $\mu$ is indeed a probability measure. ∎

*Remark 23.78.* One might think that item 5. of Theorem 23.77 is a consequence of item 2. but his is not the case. For example, take $F(x,y) := x + y - xy$ for $x, y \in [0,1]$. Then $F_x(x,y) = 1 - y \geq 0$ and $F_y(x,y) = 1 - x \geq 0$ while $F_{xy}(x,y) = -1$ so that $\Delta_1^{(a_1,b_1)} \Delta_2^{(a_2,b_2)} F = -(b_1 - a_1)(b_2 - a_2)$. This remark is not so surprising in light of Lemma 23.72.

**Lemma 23.79.** *Suppose that $F : \mathbb{R}^n \to \mathbb{R}$ is continuous from above and is increasing in each of its variables. Then the following are equivalent;*

1. *$x \in \mathcal{C}(F)$ – the continuity points of $F$,*
2. *if $\{y_n\}$ is any sequence such that $y_n < x$ and $\lim_{n\to\infty} y_n = x$, then $\lim_{n\to\infty} F(y_n) = F(x)$,*
3. *$\lim_{n\to\infty} F\left(x - \frac{1}{n}\mathbf{1}\right) = F(x)$ where $\mathbf{1} := (1, \dots, 1)$, and*
4. *there exists $y_n < x$ such that $\lim_{n\to\infty} F(y_n) = F(x)$.*

**Proof.** The implications, 1. $\implies$ 2. $\implies$ 3. $\implies$ 4. are obvious so it only remains to show 4. $\implies$ 1. Given $\varepsilon > 0$ there exists $n$ such that $F(y_n) \leq F(x) \leq F(y_n) + \varepsilon$. Moreover, since $F$ is right continuous there exists $z > x$ such that $F(x) \leq F(z) \leq F(x) + \varepsilon$. So for $a \in (y_n, z)$ we have $F(x) - \varepsilon \leq F(y_n) < F(a) < F(x) \leq F(x) + \varepsilon$ which shows that $F$ is continuous at $x$. ∎

**Exercise 23.9.** If $F(x) = \mu(S_x)$ is a CDF then $\mathcal{C}(F) = \{x \in \mathbb{R}^n : \mu(\mathrm{bd}(S_x)) = 0\}$. Moreover, there exists a countable subset $D \subset \mathbb{R}$ such that $\mu(\pi_i = t) = 0$ for all $t \notin D$ and $1 \leq i \leq n$ where $\pi_i(x) = x_i$. As $\mathrm{bd}(S_x) \subset \cup_i \{\pi_i = x_i\}$ it follows that every $x \in \mathbb{R}^n$ with $x_i \notin D$ for all $i$ is in $\mathcal{C}(F)$. In particular $\mathcal{C}(F)$ contains a countable dense subset of $\mathbb{R}^n$.

**Definition 23.80.** *Let $\{F_n\}_{n=1}^\infty$ and $F$ be CDF's. We say $F_n \implies F$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathcal{C}(F)$ – the continuity points of $F$.*

**Lemma 23.81 (See Lemma 23.67).** *Let $\{F_n\}_{n=1}^\infty$ and $F$ be CDF's such that there exists a dense set $D \subset \mathbb{R}^n$ with $\lim_{n\to\infty} F_n(y) = F(y)$ for all $y \in D$. Then $F_n \implies F$.*

**Proof.** Let $x \in \mathcal{C}(F)$ and $\varepsilon > 0$ be given. Choose $\alpha \in \mathbb{R}^n$ and $b \in D$ such that $\alpha < x < b$ and $F(\alpha) \leq F(b) \leq F(\alpha) + \varepsilon$. Then for any $a \in D$ with $\alpha < a < x < b$ we have, $F_n(a) \leq F_n(x) \leq F_n(b)$ and therefore,

$$F(a) = \liminf_{n\to\infty} F_n(a) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(b) = F(b) \leq F(x)+$$

Letting $a \uparrow x$ then implies,

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x) + \varepsilon$$

and as $\varepsilon > 0$ was arbitrary, it follows that From this it follows that $F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) = F(x)$. ■

**Lemma 23.82 (Right Continuous Versions. See Lemma 23.67).** *Suppose $G : \mathbb{R}^n \to \mathbb{R}$ is increasing in each of its variables. For $x \in \mathbb{R}^n$ let $F(x) := \inf_{y>x} G(y)$. Then $F$ is increasing in each of its variables and $F$ is right continuous.*

**Proof.** If $a \leq b$ and $y > b$ then $y > a$ so that $F(a) \leq G(y)$. Therefore $F(a) \leq \inf_{y>b} G(y) = F(b)$ and so $F$ is increasing in each of its variables. Now suppose that $\varepsilon > 0$ there exists $\beta > a$ such that $0 \leq G(\beta) - F(a) < \varepsilon$. Then for any $a < b < \beta$, we have

$$0 \leq G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon \text{ for all } b < y < \beta.$$

From this it follows that

$$0 \leq F(b) - F(a) \leq \inf_{b<y<\beta} G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon$$

which proves the right continuity of $F$. ■

**Theorem 23.83.** *Suppose that $\{\mu_n\} \cup \{\mu\}$ is a collection of probability measures on $\mathbb{R}^n$ with corresponding CDF's, $F_n$ and $F$ respectively. The notion of weak convergence given in Definition 23.80 agrees with that in Definition* **??** *of Section* **??** *below.*

**Proof.** Suppose that $F_n \implies F$ and let $D$ be a countable dense subset of $\mathbb{R}$ such that $\mu(\pi_i = t) = 0$ for all $t \in D$. Then if $a < b$ with $a, b \in D^n$, then $y^\varepsilon(a,b) \in D^n \subset \mathcal{C}(F)$ for all $\varepsilon \in \{0,1\}^n$. Therefore $\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b])$. Noting that $\mathcal{E} := \{(a,b] : a < b \text{ with } a, b \in D^n\}$ is closed under intersections we may use the inclusion exclusion formula to conclude that $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all finite unions of sets from $\mathcal{E}$. Thus if $V$ is an open subset of $\mathbb{R}^n$ we can find such $A_k$ being a finite unions of sets from $\mathcal{E}$ such that $A_k \uparrow V$ as $k \uparrow \infty$. It then follows that

$$\mu(A_k) = \lim_{n\to\infty} \mu_n(A_k) \leq \liminf_{n\to\infty} \mu_n(V)$$

and then letting $k \uparrow \infty$ we concluded that $\mu(V) \leq \liminf_{n\to\infty} \mu_n(V)$. So according to Proposition **??**, $\mu_n \implies \mu$.

Conversely if $\mu_n \implies \mu$,then, Proposition **??**, $\mu_n(S_x) \to \mu(S_x)$ for all $x \in D^n$ as $\mu(\mathrm{bd}(S_x)) = 0$ and hence $\lim_{n\to\infty} F_n(x) = F(x)$ for $x$ in a dense set. It now follows by an application of Lemma 23.81 that $F_n \implies F$. ■

**Theorem 23.84 (Helly's Selection Theorem for $\mathbb{R}^n$).** *Every sequence of tight probability measures, $\{\mu_n\}_{n=1}^\infty$, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ has a sub-sequence which is weakly convergent to a probability measure, $\mu_0$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.*

**Proof.** The proof is very similar to Theorem 23.61 provided we replace Lemma 23.59 by Lemma 23.82 and use a multi-variate version of Proposition 23.60. In a bit more detail. Choose $\{m_k\}$ such that $G(y) := \lim_{k\to\infty} F_{m_k}(y)$ exists for all $y \in \mathbb{Q}^n$. To simplify notation we suppose that we have already passed to this subsequence so that we may now assume $G(y) := \lim_{k\to\infty} F_k(y)$ exists for all $y \in \mathbb{Q}^n$. Notice that

$$\Delta_{(a,b]} G := \lim_{k\to\infty} \Delta_{(a,b]} F_k = \lim_{k\to\infty} \mu_k((a,b]) \geq 0 \text{ for all } a < b \text{ with } a, b \in \mathbb{Q}^n.$$

We then define $F(x) := \inf\{G(y) : y > x \text{ with } y \in \mathbb{Q}^n\}$ .Notice that $F(x)$ may also be defined as $\lim_{l\to\infty} G(y_l)$ where $\{y_l\}$ is any sequence in $\mathbb{Q}^n$ with $y_l > x$ and $y_l \to x$ as $l \to \infty$. Therefore for $a < b$ in $\mathbb{R}^n$ choose $a < a_l < b_l < b$ with $a_l, b_l \in \mathbb{Q}^n$ and $a_l \downarrow a$ and $b_l \uparrow b$, then

$$\begin{aligned}\Delta_{(a,b]} F &= \sum_\varepsilon (-1)^{|\varepsilon|} F(y^\varepsilon(a,b)) = \sum_\varepsilon (-1)^{|\varepsilon|} \lim_{l\to\infty} G(y^\varepsilon(a_l,b_l)) \\ &= \lim_{l\to\infty} \sum_\varepsilon (-1)^{|\varepsilon|} G(y^\varepsilon(a_l,b_l)) = \lim_{l\to\infty} \Delta_{(a_l,b_l]} G \geq 0. \qquad (23.58)\end{aligned}$$

In this way we see that $F$ defines a measure $\mu$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

The tightness assumption then guarantees that $\mu$ is a probability measure. Indeed, for $\varepsilon > 0$ there exist $a < b$ in $\mathbb{Q}^n$ such that $\mu_k((a,b]) \geq 1 - \varepsilon$ for all $k$. It then follows that

$$\Delta_{(a,b]} G := \lim_{k\to\infty} \mu_k((a,b]) \geq 1 - \varepsilon.$$

Therefore for any $\alpha < a < b < \beta$ it follows from the argument in Eq. (23.58) that $\mu((\alpha,\beta]) = \Delta_{(\alpha,\beta]} F \geq 1 - \varepsilon$.

[Clean this up!] Finally I claim that $F_n \implies F$. Indeed, suppose that $x \in \mathcal{C}(F)$ and $\varepsilon > 0$ be given. Choose $\alpha \in \mathbb{R}^n$ and $b \in \mathbb{Q}^n$ such that $\alpha < x < b$ and $F(\alpha) \leq G(b) \leq F(\alpha) + \varepsilon$. Then for any $a \in \mathbb{Q}^n$ with $\alpha < a < x < b$ we have, $F_n(a) \leq F_n(x) \leq F_n(b)$ and therefore,

$$F(\alpha) \leq G(a) = \liminf_{n\to\infty} F_n(a) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(b) = G(b)$$

Letting $\alpha \uparrow x$ then implies,

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x) + 2\varepsilon$$

and as $\varepsilon > 0$ was arbitrary, it follows that

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) = F(x).$$

■

## 23.10 Metric Space Extensions

The goal of this section is to extend the notions of weak convergence when $\mathbb{R}$ is replace by a metric space $(S,\rho)$. Standard references for the material here are [3] and [34] – also see [14] and [29]. Throughout this section, $(S,\rho)$ will be a metric space and $\mathcal{B}_S$ will be the Borel $\sigma$ – algebra on $S$, i.e. the $\sigma$ – algebra generated by the open subsets of $S$. Recall that $V\subset S$ is open if it is the union of open balls of the form

$$B(x,r) := \{y\in S : \rho(x,y) < r\}$$

where $x\in S$ and $r\geq 0$. It should be noted that if $S$ is separable (i.e. contains a countable dense set, $\Lambda\subset S$), then every open set may be written as a union of balls with $x\in\Lambda$ and $r\in\mathbb{Q}$ and so in the separable case

$$\mathcal{B}_S = \sigma(B(x,r) : x\in\Lambda,\ r\in\mathbb{Q}) = \sigma(B(x,r) : x\in S,\ r\geq 0).$$

Let us now state the theorems of this section.

**Definition 23.85.** *Let $(S,\tau)$ be a topological space, $\mathcal{B} := \sigma(\tau)$ be the Borel $\sigma$ – algebra, and $\mu$ be a probability measure on $(S,\mathcal{B})$. We say that $A\in\mathcal{B}$ is a **continuity set** for $\mu$ provided $\mu(\mathrm{bd}(A)) = 0$. Notice that this is equivalent to saying that $\mu(A^\circ) = \mu(A) = \mu(\bar{A})$.*

**Theorem 23.86 (Skorohod Theorem).** *Let $(S,\rho)$ be a separable metric space and $\{\mu_n\}_{n=0}^\infty$ be probability measures on $(S,\mathcal{B}_S)$ such that $\lim_{n\to\infty}\mu_n(A) = \mu(A)$ for all $A\in\mathcal{B}$ such that $\mu(\mathrm{bd}(A)) = 0$.*[7] *Then there exists a probability space, $(\Omega,\mathcal{B},P)$ and measurable functions, $Y_n:\Omega\to S$, such that $\mu_n = P\circ Y_n^{-1}$ for all $n\in\mathbb{N}_0 := \mathbb{N}\cup\{0\}$ and $\lim_{n\to\infty}Y_n = Y$ a.s.*

**Proposition 23.87 (The Portmanteau Theorem).** *Suppose that $S$ is a complete separable metric space and $\{\mu_n\}\cup\{\mu\}$ are probability measure on $(S,\mathcal{B} := \mathcal{B}_S)$. Then the following are equivalent:*

1. *$\mu_n \implies \mu$ as $n\to\infty$, i.e. $\mu_n(f)\to\mu(f)$ for all $f\in BC(S)$.*
2. *$\mu_n(f)\to\mu(f)$ for every $f\in BC(S)$ which is uniformly continuous.*
3. *$\limsup_{n\to\infty}\mu_n(F)\leq\mu(F)$ for all $F\sqsubset S$.*
4. *$\liminf_{n\to\infty}\mu_n(G)\geq\mu(G)$ for all $G\subset_o S$.*
5. *$\lim_{n\to\infty}\mu_n(A) = \mu(A)$ for all $A\in\mathcal{B}$ such that $\mu(\mathrm{bd}(A)) = 0$.*

**Definition 23.88.** *Let $S$ be a topological space. A collection of probability measures $\Lambda$ on $(S,\mathcal{B}_S)$ is said to be **tight** if for every $\varepsilon>0$ there exists a compact set $K_\varepsilon\in\mathcal{B}_S$ such that $\mu(K_\varepsilon)\geq 1-\varepsilon$ for all $\mu\in\Lambda$.*

[7] In Proposition 23.87 below we will see that this assumption is equivalent to assuming $\mu_n \implies \mu$.

**Theorem 23.89 (Prokhorov's Theorem).** *Suppose $S$ is a separable metrizable space and $\Lambda = \{\mu_n\}_{n=1}^\infty$ is a tight sequence of probability measures on $\mathcal{B}_{S_1}$. Then there exists a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ which is weakly convergent to a probability measure $\mu$ on $\mathcal{B}_S$.*

*Conversely, if we further assume that $(S,\rho)$ is a complete and $\Lambda$ is a sequentially compact subset of the probability measures on $(S,\mathcal{B}_S)$ with the weak topology, then $\Lambda$ is tight. (The converse direction is not so important for us.)*

For the next few exercises, let $(S_1,\rho_1)$ and $(S_2,\rho_2)$ be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel $\sigma$ – algebras on $S_1$ and $S_2$ respectively. Further define a metric, $\rho$, on $S := S_1\times S_2$ by

$$\rho((x_1,x_2),(y_1,y_2)) = \rho_1(x_1,y_1)\vee\rho_2(x_2,y_2)$$

and let $\mathcal{B}_{S_1\times S_2}$ be the Borel $\sigma$ – algebra on $S_1\times S_2$. For $i=1,2$, let $\pi_i : S_1\times S_2\to S_i$ be the projection maps and recall that

$$\mathcal{B}_{S_1}\otimes\mathcal{B}_{S_2} = \sigma(\pi_1,\pi_2) = \sigma\left(\pi_1^{-1}(\mathcal{B}_{S_1})\cup\pi_2^{-1}(\mathcal{B}_{S_2})\right).$$

**Exercise 23.10 (Continuous Mapping Theorem).** Let $(S_1,\rho_1)$ and $(S_2,\rho_2)$ be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel $\sigma$ – algebras on $S_1$ and $S_2$ respectively. Let Further suppose that $\{\mu_n\}\cup\{\mu\}$ are probability measures on $(S_1,\mathcal{B}_{S_1})$ such that $\mu_n \implies \mu$. If $f:S_1\to S_2$ is a Borel measurable function such that $\mu(\mathcal{D}(f)) = 0$ (see Notation 23.17), then $f_*\mu_n \implies f_*\mu$ where $f_*\mu := \mu\circ f^{-1}$.

**Exercise 23.11.** Prove the analogue of Lemma 8.25, namely show $\mathcal{B}_{S_1\times S_2} = \mathcal{B}_{S_1}\otimes\mathcal{B}_{S_2}$. Hint: you may find Exercise 8.6 helpful.

**Exercise 23.12.** Let $(S_1,\rho_1)$ and $(S_2,\rho_2)$ be separable metric spaces and $\mathcal{B}_{S_1}$ and $\mathcal{B}_{S_2}$ be the Borel $\sigma$ – algebras on $S_1$ and $S_2$ respectively. Further suppose that $\{\mu_n\}\cup\{\mu\}$ and $\{\nu_n\}\cup\{\nu\}$ are probability measures on $(S_1,\mathcal{B}_{S_1})$ and $(S_2,\mathcal{B}_{S_2})$ respectively. If $\mu_n \implies \mu$ and $\nu_n \implies \nu$, then $\mu_n\otimes\nu_n \implies \mu\otimes\nu$.

Exercise 23.11 and 23.12 have obvious generalizations to finite product spaces. In particular, if $\left\{X_n^{(i)}\right\}_{n=0}^\infty$ are sequences of random variables for $1\leq i\leq K$ such that for each $n$, $\left\{X_n^{(i)}\right\}_{i=1}^K$ are independent random variables with $X_n^{(i)} \implies X_0^{(i)}$ as $n\to\infty$ for each $1\leq i\leq K$, then

$$\left(X_n^{(1)},X_n^{(2)},\dots,X_n^{(K)}\right) \implies \left(X_0^{(1)},X_0^{(2)},\dots,X_0^{(K)}\right) \text{ as } n\to\infty.$$

These comments will be useful for Exercise 23.13 below.

**Definition 23.90 (Convergence of finite dimensional distributions).** *Let $\{X_n(t) : t \geq 0\}_{n=0}^{\infty}$ be a collection of random processes, $X_n(t) : \Omega \to \mathbb{R}$. We say that $X_n$ converges to $X_0$ in finite dimensional distributions and write $X_n \overset{f.d.}{\Longrightarrow} X_0$ provided for every finite subset $\Lambda := \{0 = t_0 < t_1 < t_2 < \cdots < t_K\}$ of $\mathbb{R}_+$ we have*

$$(X_n(t_0), \dots, X_n(t_K)) \implies (X_0(t_0), \dots, X_0(t_K)) \text{ as } n \to \infty.$$

**Exercise 23.13.** Let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of random variables with zero mean and $\mathrm{Var}(X_n) = 1$. For $t \geq 0$, let $W_n(t) := \frac{1}{\sqrt{n}} S_{[nt]}$ where $[nt]$ is the nearest integer to $nt$ less than or equal to $nt$ and $S_m := \sum_{k \leq m} X_k$ where $S_0 = 0$ by definition. Show that $W_n \overset{\text{f.d.}}{\Longrightarrow} B$ where $\{B(t) : t \geq 0\}$ is a Brownian motion as defined in definition 19.24. You might use the following outline.

1. For any $0 \leq s < t < \infty$, explain why $W_n(t) - W_n(s) \implies N(0, (t-s))$.
2. Given $\Lambda := \{0 = t_0 < t_1 < t_2 < \cdots < t_K\} \subset \mathbb{R}_+$ argue that $\{W_n(t_i) - W_n(t_{i-1})\}_{i=1}^K$ are independent and then show
$$\{W_n(t_i) - W_n(t_{i-1})\}_{i=1}^K \implies \{B(t_i) - B(t_{i-1})\}_{i=1}^K \text{ as } n \to \infty.$$
3. Now show that $\{W_n(t_i)\}_{i=1}^K \implies \{B(t_i)\}_{i=1}^K$ as $n \to \infty$.

The rest of this section is devoted to the proofs of the results stated at the begining of the section. (These proofs may safely be skipped on first reading.)

### 23.10.1 A point set topology review

Before getting down to business let me recall a few basic point set topology results which we will need. Recall that if $(S, \tau)$ is a topological space that $\bar{A} \subset S$, the **closure of** $A$, is defined by

$$\bar{A} := \cap \{C : A \subset C \sqsubset S\} \text{ and } A^\circ := \cup \{V : \tau \ni V \subset A\}$$

and the **interior of** $A$ is defined by

$$A^\circ = \cup \{V : \tau \ni V \subset A\}.$$

Thus $\bar{A}$ is the smallest closed set containing $A$ and $A^\circ$ is the largest open set contained in $A$. The relationship between the interior and closure operations is;

$$\begin{aligned}(A^\circ)^c &= \cap \{V^c : \tau \ni V \subset A\} \\ &= \cap \{C : A^c \subset C \sqsubset S\} = \overline{A^c}.\end{aligned}$$

Finally recall that the **topological boundary** of a set $A \subset S$ is defined by $\mathrm{bd}(A) := \bar{A} \setminus A^\circ$ which may also be expressed as

$$\mathrm{bd}(A) = \bar{A} \cap (A^\circ)^c = \bar{A} \cap \overline{A^c} \qquad (= \mathrm{bd}(A^c)).$$

In the case of a metric space we may describe $\bar{A}$ and $\mathrm{bd}(A)$ as

$$\begin{aligned}\bar{A} &= \{x \in S : \exists \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\} \text{ and} \\ \mathrm{bd}(A) &= \{x \in S : \exists \{x_n\} \subset A \text{ and } \{y_n\} \subset A^c \ni \lim_{n\to\infty} y_n = x = \lim_{n\to\infty} x_n\}.\end{aligned}$$

So the boundary of $A$ consists of those points in $S$ which are arbitrarily close to points inside of $A$ and outside of $A$. In the metric space case of most interest, the next lemma is easily proved using this characterization.

**Lemma 23.91.** *For any subsets, $A$ and $B$, of $S$ we have* $\mathrm{bd}(A \cap B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$, $\mathrm{bd}(A \setminus B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$, *and* $\mathrm{bd}(A \cup B) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B)$.

**Proof.** We begin by observing that $A^\circ \cap B^\circ \subset A \cap B \subset \bar{A} \cap \bar{B}$ from which it follows that

$$A^\circ \cap B^\circ \subset [A \cap B]^\circ \subset A \cap B \subset \overline{A \cap B} \subset \bar{A} \cap \bar{B}$$

and hence,

$$\mathrm{bd}(A \cap B) \subset \left[\bar{A} \cap \bar{B}\right] \setminus [A^\circ \cap B^\circ].$$

Combining this inclusion with

$$\begin{aligned}\left[\bar{A} \cap \bar{B}\right] \setminus [A^\circ \cap B^\circ] &= \left[\bar{A} \cap \bar{B}\right] \cap [A^\circ \cap B^\circ]^c = \left[\bar{A} \cap \bar{B}\right] \cap [(A^\circ)^c \cup (B^\circ)^c] \\ &= \left[\bar{A} \cap \bar{B} \cap (A^\circ)^c\right] \cup \left[\bar{A} \cap \bar{B} \cap (B^\circ)^c\right] \\ &\subset \left[\bar{A} \cap (A^\circ)^c\right] \cup \left[\bar{B} \cap (B^\circ)^c\right] = \mathrm{bd}(A) \cup \mathrm{bd}(B)\end{aligned}$$

completes the proof of the first assertion. The second and third assertions are easy consequence of the first because;

$$\mathrm{bd}(A \setminus B) = \mathrm{bd}(A \cap B^c) \subset \mathrm{bd}(A) \cup \mathrm{bd}(B^c) = \mathrm{bd}(A) \cup \mathrm{bd}(B)$$

and

$$\begin{aligned}\mathrm{bd}(A \cup B) &= \mathrm{bd}([A \cup B]^c) = \mathrm{bd}(A^c \cap B^c) \\ &\subset \mathrm{bd}(A^c) \cup \mathrm{bd}(B^c) = \mathrm{bd}(A) \cup \mathrm{bd}(B).\end{aligned}$$

■

### 23.10.2 Proof of Skorohod's Theorem 23.86

**Lemma 23.92.** *Let $(S,\rho)$ be a separable metric space, $\mathcal{B}$ be the Borel $\sigma$ – algebra on $S$, and $\mu$ be a probability measure on $\mathcal{B}$. Then for every $\varepsilon>0$ there exists a countable partition, $\{B_n\}_{n=1}^\infty$, of $S$ such that $B_n\in\mathcal{B}$, $\operatorname{diam}(B_n)\le\varepsilon$ and $B_n$ is a $\mu$ – continuity set (i.e. $\mu(\operatorname{bd}(B_n))=0$) for all $n$.*

**Proof.** For $x\in S$ and $r\ge 0$ let $S(x,r):=\{y\in S:\rho(x,y)=r\}$. For any finite subset, $\Gamma\subset[0,\infty)$, we have $\sum_{r\in\Gamma}S(x,r)\subset S$ and therefore,

$$\sum_{r\in\Gamma}\mu(S(x,r))\le\mu(S)=1.$$

As $\Gamma\subset_f[0,\infty)$ was arbitrary we may conclude that $\sum_{r\ge0}\mu(S(x,r))\le1<\infty$ and therefore the set $Q_x:=\{r\ge0:\mu(S(x,r))>0\}$ is at most countable.

If $B(x,r):=\{y\in S:\rho(x,y)<r\}$ and $C(x,r):=\{y\in S:\rho(x,y)\le r\}$ are the open and closed $r$ – balls about $x$ respectively, we have $S(x,r)=C(x,r)\setminus B(x,r)$. As

$$\operatorname{bd}(B(x,r))=\overline{B(x,r)}\setminus B(x,r)\subset C(x,r)\setminus B(x,r)=S(x,r),$$

it follows that $B(x,r)$ is a $\mu$ – continuity set for all $r\notin Q_x$. With these preparations in hand we are now ready to complete the proof.

Let $\{x_n\}_{n=1}^\infty$ be a countable dense subset of $S$ and let $Q:=\cup_{n=1}^\infty Q_{x_n}$ – a countable subset of $[0,\infty)$. Choose $r\in[0,\infty)\setminus Q$ such that $r\le\varepsilon/2$ and then define

$$B_n:=B(x_n,r)\setminus[B(x_1,r)\cup\dots\cup B(x_{n-1},r)].$$

It is clear that $\{B_n\}_{n=1}^\infty\subset\mathcal{B}$ is a partition of $S$ with $\operatorname{diam}(B_n)\le2r\le\varepsilon$. Moreover, we know that

$$\begin{aligned}\operatorname{bd}(B_n)&\subset\operatorname{bd}(B(x_n,r))\cup\operatorname{bd}(B(x_1,r)\cup\dots\cup B(x_{n-1},r))\\&\subset\cup_{k=1}^n\operatorname{bd}(B(x_k,r))\subset\cup_{k=1}^nS(x_k,r)\end{aligned}$$

and therefore as $r\notin Q$ we have

$$\mu(\operatorname{bd}(B_n))\le\sum_{k=1}^n\mu(S(x_k,r))=0$$

so that $B_n$ is a $\mu$ – continuity set for each $n\in\mathbb{N}$. ■

We are now ready to prove Skorohod's Theorem 23.86.

**Proof.** (of Skorohod's Theorem 23.86) We will be following the proof in Kallenberg [26, Theorem 4.30 on page 79.]. In this proof we will be using an auxiliary probability space $(\Omega_0,\mathcal{B}_0,P_0)$ which is sufficiently rich so as to support the collection of independent random variables needed in the proof.[8] The final probability space will then be given by $(\Omega,\mathcal{B},P)=(\Omega_0\times S,\mathcal{B}_0\otimes\mathcal{B}_S,P_0\otimes\mu)$ and the random variable $Y$ will be defined by $Y(\omega,x):=x$ for all $(\omega,x)\in\Omega$. Let us now start the proof.

Given $p\in\mathbb{N}$, use Lemma 23.92 to construct a partition, $\{B_n\}_{n=1}^\infty$, of $S$ such that $\operatorname{diam}(B_n)<2^{-p}$ and $\mu(\operatorname{bd}(B_n))=0$ for all $n$. Choose $m$ sufficiently large so that $\mu\left(\sum_{n=m+1}^\infty B_n\right)<2^{-p}$ and let $B_0:=\sum_{n=m+1}^\infty B_n$ so that $\{B_k\}_{k=0}^m$ is a partition of $S$. Now define

$$\kappa:=\sum_{k=0}^m k1_{B_k}(Y)=\sum_{k=0}^m k1_{Y\in B_k}$$

and let $\Theta$ be a random variable on $\Omega$ which is independent of $Y$ and has the uniform distribution on $[0,1]$. For each $n\in\mathbb{N}$, the Prenatal Skorohod Theorem 23.31 implies there exists $\tilde\kappa_n:(0,1)\times\{0,\dots,m\}\to\{0,\dots,m\}$ such that $\tilde\kappa_n(\theta,k)=k$ when $\theta\le\mu_n(B_k)/\mu(B_k)$ and $\operatorname{Law}_{m\times\{\mu(B_k)\}_{k=0}^m}(\kappa_n)=\{\mu_n(B_k)\}_{k=0}^m$. Now let $\kappa_n:=\tilde\kappa_n(\Theta,\kappa)$ so that $P(\kappa_n=k)=\mu_n(B_k)$ for all $n\in\mathbb{N}$ and $0\le k\le m$ and $\kappa_n=k$ when $\Theta\le\mu_n(B_\kappa)/\mu(B_\kappa)$. Since $\mu(\operatorname{bd}(B_k))=0$ for all $k$ it follows $\mu_n(B_k)\to\mu(B_k)$ for all $0\le k\le m$ and therefore $\lim_{n\to\infty}\kappa_n=\kappa$, $P$ – a.s.

Now choose $\xi_n^k$ independent of everything such that $P\left(\xi_n^k\in A\right)=\mu_n(A|B_k)$ for all $n$ and $0\le k\le n$. Then define

$$Y_n^p:=\xi_n^{\kappa_n(\theta,\kappa)}=\sum_{k=0}^m 1_{\kappa_n(\theta,\kappa)=k}\cdot\xi_n^k.$$

Notice that

$$\begin{aligned}P(Y_n^p\in A)&=\sum_{k=0}^m P\left(\xi_n^k\in A\ \&\ \kappa_n(\theta,\kappa)=k\right)\\&=\sum_{k=0}^m\mu_n(A|B_k)\mu_n(B_k)=\sum_{k=0}^m\mu_n(A\cap B_k)=\mu_n(A),\end{aligned}$$

and

$$\left\{\rho(Y_n^p,Y)>2^{-p}\right\}\subset\{Y\in B_0\}\cup\{\kappa\ne\kappa_n\}$$

so that

$$\begin{aligned}P\left(\cup_{n\ge N}\left\{\rho(Y_n^p,Y)>2^{-p}\right\}\right)&\le P(Y\in B_0)+P(\cup_{n\ge N}\{\kappa_n\ne\kappa\})\\&<2^{-p}+P(\cup_{n\ge N}\{\kappa_n\ne\kappa\}).\end{aligned}$$

[8] An examination of the proof will show that $\Omega_0$ can be taken to be $(0,1)\times S^{\mathbb{N}}$ equipped with a well chosen infinite product measure.

Since $\kappa_n \to \kappa$ a.s. it follows that

$$0 = P(\kappa_n \neq \kappa \text{ i.o. } n) = \lim_{N\to\infty} P\left(\cup_{n\geq N}\{\kappa_n \neq \kappa\}\right)$$

and so there exists $n_p < \infty$ such that

$$P\left(\cup_{n\geq n_p}\left\{\rho\left(Y_n^p, Y\right) > 2^{-p}\right\}\right) < 2^{-p}.$$

To finish the proof, construct $\{Y_n^p\}_{n=1}^\infty$ and $n_p \in \mathbb{N}$ as above for each $p \in \mathbb{N}$. By replacing $n_p$ by $\sum_{i=1}^p n_i$ if necessary, we may assume that $n_1 < n_2 < n_3 < \dots$ . As

$$\sum_p P\left(\cup_{n\geq n_p}\left\{\rho\left(Y_n^p, Y\right) > 2^{-p}\right\}\right) < \sum_p 2^{-p} < \infty$$

it follows from the first Borel Cantelli lemma that $P(N) = 0$ where

$$N := \left\{\left[\cup_{n\geq n_p}\left\{\rho\left(Y_n^p, Y\right) > 2^{-p}\right\}\right] \text{ i.o. } p\right\}.$$

So off the null set $N$ we have $\rho(Y_n^p, Y) \leq 2^{-p}$ for all $n \geq n_p$ and a.a. $p$. We now define $\{Y_n\}_{n=1}^\infty$ by

$$Y_n := Y_n^p \text{ for } n_p \leq n < n_{p+1} \text{ and } p \in \mathbb{N}.$$

Then by construction we have $\mathrm{Law}(Y_n) = \mu_n$ for all $n$ and $\rho(Y_n, Y) \to 0$ a.s. ∎

### 23.10.3 Proof of Proposition – The Portmanteau Theorem 23.87

**Proof.** (of Proposition 23.87.) 1. $\implies$ 2. is obvious.

For 2. $\implies$ 3., let

$$\varphi(t) := \begin{cases} 1 & \text{if } t \leq 0 \\ 1-t & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t \geq 1 \end{cases} \tag{23.59}$$

and let $f_n(x) := \varphi(n\rho(x,F))$. Then $f_n \in BC(S,[0,1])$ is uniformly continuous, $0 \leq 1_F \leq f_n$ for all $n$ and $f_n \downarrow 1_F$ as $n \to \infty$. Passing to the limit $n \to \infty$ in the equation

$$0 \leq \mu_n(F) \leq \mu_n(f_m)$$

gives

$$0 \leq \limsup_{n\to\infty} \mu_n(F) \leq \mu(f_m)$$

and then letting $m \to \infty$ in this inequality implies item 3.

3. $\iff$ 4. Assuming item 3., let $F = G^c$, then

$$\begin{aligned} 1 - \liminf_{n\to\infty} \mu_n(G) &= \limsup_{n\to\infty}(1-\mu_n(G)) = \limsup_{n\to\infty} \mu_n(G^c) \\ &\leq \mu(G^c) = 1 - \mu(G) \end{aligned}$$

which implies 4. Similarly 4. $\implies$ 3.

3. $\iff$ 5. Recall that $\mathrm{bd}(A) = \bar{A} \setminus A^o$, so if $\mu(\mathrm{bd}(A)) = 0$ and 3. (and hence also 4. holds) we have

$$\begin{aligned} \limsup_{n\to\infty} \mu_n(A) &\leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A) \text{ and} \\ \liminf_{n\to\infty} \mu_n(A) &\geq \liminf_{n\to\infty} \mu_n(A^o) \geq \mu(A^o) = \mu(A) \end{aligned}$$

from which it follows that $\lim_{n\to\infty} \mu_n(A) = \mu(A)$. Conversely, let $F \sqsubset S$ and set $F_\delta := \{x \in S : \rho(x,F) \leq \delta\}$.[9] Then

$$\mathrm{bd}(F_\delta) \subset F_\delta \setminus \{x \in S : \rho(x,F) < \delta\} = A_\delta$$

where $A_\delta := \{x \in S : \rho(x,F) = \delta\}$. Since $\{A_\delta\}_{\delta>0}$ are all disjoint, we must have

$$\sum_{\delta>0} \mu(A_\delta) \leq \mu(S) \leq 1$$

and in particular the set $\Lambda := \{\delta > 0 : \mu(A_\delta) > 0\}$ is at most countable. Let $\delta_n \notin \Lambda$ be chosen so that $\delta_n \downarrow 0$ as $n \to \infty$, then

$$\mu(F_{\delta_m}) = \lim_{n\to\infty} \mu_n(F_{\delta_m}) \geq \limsup_{n\to\infty} \mu_n(F).$$

Let $m \to \infty$ in this equation to conclude $\mu(F) \geq \limsup_{n\to\infty} \mu_n(F)$ as desired.

To finish the proof it suffices to show 5. $\implies$ 1. which is easily done using Skorohod's Theorem 23.86 just as was done in the proof of Theorem 23.32. For those not wanting to use Skorohod's theorem we also provide a direct proof that 3. $\implies$ 1.

**Alternate finish to the proof** (3. $\implies$ 1.). By an affine change of variables it suffices to consider $f \in C(S,(0,1))$ in which case we have

$$\sum_{i=1}^k \frac{(i-1)}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}} \leq f \leq \sum_{i=1}^k \frac{i}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}}. \tag{23.60}$$

Let $F_i := \left\{\frac{i}{k} \leq f\right\}$ and notice that $F_k = \emptyset$. Then for any probability $\mu$,

$$\sum_{i=1}^k \frac{(i-1)}{k}\left[\mu(F_{i-1}) - \mu(F_i)\right] \leq \mu(f) \leq \sum_{i=1}^k \frac{i}{k}\left[\mu(F_{i-1}) - \mu(F_i)\right]. \tag{23.61}$$

[9] We let $\rho(x,F) := \inf\{\rho(x,y) : y \in F\}$ so that $\rho(x,F)$ is the distance of $x$ from $F$. Recall that $\rho(\cdot,F) : S \to [0,\infty)$ is a continuous map.

Since

$$\sum_{i=1}^{k}\frac{(i-1)}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] = \sum_{i=1}^{k}\frac{(i-1)}{k}\mu(F_{i-1})-\sum_{i=1}^{k}\frac{(i-1)}{k}\mu(F_i) = \sum_{i=1}^{k-1}\frac{i}{k}\mu(F_i)-\sum_{i=1}^{k}\frac{i-1}{k}\mu(F_i)=\frac{1}{k}\sum_{i=1}^{k-1}\mu(F_i)$$

and

$$\sum_{i=1}^{k}\frac{i}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] = \sum_{i=1}^{k}\frac{i-1}{k}\left[\mu(F_{i-1})-\mu(F_i)\right]+\sum_{i=1}^{k}\frac{1}{k}\left[\mu(F_{i-1})-\mu(F_i)\right] = \sum_{i=1}^{k-1}\mu(F_i)+\frac{1}{k},$$

Eq. (23.61) becomes,

$$\frac{1}{k}\sum_{i=1}^{k-1}\mu(F_i)\le\mu(f)\le\frac{1}{k}\sum_{i=1}^{k-1}\mu(F_i)+1/k.$$

Using this equation with $\mu=\mu_n$ and then with $\mu=\mu$ we find

$$\limsup_{n\to\infty}\mu_n(f)\le\limsup_{n\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k-1}\mu_n(F_i)+1/k\right] \le\frac{1}{k}\sum_{i=1}^{k-1}\mu(F_i)+1/k\le\mu(f)+1/k.$$

Since $k$ is arbitrary, $\limsup_{n\to\infty}\mu_n(f)\le\mu(f)$. Replacing $f$ by $1-f$ in this inequality also gives $\liminf_{n\to\infty}\mu_n(f)\ge\mu(f)$ and hence we have shown $\lim_{n\to\infty}\mu_n(f)=\mu(f)$ as claimed. ■

### 23.10.4 Proof of Prokhorov's compactness Theorem 23.89

The following proof relies on results not proved in these notes up to this point. The missing results may be found by searching for "Riesz-Markov Theorem" in the notes at

http://www.math.ucsd.edu/~bdriver/240A-C-03-04/240_lecture_notes.htm.

**Proof.** (of Prokhorov's compactness Theorem 23.89) First suppose that $S$ is compact. In this case $C(S)$ is a Banach space which is separable by the Stone – Weirstrass theorem, see Exercise **??** in the analysis notes. By the Riesz theorem, Corollary **??** of the analysis notes, we know that $C(S)^*$ is in one to one correspondence with the complex measures on $(S,\mathcal{B}_S)$. We have also seen that $C(S)^*$ is metrizable and the unit ball in $C(S)^*$ is weak - * compact, see Theorem **??** of the analysis notes. Hence there exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ which is weak -* convergent to a probability measure $\mu$ on $S$. Alternatively, use the Cantor's diagonalization procedure on a countable dense set $\Gamma\subset C(S)$ so find $\{\mu_{n_k}\}_{k=1}^{\infty}$ such that $\Lambda(f):=\lim_{k\to\infty}\mu_{n_k}(f)$ exists for all $f\in\Gamma$. Then for $g\in C(S)$ and $f\in\Gamma$, we have

$$\begin{aligned}|\mu_{n_k}(g)-\mu_{n_l}(g)| &\le |\mu_{n_k}(g)-\mu_{n_k}(f)|+|\mu_{n_k}(f)-\mu_{n_l}(f)| \\ &\quad+|\mu_{n_l}(f)-\mu_{n_l}(g)| \\ &\le 2\|g-f\|_\infty+|\mu_{n_k}(f)-\mu_{n_l}(f)|\end{aligned}$$

which shows

$$\limsup_{n\to\infty}|\mu_{n_k}(g)-\mu_{n_l}(g)|\le 2\|g-f\|_\infty.$$

Letting $f\in\Lambda$ tend to $g$ in $C(S)$ shows $\limsup_{n\to\infty}|\mu_{n_k}(g)-\mu_{n_l}(g)|=0$ and hence $\Lambda(g):=\lim_{k\to\infty}\mu_{n_k}(g)$ for all $g\in C(S)$. It is now clear that $\Lambda(g)\ge 0$ for all $g\ge 0$ so that $\Lambda$ is a positive linear functional on $S$ and thus there is a probability measure $\mu$ such that $\Lambda(g)=\mu(g)$.

**General case.** By Theorem 11.71 we may assume that $S$ is a subset of a compact metric space which we will denote by $\bar S$. We now extend $\mu_n$ to $\bar S$ by setting $\bar\mu_n(A):=\bar\mu_n(A\cap S)$ for all $A\in\mathcal{B}_{\bar S}$. By what we have just proved, there is a subsequence $\{\bar\mu_k':=\bar\mu_{n_k}\}_{k=1}^{\infty}$ such that $\bar\mu_k'$ converges weakly to a probability measure $\bar\mu$ on $\bar S$. The main thing we now have to prove is that "$\bar\mu(S)=1$," this is where the tightness assumption is going to be used. Given $\varepsilon>0$, let $K_\varepsilon\subset S$ be a compact set such that $\bar\mu_n(K_\varepsilon)\ge 1-\varepsilon$ for all $n$. Since $K_\varepsilon$ is compact in $S$ it is compact in $\bar S$ as well and in particular a closed subset of $\bar S$. Therefore by Proposition 23.87

$$\bar\mu(K_\varepsilon)\ge\limsup_{k\to\infty}\bar\mu_k^{'}(K_\varepsilon)=1-\varepsilon.$$

Since $\varepsilon>0$ is arbitrary, this shows with $S_0:=\cup_{n=1}^{\infty}K_{1/n}$ satisfies $\bar\mu(S_0)=1$. Because $S_0\in\mathcal{B}_S\cap\mathcal{B}_{\bar S}$, we may view $\bar\mu$ as a measure on $\mathcal{B}_S$ by letting $\mu(A):=\bar\mu(A\cap S_0)$ for all $A\in\mathcal{B}_S$. Given a closed subset $F\subset S$, choose $\tilde F\sqsubset\bar S$ such that $F=\tilde F\cap S$. Then

$$\limsup_{k\to\infty}\mu_k'(F)=\limsup_{k\to\infty}\bar\mu_k'(\tilde F)\le\bar\mu(\tilde F)=\bar\mu(\tilde F\cap S_0)=\mu(F),$$

which shows $\mu_k'\implies\mu$.

**Converse direction.** Suppose now that $(S, \rho)$ is complete and $\Lambda$ is a sequentially compact subset of the probability measures on $(S, \mathcal{B}_S)$. We first will prove if $\{G_n\}_{n=1}^{\infty}$ is a sequence of open subsets of $S$ such that $G_n \uparrow S$, then

$$c := \sup_n \inf_{\mu \in \Lambda} \mu(G_n) = \lim_{n\to\infty} \inf_{\mu \in \Lambda} \mu(G_n) = 1.$$

Suppose for sake of contradiction that $c < 1$ and let $c' \in (c, 1)$. By our assumption we have $\inf_{\mu\in\Lambda} \mu(G_n) \le c$ for all $n$ therefore there exists $\mu_n \in \Lambda$ such that $\mu_n(G_n) \le c'$ for all $n \in \mathbb{N}$. By passing to a subsequence of $\{n\}$ and corresponding subsequence $\{G'_n\}$ of the $\{G_n\}$, we may assume that $\nu_n := \mu_{k_n} \implies \mu$ for some probability measure $\mu$ on $S$ and $\nu_n(G'_n) \le c'$ for all $n$ where $G'_n \uparrow S$ as $n \uparrow \infty$. For fixed $N \in \mathbb{N}$ we have $\nu_n(G'_N) \le \nu_n(G'_n) \le c'$ for $n \ge N$. Passing to the limit as $n \to \infty$ in these inequalities then implies

$$\mu(G'_N) \le \liminf_{n\to\infty} \nu_n(G'_N) \le c' < 1.$$

However this is absurd since $\mu(G'_N) \uparrow 1$ as $N \to \infty$ since $\mu$ is a probability measure on $S$ and $G'_N \uparrow S$ as $N \uparrow \infty$.

We may now finish the proof as follows. Let $\varepsilon > 0$ be given and let $\{x_k\}_{k=1}^{\infty}$ be a countable dense subset of $S$. For each $m \in \mathbb{N}$ the open sets $G_n := \cup_{k=1}^n B\left(x_k, \frac{1}{m}\right) \uparrow S$ and so by the above claim there exists $n_m$ such $V_m := G_{n_m}$ satisfies $\inf_k \mu_k(V_m) \ge 1 - \varepsilon 2^{-m}$. We now let $A := \cap_m V_m$ so that $\mu_k(A) \ge 1 - \varepsilon$ for all $k$. As $A$ is totally bounded and $S$ is complete, $K_\varepsilon := \bar{A}$ is the desired compact subset of $S$ such that $\mu_k(K_\varepsilon) \ge 1 - \varepsilon$ for all $k$. ■

# Characteristic Functions (Fourier Transform)

**Notation 24.1** *Given a measure $\mu$ on a measurable space, $(\Omega, \mathcal{B})$ and a function, $f \in L^1(\mu)$, we will often write $\mu(f)$ for $\int_\Omega f d\mu$.*

Let us recall Definition 10.12 here.

**Definition 24.2.** *Given a probability measure, $\mu$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, let*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\cdot x} d\mu(x)$$

*be the* ***Fourier transform or characteristic function of*** *$\mu$. If $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is a random vector on some probability space $(\Omega, \mathcal{B}, P)$, then we let $f(\lambda) := f_X(\lambda) := \mathbb{E}\left[e^{i\lambda\cdot X}\right]$. Of course, if $\mu := P \circ X^{-1}$, then $f_X(\lambda) = \hat{\mu}(\lambda)$.*

From Corollary 10.13 that we know if $\mu$ and $\nu$ are two probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ such that $\hat{\mu} = \hat{\nu}$ then $\mu = \nu$ – i.e. the Fourier transform map is injective. In this chapter we are going to, among other things, characterize those functions which are characteristic functions and we will also construct an inversion formula.

## 24.1 Basic Properties of the Characteristic Function

**Definition 24.3.** *A function $f : \mathbb{R}^n \to \mathbb{C}$ is said to be* ***positive definite****, iff $f(-\lambda) = \overline{f(\lambda)}$ for all $\lambda \in \mathbb{R}^n$ and for all $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ the matrix, $\left(\{f(\lambda_j - \lambda_k)\}_{j,.k=1}^m\right)$ is non-negative. More explicitly we require,*

$$\sum_{j,k=1}^m f(\lambda_j - \lambda_k) \xi_j \bar{\xi}_k \geq 0 \text{ for all } (\xi_1, \ldots, \xi_m) \in \mathbb{C}^m.$$

**Notation 24.4** *For $l \in \mathbb{N} \cup \{0\}$, let $C^l(\mathbb{R}^n, \mathbb{C})$ denote the vector space of functions, $f : \mathbb{R}^n \to \mathbb{C}$ which are $l$ - time continuously differentiable. More explicitly, if $\partial_j := \frac{\partial}{\partial x_j}$, then $f \in C^l(\mathbb{R}^n, \mathbb{C})$ iff the partial derivatives, $\partial_{j_1} \ldots \partial_{j_k} f$, exist and are continuous for $k = 1, 2, \ldots, l$ and all $j_1, \ldots, j_k \in \{1, 2, \ldots, n\}$.*

**Proposition 24.5 (Basic Properties of $\hat{\mu}$).** *Let $\mu$ and $\nu$ be two probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, then;*

1. *$\hat{\mu}(0) = 1$, and $|\hat{\mu}(\lambda)| \leq 1$ for all $\lambda$.*
2. *$\hat{\mu}(\lambda)$ is continuous.*
3. *$\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ for all $\lambda \in \mathbb{R}^n$ and in particular, $\hat{\mu}$ is real valued iff $\mu$ is symmetric, i.e. iff $\mu(-A) = \mu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^n}$. (If $\mu = P \circ X^{-1}$ for some random vector $X$, then $\mu$ is symmetric iff $X \overset{d}{=} -X$.)*
4. *$\hat{\mu}$ is a positive definite function.*
   *(Bochner's Theorem 24.46 below asserts that if $f$ is a function satisfying properties of $\hat{\mu}$ in items 1 – 4 above, then $f = \hat{\mu}$ for some probability measure $\mu$.)*
5. *If $\int_{\mathbb{R}^n} \|x\|^l d\mu(x) < \infty$, then $\hat{\mu} \in C^l(\mathbb{R}^n, \mathbb{C})$ and*
   $$\partial_{j_1} \ldots \partial_{j_m} \hat{\mu}(\lambda) = \int_{\mathbb{R}^n} (ix_{j_1} \ldots ix_{j_m}) e^{i\lambda\cdot x} d\mu(x) \text{ for all } m \leq l.$$
6. *If $X$ and $Y$ are independent random vectors then*
   $$f_{X+Y}(\lambda) = f_X(\lambda) f_Y(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
   *This may be alternatively expressed as*
   $$\widehat{\mu * \nu}(\lambda) = \hat{\mu}(\lambda) \hat{\nu}(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
7. *If $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, and $X : \Omega \to \mathbb{R}^n$ is a random vector, then*
   $$f_{aX+b}(\lambda) = e^{i\lambda\cdot b} f_X(a\lambda).$$

**Proof.** The proof of items 1., 2., 6., and 7. are elementary and will be left to the reader. It also easy to see that $\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ and $\hat{\mu}(\lambda) = \hat{\mu}(-\lambda)$ if $\mu$ is symmetric. Therefore if $\mu$ is symmetric, then $\hat{\mu}(\lambda)$ is real. Conversely if $\hat{\mu}(\lambda)$ is real then

$$\hat{\mu}(\lambda) = \hat{\mu}(-\lambda) = \int_{\mathbb{R}^n} e^{i\lambda\cdot x} d\nu(x) = \hat{\nu}(\lambda)$$

where $\nu(A) := \mu(-A)$. The uniqueness Corollary 10.13 then implies $\mu = \nu$, i.e. $\mu$ is symmetric. This proves item 3.

Item 5. follows by induction using Corollary 9.30. For item 4. let $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ and $(\xi_1, \ldots, \xi_m) \in \mathbb{C}^m$. Then

$$\begin{aligned}\sum_{j,k=1}^m \hat{\mu}(\lambda_j - \lambda_k)\xi_j\bar{\xi}_k &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i(\lambda_j-\lambda_k)\cdot x}\xi_j\bar{\xi}_k d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i\lambda_j\cdot x}\xi_j\overline{e^{i\lambda_k\cdot x}\xi_k} d\mu(x) \\ &= \int_{\mathbb{R}^n} \left|\sum_{j=1}^m e^{i\lambda_j\cdot x}\xi_j\right|^2 d\mu(x) \geq 0.\end{aligned}$$

■

*Example 24.6 (Example 23.3 continued.).* Let $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $\nu(A) = \mu(-A)$. Then

$$\begin{aligned}\hat{\mu}(\lambda) &= \int_0^1 e^{i\lambda x}dx = \frac{e^{i\lambda}-1}{i\lambda}, \\ \hat{\nu}(\lambda) &= \hat{\mu}(-\lambda) = \overline{\hat{\mu}(\lambda)} = \frac{e^{-i\lambda}-1}{-i\lambda}, \text{ and} \\ \widehat{\mu*\nu}(\lambda) &= \hat{\mu}(\lambda)\hat{\nu}(\lambda) = |\hat{\mu}(\lambda)|^2 = \left|\frac{e^{i\lambda}-1}{i\lambda}\right|^2 = \frac{2}{\lambda^2}[1-\cos\lambda]. \qquad (24.1)\end{aligned}$$

According to Example 23.3 we also have $d(\mu * \nu)(x) = (1-|x|)_+\,dx$ and so we may directly verify Eq. (24.1) as follows;

$$\begin{aligned}\widehat{\mu*\nu}(\lambda) &= \int_{\mathbb{R}} e^{i\lambda x}(1-|x|)_+\,dx = \int_{\mathbb{R}} \cos(\lambda x)(1-|x|)_+\,dx \\ &= 2\int_0^1 (1-x)\cos\lambda x\ dx = 2\int_0^1 (1-x)\,d\frac{\sin\lambda x}{\lambda} \\ &= -2\int_0^1 d(1-x)\,\frac{\sin\lambda x}{\lambda} = 2\int_0^1 \frac{\sin\lambda x}{\lambda}dx = 2\frac{-\cos\lambda x}{\lambda^2}|_{x=0}^{x=1} \\ &= 2\frac{1-\cos\lambda}{\lambda^2}.\end{aligned}$$

For the most part we are now going to stick to the one dimensional case, i.e. $X$ will be a random variable and $\mu$ will be a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The following Lemma is a special case of item 4. of Proposition 24.5.

**Lemma 24.7.** *Suppose $n \in \mathbb{N}$ and $X$ is random variables such that $\mathbb{E}[|X|^n] < \infty$. If $\mu = P \circ X^{-1}$ is the distribution of $X$, then $\hat{\mu}(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ is $C^n$ – differentiable and*

$$\hat{\mu}^{(l)}(\lambda) = \mathbb{E}\left[(iX)^l e^{i\lambda X}\right] = \int_{\mathbb{R}} (ix)^l e^{i\lambda x}d\mu(x) \text{ for } l = 0, 1, 2, \ldots, n.$$

*In particular it follows that*

$$\mathbb{E}\left[X^l\right] = \frac{\hat{\mu}^{(l)}(0)}{i^l}.$$

The following theorem is a partial converse to this lemma. Hence the combination of Lemma 24.7, Theorem 24.8, and Corollary 24.19 (see also Corollary 24.38 below) shows that there is a correspondence between the number of moments of $X$ and the differentiability of $f_X$.

**Theorem 24.8 (Smoothness implies integrability I).** *Let $X$ be a random variable, $m \in \{0, 1, 2, \ldots\}$, $f(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$. If $f \in C^{2m}(\mathbb{R}, \mathbb{C})$ such that $g := f^{(2m)}$ is differentiable in a neighborhood of 0 and $g''(0) = f^{(2m+2)}(0)$ exists. Then $\mathbb{E}\left[X^{2m+2}\right] < \infty$ and $f \in C^{2m+2}(\mathbb{R}, \mathbb{C})$.*

**Proof.** This will be proved by induction on $m$. Let $m \in \mathbb{N}_0$ be given and suppose that

$$u(\lambda) = \mathbb{E}\left[X^{2m}\cos(\lambda X)\right] = \mathrm{Re}\,\mathbb{E}\left[X^{2m}e^{i\lambda X}\right]$$

is differentiable in a neighborhood of 0 and further suppose that $u''(0)$ exists. Since $u$ is an even function of $\lambda$, $u'$ is an odd function of $\lambda$ near 0 and therefore $u'(0) = 0$. By the mean value theorem, to each $\lambda > 0$ with $\lambda$ near 0, there exists $0 < c_\lambda < \lambda$ such that

$$\frac{u(\lambda)-u(0)}{\lambda} = u'(c_\lambda) = u'(c_\lambda) - u'(0)$$

and so

$$\frac{u(0)-u(\lambda)}{\lambda c_\lambda} = -\frac{u'(c_\lambda)-u'(0)}{c_\lambda} \to -u''(0) \text{ as } \lambda \downarrow 0. \qquad (24.2)$$

Using

$$\lim_{\lambda\downarrow 0}\frac{1-\cos(\lambda X)}{\lambda^2} = \frac{X^2}{2}$$

and Fatou's lemma, we may pass to the limit as $\lambda \downarrow 0$ in the inequality,

$$\mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda^2}\right] \leq \mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda c_\lambda}\right] = \frac{u(0)-u(\lambda)}{\lambda c_\lambda},$$

to find

$$\frac{1}{2}\mathbb{E}\left[X^{2m+2}\right] \leq \liminf_{\lambda\downarrow 0}\frac{u(0)-u(\lambda)}{\lambda c_\lambda} = -u''(0) < \infty,$$

where the last equality is a consequence of Eq. (24.2).

With this result in hand, the theorem is now easily proved by induction. We start with $m = 0$ and recall from Proposition 24.5 that $f \in C(\mathbb{R}, \mathbb{C}))$. Assuming $f$ is differentiable in a neighborhood of 0 and $f''(0)$ exists we may apply the above result with $u = \operatorname{Re} f$ in order to learn $\mathbb{E}\left[X^2\right] < \infty$. An application of Lemma 24.7 then implies that $f \in C^2(\mathbb{R}, \mathbb{C})$. The induction step is handled in much the same way upon noting,

$$f^{(2m)}(\lambda) = (-1)^m \mathbb{E}\left[X^{2m} e^{i\lambda X}\right]$$

so that

$$u(\lambda) := (-1)^m \operatorname{Re} f^{(2m)}(\lambda) = \mathbb{E}\left[X^{2m} \cos(\lambda X)\right].$$

■

**Corollary 24.9.** *Suppose that $X$ is a $\mathbb{R}^d$ – valued random vector such that for all $\lambda \in \mathbb{R}^d$ the function*

$$f_\lambda(t) := f_X(t\lambda) = \mathbb{E}\left[e^{it\lambda \cdot X}\right]$$

*is $2m$ – times differentiable in a neighborhood of $t = 0$, then $\mathbb{E}\|X\|^{2m} < \infty$ and $f_X \in C^{2m}\left(\mathbb{R}^d, \mathbb{C}\right)$.*

**Proof.** Applying Theorem 24.8 with $X$ replaced by $\lambda \cdot X$ shows that $\mathbb{E}\left[|\lambda \cdot X|^{2m}\right] < \infty$ for all $\lambda \in \mathbb{R}^d$. In particular, taking $\lambda = e_i$ (the $i^{\text{th}}$ – standard basis vector) implies that $\mathbb{E}\left[|X_i|^{2m}\right] < \infty$ for $1 \le i \le d$. So by Minkowski's inequality;

$$\|X\|_{L^{2m}(P)} = \left\|\sum_{i=1}^d X_i e_i\right\|_{L^{2m}(P)} \le \left\|\sum_{i=1}^d |X_i| \, \|e_i\|\right\|_{L^{2m}(P)} \le \sum_{i=1}^d \|e_i\| \cdot \|X_i\|_{L^{2m}(P)} < \infty,$$

i.e. $\mathbb{E}\|X\|^{2m} < \infty$. The fact that $f_X \in C^{2m}\left(\mathbb{R}^d, \mathbb{C}\right)$ now follows from Proposition 24.5. ■

## 24.2 Examples

*Example 24.10.* If $-\infty < a < b < \infty$ and $d\mu(x) = \frac{1}{b-a} 1_{[a,b]}(x)\, dx$ then

$$\hat{\mu}(\lambda) = \frac{1}{b-a} \int_a^b e^{i\lambda x} dx = \frac{e^{i\lambda b} - e^{i\lambda a}}{i\lambda(b-a)}.$$

If $a = -c$ and $b = c$ with $c > 0$, then

$$\hat{\mu}(\lambda) = \frac{\sin \lambda c}{\lambda c}.$$

Observe that

$$\hat{\mu}(\lambda) = 1 - \frac{1}{3!}\lambda^2 c^2 + \dots$$

and therefore, $\hat{\mu}'(0) = 0$ and $\hat{\mu}''(0) = -\frac{1}{3}c^2$ and hence it follows that

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = \frac{1}{3}c^2.$$

*Example 24.11.* Suppose $Z$ is a Poisson random variable with mean $a > 0$, i.e. $P(Z = n) = e^{-a}\frac{a^n}{n!}$. Then

$$f_Z(\lambda) = \mathbb{E}\left[e^{i\lambda Z}\right] = e^{-a}\sum_{n=0}^\infty e^{i\lambda n}\frac{a^n}{n!} = e^{-a}\sum_{n=0}^\infty \frac{\left(ae^{i\lambda}\right)^n}{n!} = \exp\left(a\left(e^{i\lambda} - 1\right)\right).$$

Differentiating this result gives,

$$f_Z'(\lambda) = iae^{i\lambda}\exp\left(a\left(e^{i\lambda} - 1\right)\right) \text{ and}$$
$$f_Z''(\lambda) = \left(-a^2 e^{i2\lambda} - ae^{i\lambda}\right)\exp\left(a\left(e^{i\lambda} - 1\right)\right)$$

from which we conclude,

$$\mathbb{E}Z = \frac{1}{i}f_Z'(0) = a \text{ and } \mathbb{E}Z^2 = -f_Z''(0) = a^2 + a.$$

Therefore, $\mathbb{E}Z = a = \operatorname{Var}(Z)$.

*Example 24.12.* Suppose $T \overset{d}{=} \exp(a)$, i.e. $T \ge 0$ a.s. and $P(T \ge t) = e^{-at}$ for all $t \ge 0$. Recall that $\mu = \operatorname{Law}(T)$ is given by

$$d\mu(t) = F_T'(t)\, dt = ae^{-at}1_{t\ge 0}dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right] = \int_0^\infty ae^{-at}e^{i\lambda t}dt = \frac{a}{a - i\lambda} = \hat{\mu}(\lambda).$$

Since

$$\hat{\mu}'(\lambda) = i\frac{a}{(a - i\lambda)^2} \text{ and } \hat{\mu}''(\lambda) = -2\frac{a}{(a - i\lambda)^3}$$

it follows that

$$\mathbb{E}T = \frac{\hat{\mu}'(0)}{i} = a^{-1} \text{ and } \mathbb{E}T^2 = \frac{\hat{\mu}''(0)}{i^2} = \frac{2}{a^2}$$

and hence $\operatorname{Var}(T) = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = a^{-2}$.

*Example 24.13.* From Exercise 9.15, if $d\mu(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, then $\hat{\mu}(\lambda) = e^{-\lambda^2/2}$ and we may deduce

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = 1.$$

Recall from Section 11.9 that we have defined a random vector, $X \in \mathbb{R}^d$, to be Gaussian iff

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right).$$

We define a probability measure, $\mu$, on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ to be **Gaussian** iff there is a Gaussian random vector, $X$, such that $\operatorname{Law}(X) = \mu$. This can be expressed directly in terms of $\mu$ as; $\mu$ is **Gaussian** iff

$$\hat{\mu}(\lambda) = \exp\left(-\frac{1}{2}q(\lambda, \lambda) + i\lambda \cdot m\right) \text{ for all } \lambda \in \mathbb{R}^d$$

where

$$m := \int_{\mathbb{R}^d} x d\mu(x) \text{ and } q(\lambda, \lambda) := \int_{\mathbb{R}^d} (\lambda \cdot x)^2 d\mu(x) - (\lambda \cdot m)^2.$$

*Example 24.14.* If $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $n \in \mathbb{N}$, then $\hat{\mu}^n$ is the characteristic function of the probability measure, namely the measure

$$\mu^{*n} := \overbrace{\mu * \cdots * \mu}^{n \text{ times}}. \tag{24.3}$$

Alternatively put, if $\{X_k\}_{k=1}^n$ are i.i.d. random variables with $\mu = P \circ X_k^{-1}$, then

$$f_{X_1 + \cdots + X_n}(\lambda) = f_{X_1}^n(\lambda).$$

*Example 24.15.* Suppose that $\{\mu_n\}_{n=0}^\infty$ are probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\{p_n\}_{n=0}^\infty \subset [0,1]$ such that $\sum_{n=0}^\infty p_n = 1$. Then $\sum_{n=0}^\infty p_n \hat{\mu}_n$ is the characteristic function of the probability measure,

$$\mu := \sum_{n=0}^\infty p_n \mu_n.$$

Here is a more interesting interpretation of $\mu$. Let $\{X_n\}_{n=0}^\infty \cup \{T\}$ be independent random variables with $P \circ X_n^{-1} = \mu_n$ and $P(T = n) = p_n$ for all $n \in \mathbb{N}_0$. Then $\mu(A) = P(X_T \in A)$, where $X_T(\omega) := X_{T(\omega)}(\omega)$. Indeed,

$$\mu(A) = P(X_T \in A) = \sum_{n=0}^\infty P(X_T \in A, T = n) = \sum_{n=0}^\infty P(X_n \in A, T = n)$$
$$= \sum_{n=0}^\infty P(X_n \in A, T = n) = \sum_{n=0}^\infty p_n \mu_n(A).$$

Let us also observe that

$$\hat{\mu}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] = \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_T} : T = n\right] = \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_n} : T = n\right]$$
$$= \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X_n}\right] P(T = n) = \sum_{n=0}^\infty p_n \hat{\mu}_n(\lambda).$$

*Example 24.16.* If $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ then $\sum_{n=0}^\infty p_n \hat{\mu}^n$ is the characteristic function of a probability measure, $\nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. In this case, $\nu = \sum_{n=0}^\infty p_n \mu^{*n}$ where $\mu^{*n}$ is defined in Eq. (24.3). As an explicit example, if $a > 0$ and $p_n = \frac{a^n}{n!} e^{-a}$, then

$$\sum_{n=0}^\infty p_n \hat{\mu}^n = \sum_{n=0}^\infty \frac{a^n}{n!} e^{-a} \hat{\mu}^n = e^{-a} e^{a\hat{\mu}} = e^{a(\hat{\mu} - 1)}$$

is the characteristic function of a probability measure. In other words,

$$f_{X_T}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] = \exp\left(a\left(f_{X_1}(\lambda) - 1\right)\right).$$

## 24.3 Tail Estimates

**Lemma 24.17 (Tail Estimate).** *Let $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}$ be a random variable and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ be its characteristic function. Then for $a > 0$,*

$$P(|X| \geq a) \leq \frac{a}{2} \int_{-2/a}^{2/a} (1 - f_X(\lambda)) d\lambda = \frac{a}{2} \int_{-2/a}^{2/a} (1 - \operatorname{Re} f_X(\lambda)) d\lambda = 2 \int_{-1}^{1} \Big(1 - \operatorname{Re} f_X \tag{24.4}$$

**Proof.** Recall that the Fourier transform of the uniform distribution on $[-c, c]$ is $\frac{\sin \lambda c}{\lambda c}$ and hence

$$\frac{1}{2c} \int_{-c}^{c} f_X(\lambda) d\lambda = \frac{1}{2c} \int_{-c}^{c} \mathbb{E}\left[e^{i\lambda X}\right] d\lambda = \mathbb{E}\left[\frac{\sin cX}{cX}\right].$$

Therefore,

$$\frac{1}{2c}\int_{-c}^{c}(1-f_X(\lambda))\,d\lambda = 1-\mathbb{E}\left[\frac{\sin cX}{cX}\right] = \mathbb{E}[Y_c] \tag{24.5}$$

where

$$Y_c := 1-\frac{\sin cX}{cX}.$$

Notice that $Y_c \geq 0$ (see Eq. (24.51)) and moreover, $Y_c \geq 1/2$ if $|cX| \geq 2$ ($|\sin cX|/|cX| \leq |\sin cX|/2 \leq 1/2$ if $|cX| \geq 2$). Hence we may conclude

$$\mathbb{E}[Y_c] \geq \mathbb{E}[Y_c : |cX| \geq 2] \geq \mathbb{E}\left[\frac{1}{2} : |cX| \geq 2\right] = \frac{1}{2}P(|X| \geq 2/c).$$

Combining this estimate with Eq. (24.5) shows,

$$\frac{1}{2c}\int_{-c}^{c}(1-f_X(\lambda))\,d\lambda \geq \frac{1}{2}P(|X| \geq 2/c).$$

Taking $c = 2/a$ in this estimate proves Eq. (24.4). ∎

*Remark 24.18.* The above proof is more mysterious than it need be. To begin with if $\varphi \in L^1(\mathbb{R}^d, m)$ and $\mu$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, then by Fubini-Tonelli,

$$\begin{aligned}\int_{\mathbb{R}^d}\varphi(k)\,\hat{\mu}(k)\,dk &= \int_{\mathbb{R}^d}dk\varphi(k)\int_{\mathbb{R}^d}d\mu(x)\,e^{ik\cdot x} = \int_{\mathbb{R}^d}d\mu(x)\int_{\mathbb{R}^d}dk\varphi(k)\,e^{ik\cdot x}\\ &= \int_{\mathbb{R}^d}\hat{\varphi}(x)\,d\mu(x)\end{aligned}$$

which proves the duality relation,

$$\int_{\mathbb{R}^d}\hat{\varphi}(k)\,d\mu(k) = \int_{\mathbb{R}^d}\varphi(k)\,\hat{\mu}(k)\,dk.$$

Let us take

$$\varphi(x) = \delta_\varepsilon(x) := \frac{1}{\varepsilon^d}\delta\left(\frac{x}{\varepsilon}\right)$$

where $\delta \in L^1(\mathbb{R}^d, m)$ is non-negative and $\int_{\mathbb{R}^d}\delta(x)\,dx = 1$. In this case, making a change of variables, we find

$$\begin{aligned}\hat{\delta}_\varepsilon(k) &= \int_{\mathbb{R}^d}\delta_\varepsilon(x)\,e^{ik\cdot x}dx = \int_{\mathbb{R}^d}\frac{1}{\varepsilon^d}\delta\left(\frac{x}{\varepsilon}\right)e^{ik\cdot x}dx\\ &= \int_{\mathbb{R}^d}\delta(y)\,e^{ik\cdot\varepsilon y}dy = \hat{\delta}(\varepsilon k).\end{aligned}$$

Thus the duality relation now becomes,

$$\int_{\mathbb{R}^d}\hat{\delta}(\varepsilon k)\,d\mu(k) = \int_{\mathbb{R}^d}\delta_\varepsilon(k)\,\hat{\mu}(k)\,dk = \int_{\mathbb{R}^d}\delta(k)\,\hat{\mu}(\varepsilon k)\,dk.$$

Subtracting this equation from the simple identity,

$$\int_{\mathbb{R}^d}1d\mu(k) = \mu(\mathbb{R}^d) = 1 = \int_{\mathbb{R}^d}\delta(k)\,dk,$$

then shows,

$$\int_{\mathbb{R}^d}\left[1-\hat{\delta}(\varepsilon k)\right]d\mu(k) = \int_{\mathbb{R}^d}\left[1-\hat{\mu}(\varepsilon k)\right]\delta(k)\,dk$$

and by taking the real part of the identity,

$$\int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\delta}(\varepsilon k)\right]d\mu(k) = \int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\mu}(\varepsilon k)\right]\delta(k)\,dk.$$

Since (by the Riemann Lebesgue lemma) $\lim_{|k|\to\infty}\hat{\delta}(k) = 0$, there exists $M = M(\delta) < \infty$ such that $\left|\hat{\delta}(k)\right| \leq \frac{1}{2}$ if $|k| \geq M$ and hence

$$1-\operatorname{Re}\hat{\delta}(\varepsilon k) \geq \frac{1}{2} \text{ if } |\varepsilon k| \geq M, \text{ i.e. if } |k| \geq \frac{M}{\varepsilon}.$$

Since $\left|\hat{\delta}(k)\right| \leq 1$ for all $k$ it also follows that $1-\operatorname{Re}\hat{\delta}(\varepsilon k) \geq 0$ for all $k$. From these remarks we may conclude that

$$\int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\delta}(\varepsilon k)\right]d\mu(k) \geq \int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\delta}(\varepsilon k)\right]1_{|k|\geq\frac{M}{\varepsilon}}d\mu(k) \geq \frac{1}{2}\mu\left(|k| \geq \frac{M}{\varepsilon}\right).$$

Thus we arrive at the tail estimate

$$\mu\left(|k| \geq \frac{M}{\varepsilon}\right) \leq 2\int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\mu}(\varepsilon k)\right]\delta(k)\,dk = 2\int_{\mathbb{R}^d}\left[1-\operatorname{Re}\hat{\mu}(k)\right]\delta_\varepsilon(k)\,dk$$

which is valid for all $\varepsilon > 0$. Specializing to the case where $\delta(x) = \frac{1}{2}1_{[-1,1]}(x)$ leads to the tail estimate from before.

Another quite reasonable choice would be to take

$$\delta(x) = p_1(x) = \left(\frac{1}{2\pi}\right)^{d/2}e^{-\frac{1}{2}|x|^2} \implies \hat{\delta}(k) = e^{-\frac{1}{2}|k|^2}.$$

In this cases

$$\frac{1}{2} = e^{-\frac{1}{2}M^2} \implies \ln 2 = \frac{1}{2}M^2, \text{ i.e. } M = \sqrt{2\ln 2}$$

and so we find

$$\mu\left(|k| \geq \frac{\sqrt{2\ln 2}}{\varepsilon}\right) \leq 2\int_{\mathbb{R}^d}\left[1 - \operatorname{Re}\hat{\mu}(\varepsilon k)\right] p_1(k)\,dk$$
$$= 2\int_{\mathbb{R}^d}\left[1 - \operatorname{Re}\hat{\mu}(k)\right] p_{\varepsilon^2}(k)\,dk.$$

or by setting $a := \frac{\sqrt{2\ln 2}}{\varepsilon}$ so that $\varepsilon = \frac{\sqrt{2\ln 2}}{a}$ we find,

$$\mu(|k| \geq a) \leq 2\int_{\mathbb{R}^d}\left[1 - \operatorname{Re}\hat{\mu}\left(\frac{\sqrt{2\ln 2}}{a}k\right)\right] p_1(k)\,dk$$
$$= 2\int_{\mathbb{R}^d}\left[1 - \operatorname{Re}\hat{\mu}(k)\right] p_{2\ln 2\cdot a^{-2}}(k)\,dk.$$

Let us further note that if $\hat{\mu}(k)$ is continuously differentiable at 0, then $|\hat{\mu}(k) - 1| \leq C(|k| \wedge 1)$ and we conclude that

$$\mu\left(|k| \geq \frac{M}{\varepsilon}\right) \leq 2C\int_{\mathbb{R}^d} |\varepsilon k| \wedge 1 \cdot \delta(k)\,dk$$

and so by choosing $\delta$ to be compactly supported (say assume that $\delta(k) = 0$ if $|k| \geq 1$) it follows for $\varepsilon > 0$ sufficiently small that

$$\mu\left(|k| \geq \frac{M}{\varepsilon}\right) \leq 2C\varepsilon\int_{\mathbb{R}^d} |k| \cdot \delta(k)\,dk$$

which leads to an estimate of the form,

$$\mu(|k| \geq a) \leq \frac{K}{a} \text{ for some } K < \infty.$$

**Exercise 24.1.** Suppose now $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}^d$ is a random vector and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ is its characteristic function. Show for $a > 0$,

$$P(|X|_\infty \geq a) \leq 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} (1 - f_X(\lambda))\,d\lambda$$
$$= 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} (1 - \operatorname{Re} f_X(\lambda))\,d\lambda \qquad (24.6)$$

where $|X|_\infty = \max_i |X_i|$ and $d\lambda = d\lambda_1, \dots, d\lambda_d$.

**Exercise 24.2 (Smoothness implies integrability II).** Keeping the notation in Lemma 24.17 and letting $u(\lambda) = \operatorname{Re} f_X(\lambda) = \mathbb{E}\cos(\lambda X)$. Further suppose there exists and $\varepsilon > 0$ such that $u'(\lambda)$ exists for $|\lambda| < \varepsilon$. Since $u$ is an even function, $u'(0) = 0$. Let us further suppose there exists $\delta > 0$ and $C < \infty$ such that $|u'(\lambda)| \leq C|\lambda|^\delta$ for $|\lambda| < \varepsilon$. Show there exists $K = K(C, \delta) < \infty$ such that

$$P(|X| \geq a) \leq K(C, \delta)\left(\frac{1}{a}\right)^{1+\delta} \text{ for all } a > 2/\varepsilon. \qquad (24.7)$$

Use this estimate to show $\mathbb{E}|X| < \infty$.

**Corollary 24.19 (Smoothness implies integrability III).** *Suppose that $X$ is a random variable such that $u(\lambda) = \operatorname{Re} f_X(\lambda) = \mathbb{E}[\cos\lambda X]$ is a $C^{2m+1}(\mathbb{R})$ – function of $\lambda$. If there exists $\varepsilon, \delta, C > 0$ so that*

$$\left|u^{(2m+1)}(\lambda)\right| \leq C|\lambda|^\delta \text{ for all } |\lambda| < \varepsilon,$$

*then $\mathbb{E}|X|^{2m+1} < \infty$.*

**Proof.** By Theorem 24.8, we already know $\mathbb{E}|X|^{2m} < \infty$ and therefore,

$$u^{(2m)}(\lambda) = (-1)^m\,\mathbb{E}\left[X^{2m}\cos\lambda X\right].$$

Now let $Y$ be a random variable whose distribution is determined by

$$\mathbb{E}h(Y) = \frac{\mathbb{E}\left[X^{2m}h(X)\right]}{\mathbb{E}\left[X^{2m}\right]} \;\forall \text{ bounded } h.$$

Then

$$u_Y(\lambda) := \operatorname{Re} f_Y(\lambda) = \mathbb{E}[\cos\lambda Y] = \frac{\mathbb{E}\left[X^{2m}\cos\lambda X\right]}{\mathbb{E}\left[X^{2m}\right]} = (-1)^m\frac{u^{(2m)}(\lambda)}{\mathbb{E}\left[X^{2m}\right]}$$

and so we see that $u_Y(\lambda)$ satisfies the assumptions of Exercise 24.2. This allows us to conclude that $\mathbb{E}|Y| < \infty$ which completes the proof since,

$$\mathbb{E}|Y| = \frac{\mathbb{E}\left[X^{2m}|X|\right]}{\mathbb{E}\left[X^{2m}\right]} = \frac{\mathbb{E}|X|^{2m+1}}{\mathbb{E}\left[X^{2m}\right]}.$$

■

## 24.4 Continuity Theorem

**Theorem 24.20 (Continuity Theorem).** *Suppose that $\{\mu_n\}_{n=1}^\infty$ is a sequence of probability measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ and suppose that $f(\lambda) := \lim_{n\to\infty}\hat{\mu}_n(\lambda)$ exists for all $\lambda \in \mathbb{R}^d$. If $f$ is continuous at $\lambda = 0$, then $f$ is the characteristic function of a unique probability measure, $\mu$, on $\mathcal{B}_{\mathbb{R}^d}$ and $\mu_n \implies \mu$ as $n \to \infty$.*

**Proof.** I will give the proof when $d = 1$ and leave the straight forward extension to the $d$ – dimensional case to the reader.

By the continuity of $f$ at $\lambda = 0$, for ever $\varepsilon > 0$ we may choose $a_\varepsilon$ sufficiently large so that

$$\frac{1}{2}a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \le \varepsilon/2.$$

According to Lemma 24.17 and the DCT,

$$\begin{aligned}\mu_n(\{x : |x| \ge a_\varepsilon\}) &\le \frac{1}{2}a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re}\hat{\mu}_n(\lambda))\, d\lambda \\ &\to \frac{1}{2}a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \le \varepsilon/2 \text{ as } n \to \infty.\end{aligned}$$

Hence $\mu_n(\{x : |x| \ge a_\varepsilon\}) \le \varepsilon$ for all sufficiently large $n$, say $n \ge N$. By increasing $a_\varepsilon$ if necessary we can assure that $\mu_n(\{x : |x| \ge a_\varepsilon\}) \le \varepsilon$ for all $n$ and hence $\Gamma := \{\mu_n\}_{n=1}^\infty$ is tight.

By Theorem 23.64, we may find a subsequence, $\{\mu_{n_k}\}_{k=1}^\infty$ and a probability measure $\mu$ on $\mathcal{B}_\mathbb{R}$ such that $\mu_{n_k} \implies \mu$ as $k \to \infty$. Since $x \to e^{i\lambda x}$ is a bounded and continuous function, it follows that

$$\hat{\mu}(\lambda) = \lim_{k\to\infty} \hat{\mu}_{n_k}(\lambda) = f(\lambda) \text{ for all } \lambda \in \mathbb{R},$$

that is $f$ is the characteristic function of a probability measure, $\mu$.

We now claim that $\mu_n \implies \mu$ as $n \to \infty$. If not, we could find a bounded continuous function, $g$, such that $\lim_{n\to\infty} \mu_n(g) \ne \mu(g)$ or equivalently, there would exists $\varepsilon > 0$ and a subsequence $\{\mu'_k := \mu_{n_k}\}$ such that

$$|\mu(g) - \mu'_k(g)| \ge \varepsilon \text{ for all } k \in \mathbb{N}.$$

However by Theorem 23.64 again, there is a further subsequence, $\mu''_l = \mu'_{k_l}$ of $\mu'_k$ such that $\mu''_l \implies \nu$ for some probability measure $\nu$. Since $\hat{\nu}(\lambda) = \lim_{l\to\infty} \hat{\mu}''_l(\lambda) = f(\lambda) = \hat{\mu}(\lambda)$, it follows that $\mu = \nu$. This leads to a contradiction since,

$$\varepsilon \le \lim_{l\to\infty} |\mu(g) - \mu''_l(g)| = |\mu(g) - \nu(g)| = 0.$$

■

*Remark 24.21.* One could also use Proposition 24.43 and Bochner's Theorem 24.46 below to conclude; if $f(\lambda) := \lim_{n\to\infty} \hat{\mu}_n(\lambda)$ exists and is continuous at 0, then $f$ is the characteristic function of a probability measure. Indeed, the condition of a function being positive definite is preserved under taking pointwise limits.

*Example 24.22 (Example 24.10 continued).* For $c > 0$, let $d\mu_c(x) = \frac{1}{2c}1_{[-c,c]}(x)\, dx$. As in Example 24.10, we know $\hat{\mu}_c(\lambda) = \frac{\sin \lambda c}{\lambda c}$. In this case,

$$\mu_\infty(\lambda) := \lim_{c\uparrow\infty} \hat{\mu}_c(\lambda) = \begin{cases} 1 \text{ if } \lambda = 0 \\ 0 \text{ if } \lambda \ne 0 \end{cases}.$$

Noice that the limiting function is discontinuous and is not the characteristic function of a measure. Moreover, since

$$F_c(x) := \mu_c((-\infty, x]) = \begin{cases} 0 \text{ if } x \le -c \\ 1 \text{ if } \ x \ge c \\ \frac{x+c}{2c} \text{ if } |x| \le c \end{cases} \to F_\infty(x) = \frac{1}{2} \text{ as } c \uparrow \infty.$$

**Corollary 24.23.** *Suppose that $\{X_n\}_{n=1}^\infty \cup \{X\}$ are random vectors in $\mathbb{R}^d$, then $X_n \implies X$ iff $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ for all $\lambda \in \mathbb{R}^d$.*

**Proof.** Since $f(x) := e^{i\lambda\cdot x}$ is in $BC(\mathbb{R}^d)$ for all $\lambda \in \mathbb{R}^d$, if $X_n \implies X$ then $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$. Conversely if $\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\cdot X_n}\right] = \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ for all $\lambda \in \mathbb{R}^d$ and $\mu_n := \operatorname{Law}(X_n)$ and which is equivalent to $X_n \implies X$. ■

The proof of the next corollary is a straightforward consequence of Corollary 24.23 used for dimension $d$ and dimension 1.

**Corollary 24.24.** *Suppose that $\{X_n\}_{n=1}^\infty \cup \{X\}$ are random vectors in $\mathbb{R}^d$, then $X_n \implies X$ iff $\lambda \cdot X_n \implies \lambda \cdot X$ for all $\lambda \in \mathbb{R}^d$.*

**Lemma 24.25.** *If $\{\mu_n\}_{n=1}^\infty$ is a tight sequence of probability measures on $\mathbb{R}^d$, then the corresponding characteristic functions, $\{\hat{\mu}_n\}_{n=1}^\infty$, are equicontinuous on $\mathbb{R}^d$.*

**Proof.** By the tightness of the $\{\mu_n\}_{n=1}^\infty$, given $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that $\mu_n\left(\mathbb{R}^d \setminus [-M_\varepsilon, M_\varepsilon]^n\right) \le \varepsilon$ for all $n$. Let $\lambda, h \in \mathbb{R}^d$, then

$$\begin{aligned}|\hat{\mu}_n(\lambda + h) - \hat{\mu}_n(\lambda)| &\le \int_{\mathbb{R}^d} \left|e^{ix\cdot(\lambda+h)} - e^{ix\cdot\lambda}\right| d\mu_n(x) \\ &= \int_{\mathbb{R}^d} \left|e^{ix\cdot h} - 1\right| d\mu_n(x) \\ &\le 2\varepsilon + \sup_{x\in[-M_\varepsilon, M_\varepsilon]^n} \left|e^{ix\cdot h} - 1\right|.\end{aligned}$$

Therefore it follows that

$$\limsup_{h\to 0} \sup_{\lambda\in\mathbb{R}^d} |\hat{\mu}_n(\lambda + h) - \hat{\mu}_n(\lambda)| \le 2\varepsilon$$

and as $\varepsilon > 0$ was arbitrary the result follows. ■

**Corollary 24.26 (Uniform Convergence).** *If $\mu_n \implies \mu$ as $n \to \infty$ then $\hat{\mu}_n(\lambda) \to \hat{\mu}(\lambda)$ uniformly on compact subsets of $\mathbb{R}$ ($\mathbb{R}^n$).*

**Proof.** This is a consequence of Theorem 24.20, Lemma 24.25, and the Arzela - Ascoli Theorem **??**. For completeness here is a sketch of the proof.

Let $K$ be a compact subset of $\mathbb{R}$ ($\mathbb{R}^n$) and $\varepsilon > 0$ be given. Applying Lemma 24.25 to $\{\mu\} \cup \{\hat{\mu}_n\}$ we know that there exists $\delta > 0$ such that

$$\sup_\lambda |\hat{\mu}_n(\lambda + h) - \hat{\mu}_n(\lambda)| \le \varepsilon \text{ and } \sup_\lambda |\hat{\mu}(\lambda + h) - \hat{\mu}(\lambda)| \le \varepsilon \tag{24.8}$$

whenever $\|h\| \le \delta$. Let $F \subset K$ be a finite set such that $K \subset \cup_{\xi \in F} B(\xi, \delta)$. Since we already know that $\hat{\mu}_n \to \hat{\mu}$ pointwise we will have

$$\lim_{n\to\infty} \max_{\xi \in F} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)| = 0.$$

Since every point $\lambda \in K$ is within $\delta$ of a point in $F$ we may use Eq. (24.8) to conclude that

$$\sup_{\lambda \in K} |\hat{\mu}_n(\lambda) - \hat{\mu}(\lambda)| \le 2\varepsilon + \max_{\xi \in F} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)|$$

and therefore, $\limsup_{n\to\infty} \sup_{\lambda \in K} |\hat{\mu}_n(\lambda) - \hat{\mu}(\lambda)| \le 2\varepsilon$. As $\varepsilon > 0$ was arbitrary the result follows. ■

The following lemma will be needed before giving our first applications of the continuity theorem.

**Lemma 24.27.** *Suppose that $\{z_n\}_{n=1}^\infty \subset \mathbb{C}$ satisfies, $\lim_{n\to\infty} n z_n = \xi \in \mathbb{C}$, then*

$$\lim_{n\to\infty} (1 + z_n)^n = e^\xi.$$

**Proof.** Since $n z_n \to \xi$, it follows that $z_n \sim \frac{\xi}{n} \to 0$ as $n \to \infty$ and therefore by Lemma 24.48 below, $(1 + z_n) = e^{\ln(1+z_n)}$ and

$$\ln(1 + z_n) = z_n + O\left(z_n^2\right) = z_n + O\left(\frac{1}{n^2}\right).$$

Therefore,

$$(1 + z_n)^n = \left[e^{\ln(1+z_n)}\right]^n = e^{n \ln(1+z_n)} = e^{n\left(z_n + O\left(\frac{1}{n^2}\right)\right)} \to e^\xi \text{ as } n \to \infty.$$

■

**Proposition 24.28 (Weak Law of Large Numbers revisited).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. integrable random variables. Then $\frac{S_n}{n} \xrightarrow{P} \mathbb{E}X_1 =: c$.*

**Proof.** Let $f(\lambda) := f_{X_1}(\lambda) = \mathbb{E}\left[e^{i\lambda X_1}\right]$ in which case

$$f_{\frac{S_n}{n}}(\lambda) = \left[f\left(\frac{\lambda}{n}\right)\right]^n.$$

By Taylor's theorem (see Appendix 24.8), $f(\lambda) = 1 + k(\lambda)\lambda$ where

$$\lim_{\lambda\to 0} k(\lambda) = k(0) = f'(0) = i\mathbb{E}[X_1].$$

It now follows from Lemma 24.27 that

$$f_{\frac{S_n}{n}}(\lambda) = \left[1 + k\left(\frac{\lambda}{n}\right)\frac{\lambda}{n}\right]^n \to e^{ic\lambda} \text{ as } n \to \infty$$

which is the characteristic function of the constant random variable, $c$. By the continuity Theorem 24.20, it follows that $\frac{S_n}{n} \implies c$ and since $c$ is constant we may apply Lemma 23.28 to conclude $\frac{S_n}{n} \xrightarrow{P} c = \mathbb{E}X_1$. ■

We are now ready to continue are investigation of central limit theorems that was begun with Theorem 12.39 above.

**Theorem 24.29 (The Basic Central Limit Theorem).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random variables such that $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$. Then $\frac{S_n}{\sqrt{n}} \implies N(0,1)$.*

**Proof.** If $f(\lambda) := \mathbb{E}\left[e^{i\lambda X_1}\right]$, then by Taylor's theorem (see Appendix 24.8),

$$f(\lambda) = f(0) + f'(0)\lambda + \frac{1}{2}k(\lambda)\lambda^2 = 1 + \frac{1}{2}k(\lambda)\lambda^2 \tag{24.9}$$

where

$$\lim_{\lambda\to 0} k(\lambda) = k(0) = f''(0) = -\mathbb{E}\left[X_1^2\right] = -1.$$

Hence, using Lemma 24.27, we find

$$\mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = \left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{1}{2}k\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\lambda^2}{n}\right]^n \to e^{-\lambda^2/2}.$$

Since $e^{-\lambda^2/2}$ is the characteristic function of $N(0,1)$ (Example 24.13), the result now follows from the continuity Theorem 24.20.

**Alternative proof.** Again it suffices to show

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = e^{-\lambda^2/2} \text{ for all } \lambda \in \mathbb{R}.$$

We do this using Lemma 25.14 below as follows;

$$\begin{aligned}\left|f_{\frac{S_n}{\sqrt{n}}}(\lambda)-e^{-\lambda^2/2}\right| &= \left|\left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n-\left[e^{-\lambda^2/2n}\right]^n\right| \\ &\le n\left|f\left(\frac{\lambda}{\sqrt{n}}\right)-e^{-\lambda^2/2n}\right| \\ &= n\left|1-\frac{1}{2}\left(1+\varepsilon\left(\frac{\lambda}{\sqrt{n}}\right)\right)\frac{\lambda^2}{n}-\left(1-\frac{\lambda^2}{2n}+O\left(\frac{1}{n^2}\right)\right)\right| \\ &\to 0 \text{ as } n\to\infty.\end{aligned}$$

■

**Corollary 24.30.** *If $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random variables such that $\mathbb{E}X_1=0$ and $\mathbb{E}X_1^2=1$, then*

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}}\le y\right)-P\left(N(0,1)\le y\right)\right|\to 0 \text{ as } n\to\infty. \tag{24.10}$$

**Proof.** This is a direct consequence of Theorem 24.29 and Exercise 23.7. ■

Berry (1941) and Esseñ (1942) showed there exists a constant, $C<\infty$, such that; if $\rho^3:=\mathbb{E}|X_1|^3<\infty$, then

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}}\le y\right)-P\left(N(0,1)\le y\right)\right|\le C\left(\frac{\rho}{\sigma}\right)^3/\sqrt{n}.$$

In particular the rate of convergence is $n^{-1/2}$. The exact value of the best constant $C$ is still unknown but it is known to be less than 1. We will not prove this theorem here. However we have seen a hint that such a result should be true in Theorem 12.39 above.

*Remark 24.31 (Why normal?).* It is now a reasonable question to ask "why" is the limiting random variable normal in Theorem 24.29. One way to understand this is, if under the assumptions of Theorem 24.29, we know $\frac{S_n}{\sqrt{n}}\implies L$ where $L$ is some random variable with $\mathbb{E}L=0$ and $\mathbb{E}L^2=1$, then

$$\begin{aligned}\frac{S_{2n}}{\sqrt{2n}} &= \frac{1}{\sqrt{2}}\left(\frac{\sum_{k=1,\ k\text{ odd}}^{2n}X_j}{\sqrt{n}}+\frac{\sum_{k=1,\ k\text{ even}}^{2n}X_j}{\sqrt{n}}\right) \\ &\implies \frac{1}{\sqrt{2}}(L_1+L_2)\end{aligned} \tag{24.11}$$

where $L_1\overset{d}{=}L\overset{d}{=}L_2$ and $L_1$ and $L_2$ are independent – see Exercise 23.12. In particular this implies that

$$f(\lambda)=\left[f\left(\frac{\lambda}{\sqrt{2}}\right)\right]^2 \text{ for all } \lambda\in\mathbb{R}. \tag{24.12}$$

We could also arrive at Eq. (24.12) by passing to the limit in the identity,

$$f_{\frac{S_{2n}}{\sqrt{2n}}}(\lambda)=f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right)f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right).$$

Iterating Eq. (24.12) and then Eq. (24.9) and Lemma 24.27 above we again deduce that,

$$\begin{aligned}f(\lambda) &= \left[f\left(\frac{\lambda}{(\sqrt{2})^n}\right)\right]^{2^n} \\ &= \left[1+\frac{1}{2}k\left(\frac{\lambda}{2^{n/2}}\right)\frac{\lambda^2}{2^n}\right]^{2^n}\to e^{-\frac{1}{2}\lambda^2}=f_{N(0,1)}(\lambda).\end{aligned}$$

That is we must have $L\overset{d}{=}N(0,1)$. What we have proved is that if $L$ is any square integrable random variable with zero mean and variance equal to one such that $L\overset{d}{=}\frac{1}{\sqrt{2}}(L_1+L_2)$ where $L_1$ and $L_2$ are two independent copies of $L$, then $L\overset{d}{=}N(0,1)$.

**Theorem 24.32 (The multi-dimensional Central Limit Theorem).** *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. square integrable random vectors in $\mathbb{R}^d$ and let $m:=\mathbb{E}X_1$ and $Q=\mathbb{E}\left[(X_1-m)(X_1-m)^{\mathrm{tr}}\right]$, that is $m\in\mathbb{R}^d$ and $Q$ is the $d\times d$ matrix defined by*

$$\begin{aligned}&m_j:=\mathbb{E}(X_1)_j \text{ and} \\ &Q_{ij}:=\mathbb{E}\left[(X_1-m)_i(X_1-m)_j\right]=\operatorname{Cov}\left((X_1)_i,(X)_j\right)\end{aligned}$$

*for all $1\le i,j\le d$. Then*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n(X_k-m)\implies Z \tag{24.13}$$

*where $Z\overset{d}{=}N(0,Q)$, i.e. $Z$ is a random vector such that*

$$\mathbb{E}\left[e^{i\lambda\cdot Z}\right]=\exp\left(-\frac{1}{2}Q\lambda\cdot\lambda\right) \text{ for all } \lambda\in\mathbb{R}^d.$$

**Proof.** Let $\lambda\in\mathbb{R}^d$, then

$$\lambda \cdot Z \stackrel{d}{=} N(0, Q\lambda \cdot \lambda) \stackrel{d}{=} \sqrt{Q\lambda \cdot \lambda} \cdot N(0,1)$$

and $\{\lambda \cdot X_k\}_{k=1}^{\infty}$ are i.i.d random variables with $\mathbb{E}[\lambda \cdot X_k] = \lambda \cdot m$ and $\mathrm{Var}(\lambda \cdot X_k) = Q\lambda \cdot \lambda$. If $Q\lambda \cdot \lambda = 0$ then $\lambda \cdot X_k = \lambda \cdot m$ a.s. and $\lambda \cdot Z = 0$ a.s. and we will have

$$\lambda \cdot \left( \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \right) = 0 \implies 0 = \lambda \cdot Z. \tag{24.14}$$

If $Q\lambda \cdot \lambda > 0$ then $\left\{ \frac{\lambda \cdot X_k - \lambda \cdot m}{\sqrt{Q\lambda \cdot \lambda}} \right\}_{k=1}^{\infty}$ satisfy the hypothesis of Theorem 24.29 and therefore,

$$\frac{1}{\sqrt{Q\lambda \cdot \lambda}} \lambda \cdot \left[ \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \right] = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\lambda \cdot X_k - \lambda \cdot m}{\sqrt{Q\lambda \cdot \lambda}} \implies N(0,1)$$

which combined with Eq. (24.14) implies, for all $\lambda \in \mathbb{R}^d$ we have

$$\lambda \cdot \left[ \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \right] \implies \sqrt{Q\lambda \cdot \lambda} \cdot N(0,1) \stackrel{d}{=} \lambda \cdot Z.$$

We may now apply Corollary 24.24 to conclude that Eq. (24.13) holds. ■

## 24.5 A Fourier Transform Inversion Formula

Corollary 10.13 guarantees the injectivity of the Fourier transform on the space of probability measures. Our next goal is to find an inversion formula for the Fourier transform. To motivate the construction below, let us first recall a few facts about Fourier series. To keep our exposition as simple as possible, we now restrict ourselves to the one dimensional case.

For $L > 0$, let $e_n^L(x) := e^{-i\frac{n}{L}x}$ and let

$$(f,g)_L := \frac{1}{2\pi L} \int_{-\pi L}^{\pi L} f(x) \bar{g}(x)\, dx$$

for $f, g \in L^2([-\pi L, \pi L], dx)$. Then it is well known (and fairly elementary to prove) that $\left\{e_n^L : n \in \mathbb{Z}\right\}$ is an orthonormal basis for $L^2([-\pi L, \pi L], dx)$. In particular, if $f \in C_c(\mathbb{R})$ with $\mathrm{supp}(f) \subset [-\pi L, \pi L]$, then for $x \in [-\pi L, \pi L]$,

$$\begin{aligned} f(x) &= \sum_{n \in \mathbb{Z}} \left(f, e_n^L\right)_L e_n^L(x) = \frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \left( \int_{-\pi L}^{\pi L} f(y) e^{i\frac{n}{L}y} dy \right) e^{-i\frac{n}{L}x} \\ &= \frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{L}\right) e^{-i\frac{n}{L}x} \end{aligned} \tag{24.15}$$

where

$$\hat{f}(\lambda) = \int_{-\infty}^{\infty} f(y) e^{i\lambda y} dy.$$

Letting $L \to \infty$ in Eq. (24.15) then suggests that

$$\frac{1}{2\pi L} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{L}\right) e^{-i\frac{n}{L}x} \to \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\lambda) e^{-i\lambda x} d\lambda$$

and we are lead to expect,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\lambda) e^{-i\lambda x} d\lambda. \tag{24.16}$$

Now suppose that $f(x) = \rho(x)$ where $\rho(x)$ is a probability density for a measure $\mu$ (i.e. $d\mu(x) := \rho(x)\, dx$) so that $\hat{\rho}(\lambda) = \hat{\mu}(\lambda)$. From Eq. (24.16) we expect that

$$\begin{aligned} \mu((a,b]) &= \int_a^b \rho(x)\, dx = \int_a^b \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(\lambda) e^{-i\lambda x} d\lambda \right) dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(\lambda) \left( \int_a^b e^{-i\lambda x} dx \right) d\lambda \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(\lambda) \left( \frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda \\ &= \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \hat{\mu}(\lambda) \left( \frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda. \end{aligned} \tag{24.17}$$

We will prove this formula is essentially correct in Theorem 24.34 below. The following lemma is the key to computing the limit appearing in Eq. (24.17) which will be the heart of the proof of the inversion formula.

**Lemma 24.33.** *For $c > 0$, let*

$$S(c) := \int_{-c}^{c} \frac{\sin \lambda}{\lambda} d\lambda. \tag{24.18}$$

*Then $S(c)$ is a continuous function such that $S(c) \to \pi$ boundedly as $c \to \infty$, see Figure 24.1. Moreover for any $y \in \mathbb{R}$ we have*

$$\int_{-c}^{c} \frac{\sin \lambda y}{\lambda} d\lambda = \mathrm{sgn}(y) S(c|y|) \tag{24.19}$$

*where*

$$\mathrm{sgn}(y) = \begin{cases} 1 & \text{if } y > 0 \\ -1 & \text{if } y < 0 \\ 0 & \text{if } y = 0 \end{cases}.$$

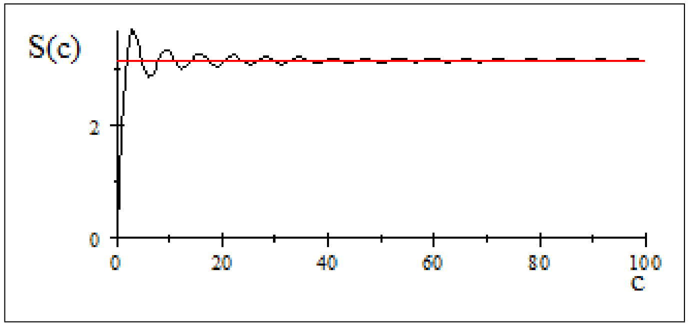



**Fig. 24.1.** The graph of $S(c)$ in black and $\pi$ in red.

**Proof.** The first assertion has already been dealt with in Example 11.11. We will repeat the argument here for the reader's convenience. By symmetry and Fubini's theorem,

$$\begin{aligned} S(c) &= 2\int_0^c \frac{\sin\lambda}{\lambda} d\lambda = 2\int_0^c \sin\lambda \cdot \left(\int_0^\infty e^{-\lambda t} dt\right) d\lambda \\ &= 2\int_0^\infty \left(\int_0^c \sin\lambda e^{-\lambda t} d\lambda\right) dt \\ &= 2\int_0^\infty \left(\frac{1}{1+t^2}\left[1 - e^{-tc}(\cos c + t\sin c)\right]\right) dt \\ &= \pi - 2\int_0^\infty \frac{1}{1+t^2} e^{-tc}\left[\cos c + t\sin c\right] dt. \end{aligned} \tag{24.20}$$

The the integral in Eq. (24.20) tends to 0 as $c \to \infty$ by the dominated convergence theorem. The second assertion in Eq. (24.19) is a consequence of the change of variables, $z = \lambda y$. ∎

**Theorem 24.34 (Fourier Inversion Formula).** *If $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $-\infty < a < b < \infty$, then*

$$\lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^c \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) d\lambda = \mu((a,b)) + \frac{1}{2}\left(\mu(\{a\}) + \mu(\{b\})\right). \tag{24.21}$$

*(At the end points, the limit picks up only half of the mass.)*

**Proof.** Let $I(c)$ denote the integral appearing in Eq. (24.21). By Fubini's theorem and Lemma 24.33,

$$\begin{aligned} I(c) &:= \int_{-c}^c \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) d\lambda \\ &= \int_{-c}^c \left(\int_\mathbb{R} e^{i\lambda x} d\mu(x)\right)\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) d\lambda \\ &= \int_\mathbb{R} d\mu(x) \int_{-c}^c d\lambda\, e^{i\lambda x}\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) \\ &= \int_\mathbb{R} d\mu(x) \int_{-c}^c d\lambda \left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda}\right). \end{aligned} \tag{24.22}$$

Since

$$\frac{e^{-i\lambda(\alpha-x)}}{i\lambda} = -\frac{i}{\lambda}\cos(\lambda(\alpha - x)) - \frac{1}{\lambda}\sin(\lambda(\alpha - x))$$

it follows that $\operatorname{Im}\left(\left[e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}\right]/i\lambda\right)$ is an odd function of $\lambda$ and

$$\operatorname{Re}\left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda}\right) = \frac{1}{\lambda}\left[\sin(\lambda(x-a)) - \sin(\lambda(x-b))\right],$$

and therefore (using Lemma 24.33)

$$\begin{aligned} I(c) &= \int_\mathbb{R} d\mu(x)\int_{-c}^c d\lambda \operatorname{Re}\left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda}\right) \\ &= \int_\mathbb{R} d\mu(x)\int_{-c}^c d\lambda\left(\frac{\sin\lambda(x-a) - \sin\lambda(x-b)}{\lambda}\right) \\ &= \int_\mathbb{R} d\mu(x)\left[\operatorname{sgn}(x-a)S(c|x-a|) - \operatorname{sgn}(x-b)S(c|x-b|)\right]. \end{aligned}$$

Using Lemma 24.33 again along with the DCT we may pass to the limit as $c \uparrow \infty$ in the previous identity to get the result;

$$\begin{aligned} \lim_{c\to\infty} \frac{1}{2\pi} I(c) &= \frac{1}{2}\int_\mathbb{R} d\mu(x)\left[\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)\right] \\ &= \frac{1}{2}\int_\mathbb{R} d\mu(x)\left[2 \cdot 1_{(a,b)}(x) + 1_{\{a\}}(x) + 1_{\{b\}}(x)\right] \\ &= \mu((a,b)) + \frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right]. \end{aligned}$$

∎

**Corollary 24.35.** *Suppose that $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\hat{\mu} \in L^1(m)$, then $d\mu = \rho dm$ where $\rho$ is the continuous probability density on $\mathbb{R}$ given by*

$$\rho(x) := \frac{1}{2\pi}\int_\mathbb{R} \hat{\mu}(\lambda) e^{-i\lambda x} d\lambda. \tag{24.23}$$

**Proof.** The function $\rho$ defined in Eq. (24.23) is continuous by the dominated convergence theorem. Moreover for any $-\infty < a < b < \infty$ we have

$$\begin{aligned}\int_a^b \rho(x)\,dx &= \frac{1}{2\pi}\int_a^b dx \int_{\mathbb{R}} d\lambda\ \hat{\mu}(\lambda)\, e^{-i\lambda x} \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} d\lambda\ \hat{\mu}(\lambda) \int_a^b dx e^{-i\lambda x} \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} d\lambda\ \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right) \\ &= \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\right)d\lambda \\ &= \mu((a,b)) + \frac{1}{2}\left[\mu(\{a\})+\mu(\{b\})\right],\end{aligned}$$

wherein we have used Theorem 24.34 to evaluate the limit. Letting $a \uparrow b$ over $a \in \mathbb{R}$ such that $\mu(\{a\}) = 0$ in this identity shows $\mu(\{b\}) = 0$ for all $b \in \mathbb{R}$. Therefore we have shown

$$\mu((a,b]) = \int_a^b \rho(x)\ dx \text{ for all } -\infty < a < b < \infty.$$

Using one of the multiplicative systems theorems, it is now easy to verify that $\mu(A) = \int_A \rho(x)\ dx$ for all $A \in \mathcal{B}_{\mathbb{R}}$ or $\int_{\mathbb{R}} h d\mu = \int_{\mathbb{R}} h\rho\, d\mu$ for all bounded measurable functions $h : \mathbb{R} \to \mathbb{R}$. This then implies that $\rho \geq 0$, $m$ – a.e.[1] and the $d\mu = \rho dm$. ■

*Example 24.36.* Let $\rho(x) = (1-|x|)_+$ be the triangle density in Figure 24.2

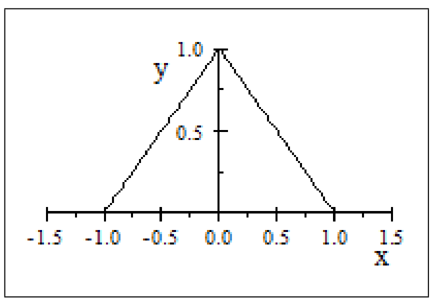



**Fig. 24.2.** The triangular density function.

Recall from Example 24.6 that

$$\int_{\mathbb{R}} e^{i\lambda x}(1-|x|)_+\,dx = 2\frac{1-\cos\lambda}{\lambda^2}.$$

Alternatively by direct calculation,

$$\begin{aligned}\int_{\mathbb{R}} e^{i\lambda x}(1-|x|)_+\,dx &= 2\,\mathrm{Re}\int_0^1 e^{i\lambda x}(1-x)\,dx \\ &= 2\,\mathrm{Re}\left[\left(I - \frac{1}{i}\frac{d}{d\lambda}\right)\int_0^1 e^{i\lambda x}dx\right] \\ &= 2\,\mathrm{Re}\left[\left(I - \frac{1}{i}\frac{d}{d\lambda}\right)\frac{e^{i\lambda}-1}{i\lambda}\right] \\ &= 2\frac{1-\cos\lambda}{\lambda^2}.\end{aligned}$$

Hence it follows[2] from Corollary 24.35 that

$$(1-|x|)_+ = \frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos\lambda}{\lambda^2}e^{-i\lambda x}d\lambda. \tag{24.24}$$

Evaluating Eq. (24.24) at $x = 0$ gives the identity

$$1 = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos\lambda}{\lambda^2}d\lambda. \tag{24.25}$$

from which we deduce that

$$d\mu(x) := \frac{1}{\pi}\frac{1-\cos x}{x^2}dx \tag{24.26}$$

is a probability measure such that (from Eq. (24.24)) has characteristic function,

$$\hat{\mu}(\lambda) = (1-|\lambda|)_+. \tag{24.27}$$

**Corollary 24.37.** *For all random variables, $X$, we have*

$$\mathbb{E}|X| = \frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\mathrm{Re}\,f_X(\lambda)}{\lambda^2}d\lambda. \tag{24.28}$$

**Proof.** For $M \in \mathbb{R}\setminus\{0\}$, make the change of variables, $\lambda \to M\lambda$ in Eq. (24.25) for find

$$|M| = \frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos(\lambda M)}{\lambda^2}d\lambda. \tag{24.29}$$

[1] Since $\rho$ is continuous we may further conclude that $\rho(x) \geq 0$ for every $x \in \mathbb{R}$.

[2] This identity could also be verified directly using residue calculus techniques from complex variables.

Observe the identity holds for $M=0$ as well. Taking $M=X$ in Eq. (24.29) and then taking expectations implies,

$$\mathbb{E}|X| = \frac{1}{\pi}\int_{\mathbb{R}} \mathbb{E}\frac{1-\cos\lambda X}{\lambda^2}d\lambda = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1-\operatorname{Re} f_X(\lambda)}{\lambda^2}d\lambda.$$

■

Suppose that we did not know the value of $c := \int_{-\infty}^{\infty}\frac{1-\cos\lambda}{\lambda^2}d\lambda$ is $\pi$, we could still proceed as above to learn

$$\mathbb{E}|X| = \frac{1}{c}\int_{\mathbb{R}} \frac{1-\operatorname{Re} f_X(\lambda)}{\lambda^2}d\lambda.$$

We could then evaluate $c$ by making a judicious choice of $X$. For example if $X \overset{d}{=} N(0,1)$, we would have on one hand

$$\mathbb{E}|X| = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x|\, e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}\int_0^\infty x e^{-x^2/2}dx = \sqrt{\frac{2}{\pi}}.$$

On the other hand, $f_X(\lambda) = e^{-\lambda^2/2}$ and so

$$\begin{aligned}\sqrt{\frac{2}{\pi}} &= -\frac{1}{c}\int_{\mathbb{R}}\left(1-e^{-\lambda^2/2}\right)d\left(\lambda^{-1}\right) = \frac{1}{c}\int_{\mathbb{R}} d\left(1-e^{-\lambda^2/2}\right)\left(\lambda^{-1}\right) \\ &= \frac{1}{c}\int_{\mathbb{R}} e^{-\lambda^2/2}d\lambda = \frac{\sqrt{2\pi}}{c}\end{aligned}$$

from which it follows, again, that $c=\pi$.

**Corollary 24.38.** *Suppose $X$ is a random variable and there exists $\varepsilon>0$ such that $u(\lambda) := \operatorname{Re} f_X(\lambda) = \mathbb{E}[\cos\lambda X]$ is continuously differentiable for $\lambda\in(-2\varepsilon,2\varepsilon)$. If we further assume that*

$$\int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda}d\lambda < \infty, \tag{24.30}$$

*then $\mathbb{E}|X|<\infty$ and $f_X\in C^1(\mathbb{R},\mathbb{C})$. (Since $u$ is even, $u'$ is odd and $u'(0)=0$. Hence if $u'(\lambda)$ were $\alpha$ – Hölder continuous for some $\alpha>0$, then Eq. (24.30) would hold.)*

**Proof.** According to Eq. (24.28)

$$\pi\cdot\mathbb{E}|X| = \int_{\mathbb{R}}\frac{1-u(\lambda)}{\lambda^2}d\lambda = \int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda + \int_{|\lambda|>\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda.$$

Since $0\le 1-u(\lambda)\le 2$ and $2/\lambda^2$ is integrable for $|\lambda|>\varepsilon$, to show $\mathbb{E}|X|<\infty$ we must show,

$$\infty > \int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda = \lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda.$$

By an integration by parts we find

$$\begin{aligned}\int_{\delta\le|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda &= \int_{\delta\le|\lambda|\le\varepsilon}(1-u(\lambda))\,d\left(-\lambda^{-1}\right) \\ &= \frac{u(\lambda)-1}{\lambda}|_\delta^\varepsilon + \frac{u(\lambda)-1}{\lambda}|_{-\varepsilon}^{-\delta} - \int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda \\ &= -\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)-1}{\varepsilon} - \frac{u(-\varepsilon)-1}{-\varepsilon} \\ &\quad + \frac{u(-\delta)-1}{-\delta} - \frac{u(\delta)-1}{\delta}. \\ &\to -\lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} + u'(0) - u'(0) \\ &\le \int_{|\lambda|\le\varepsilon}\frac{|u'(\lambda)|}{|\lambda|}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} \\ &= 2\int_0^\varepsilon\frac{|u'(\lambda)|}{\lambda}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} < \infty.\end{aligned}$$

Passing the limit as $\delta\downarrow 0$ using the fact that $u'(\lambda)$ is an odd function, we learn

$$\begin{aligned}\int_{|\lambda|\le\varepsilon}\frac{1-u(\lambda)}{\lambda^2}d\lambda &= \lim_{\delta\downarrow 0}\int_{\delta\le|\lambda|\le\varepsilon}\lambda^{-1}u'(\lambda)\,d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} \\ &\le 2\int_0^\varepsilon\frac{|u'(\lambda)|}{\lambda}d\lambda + \frac{u(\varepsilon)+u(-\varepsilon)}{\varepsilon} < \infty.\end{aligned}$$

■

## 24.6 Exercises

**Exercise 24.3.** For $x,\lambda\in\mathbb{R}$, let (also see Eq. (24.33))

$$\varphi(\lambda,x) := \begin{cases}\frac{e^{i\lambda x}-1-i\lambda x}{x^2} & \text{if } x\ne 0 \\ -\frac{1}{2}\lambda^2 & \text{if } x=0.\end{cases} \tag{24.31}$$

Let $\{x_k\}_{k=1}^n\subset\mathbb{R}\setminus\{0\}$, $\{Z_k\}_{k=1}^n\cup\{N\}$ be independent random variables with $N\overset{d}{=}N(0,1)$ and $Z_k$ being Poisson random variables with mean $a_k>0$, i.e. $P(Z_k=n) = e^{-a_k}\frac{a_k^n}{n!}$ for $n=0,1,2\dots$. With $Y := \sum_{k=1}^n x_k(Z_k-a_k)+\alpha N$, show

$$f_Y(\lambda) := \mathbb{E}\left[e^{i\lambda Y}\right] = \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right)$$

where $\nu$ is the discrete measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ given by

$$\nu = \alpha^2 \delta_0 + \sum_{k=1}^{n} a_k x_k^2 \delta_{x_k}. \tag{24.32}$$

[**Remark:** It is easy to see that $\varphi(\lambda, 0) = \lim_{x \to 0} \varphi(\lambda, x)$. In fact by Taylor's theorem with integral remainder we have

$$\varphi(\lambda, x) = -\lambda^2 \int_0^1 e^{it\lambda x}(1-t)\, dt. \tag{24.33}$$

From this formula it is clear that $\varphi$ is a smooth function of $(\lambda, x)$.]

**Exercise 24.4.** To each finite and compactly supported measure, $\nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ show there exists a sequence $\{\nu_n\}_{n=1}^{\infty}$ of finitely supported finite measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu_n \implies \nu$. Here we say $\nu$ is compactly supported if there exists $M < \infty$ such that $\nu(\{x : |x| \geq M\}) = 0$ and we say $\nu$ is finitely supported if there exists a finite subset, $\Lambda \subset \mathbb{R}$ such that $\nu(\mathbb{R} \setminus \Lambda) = 0$.

**Exercise 24.5.** Show that if $\nu$ is a finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, then

$$f(\lambda) := \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right) \tag{24.34}$$

is the characteristic function of a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Here is an outline to follow. (You may find the calculus estimates in Section 24.8 to be of help.)

1. Show $f(\lambda)$ is continuous.
2. Now suppose that $\nu$ is compactly supported. Show, using Exercises 24.3, 24.4, and the continuity Theorem 24.20 that $\exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right)$ is the characteristic function of a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
3. For the general case, approximate $\nu$ by a sequence of finite measures with compact support as in item 2.

**Exercise 24.6 (Exercise 2.3 in [44]).** Let $\mu$ be the probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, such that $\mu(\{n\}) = p(n) = c\frac{1}{n^2 \ln|n|} 1_{|n| \geq 2}$ with $c$ chosen so that $\sum_{n \in \mathbb{Z}} p(n) = 1$. Show that $\hat{\mu} \in C^1(\mathbb{R}, \mathbb{C})$ even though $\int_{\mathbb{R}} |x|\, d\mu(x) = \infty$. To do this show,

$$g(t) := \sum_{n \geq 2} \frac{1 - \cos nt}{n^2 \ln n}$$

is continuously differentiable.

**Exercise 24.7 (Polya's Criterion [2, Problem 26.3 on p. 305.] and [12, p. 104-107.]).** Suppose $\varphi(\lambda)$ is a non-negative symmetric continuous function such that $\varphi(0) = 1$, $\varphi(\lambda)$ is non-increasing and convex for $\lambda \geq 0$. Show $\varphi(\lambda) = \hat{\nu}(\lambda)$ for some probability measure, $\nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

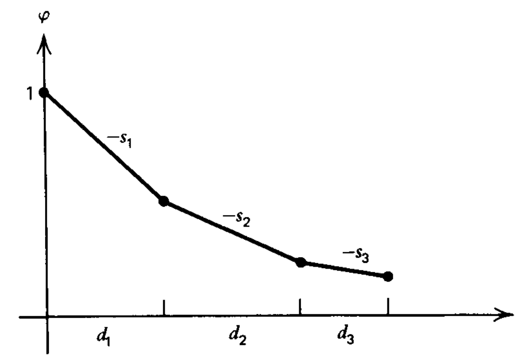



**Fig. 24.3.** Here is a piecewise linear convex function. We will assume that $d_n > 0$ for all $n$ and that $\varphi(\lambda) = 0$ for $\lambda$ sufficiently large. This last restriction may be removed later by a limiting argument.

**Exercise 24.8.** Let $d\mu(x) = \frac{1}{2} e^{-|x|} dx$. Find $\hat{\mu}(\lambda)$ and use your result to conclude,

$$\frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{1 + \lambda^2} e^{-i\lambda x} d\lambda = e^{-|x|} \ \forall\ x \in \mathbb{R}. \tag{24.35}$$

**Remark:** Another standard way to compute this integral is to use residue calculus from complex variable theory.

**Exercise 24.9.** Let

$$c := \int_{-\infty}^{\infty} \frac{(1 - \cos \lambda)^2}{\lambda^4} d\lambda, \tag{24.36}$$

and $u(\lambda) := \operatorname{Re} f_X(\lambda)$ where $f_X$ is the characteristic function of a random variable $X$. Show, using the ideas in and after Corollary 24.37 that

$$\mathbb{E}|X|^3 = \frac{1}{c} \int_{\mathbb{R}} \frac{\frac{3}{2} - 2u(\lambda) + \frac{1}{2} u(2\lambda)}{\lambda^4} d\lambda. \tag{24.37}$$

[If you are adventurous you might try to find a similar (but more complicated formula) to compute $\mathbb{E}|X|^{2k+1}$ for all $k \in \mathbb{N}_0$.]

## 24.7 Appendix: Bochner's Theorem

**Definition 24.39.** *A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have* ***rapid decay*** *or* ***rapid decrease*** *if*

$$\sup_{x\in\mathbb{R}^n} (1+|x|)^N |f(x)| < \infty \text{ for } N = 1, 2, \ldots.$$

*Equivalently, for each $N \in \mathbb{N}$ there exists constants $C_N < \infty$ such that $|f(x)| \le C_N(1+|x|)^{-N}$ for all $x \in \mathbb{R}^n$. A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have (at most)* ***polynomial growth*** *if there exists $N < \infty$ such*

$$\sup (1+|x|)^{-N} |f(x)| < \infty,$$

*i.e. there exists $N \in \mathbb{N}$ and $C < \infty$ such that $|f(x)| \le C(1+|x|)^N$ for all $x \in \mathbb{R}^n$.*

**Definition 24.40 (Schwartz Test Functions).** *Let $\mathcal{S}$ denote the space of functions $f \in C^\infty(\mathbb{R}^n)$ such that $f$ and all of its partial derivatives have rapid decay and let*

$$\|f\|_{N,\alpha} = \sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N \partial^\alpha f(x)\right|$$

*so that*

$$\mathcal{S} = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\alpha} < \infty \text{ for all } N \text{ and } \alpha \right\}.$$

*Also let $\mathcal{P}$ denote those functions $g \in C^\infty(\mathbb{R}^n)$ such that $g$ and all of its derivatives have at most polynomial growth, i.e. $g \in C^\infty(\mathbb{R}^n)$ is in $\mathcal{P}$ iff for all multi-indices $\alpha$, there exists $N_\alpha < \infty$ such*

$$\sup (1+|x|)^{-N_\alpha} |\partial^\alpha g(x)| < \infty.$$

*(Notice that any polynomial function on $\mathbb{R}^n$ is in $\mathcal{P}$.)*

**Definition 24.41.** *A function $\chi : \mathbb{R}^n \to \mathbb{C}$ is said to be* ***positive (semi) definite*** *iff the matrices $A := \{\chi(\xi_k - \xi_j)\}_{k,j=1}^m$ are positive definite for all $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$.*

**Lemma 24.42.** *If $\mu$ is a finite positive measure on $\mathcal{B}_{\mathbb{R}^n}$, then $\chi := \hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$ is a positive definite function.*

**Proof.** The dominated convergence theorem implies $\hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$. Since $\mu$ is a positive measure (and hence real),

$$\hat{\mu}(-\xi) = \int_{\mathbb{R}^n} e^{i\xi\cdot x} d\mu(x) = \overline{\int_{\mathbb{R}^n} e^{-i\xi\cdot x} d\mu(x)} = \overline{\hat{\mu}(-\xi)}.$$

From this it follows that for any $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$, the matrix $A := \{\hat{\mu}(\xi_k - \xi_j)\}_{k,j=1}^m$ is self-adjoint. Moreover if $\lambda \in \mathbb{C}^m$,

$$\begin{aligned}\sum_{k,j=1}^m \hat{\mu}(\xi_k - \xi_j)\lambda_k \bar{\lambda}_j &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i(\xi_k-\xi_j)\cdot x} \lambda_k \bar{\lambda}_j d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i\xi_k\cdot x}\lambda_k \overline{e^{-i\xi_j\cdot x}\lambda_j} d\mu(x) \\ &= \int_{\mathbb{R}^n} \left|\sum_{k=1}^m e^{-i\xi_k\cdot x}\lambda_k\right|^2 d\mu(x) \ge 0\end{aligned}$$

showing $A$ is positive definite. ■

**Proposition 24.43.** *Suppose that $\chi : \mathbb{R}^n \to \mathbb{C}$ is positive definite with $\chi(0) = 1$. If $\chi$ is continuous at 0 then in fact $\chi$ is uniformly continuous on all of $\mathbb{R}^n$.*

**Proof.** Taking $\xi_1 = x$, $\xi_2 = y$ and $\xi_3 = 0$ in Definition 24.41 we conclude that

$$A := \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \chi(y-x) & 1 & \chi(y) \\ \chi(-x) & \chi(-y) & 1 \end{bmatrix} = \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \bar{\chi}(x-y) & 1 & \chi(y) \\ \bar{\chi}(x) & \bar{\chi}(y) & 1 \end{bmatrix}$$

is positive definite. In particular,

$$\begin{aligned} 0 \le \det A = 1 &+ \chi(x-y)\chi(y)\bar{\chi}(x) + \chi(x)\bar{\chi}(x-y)\bar{\chi}(y) \\ &- |\chi(x)|^2 - |\chi(y)|^2 - |\chi(x-y)|^2. \end{aligned}$$

Combining this inequality with the identity,

$$|\chi(x) - \chi(y)|^2 = |\chi(x)|^2 + |\chi(y)|^2 - \chi(x)\bar{\chi}(y) - \chi(y)\bar{\chi}(x),$$

gives

$$\begin{aligned} 0 &\le 1 - |\chi(x-y)|^2 + \chi(x-y)\chi(y)\bar{\chi}(x) + \chi(x)\bar{\chi}(x-y)\bar{\chi}(y) \\ &\quad - \left\{ |\chi(x) - \chi(y)|^2 + \chi(x)\bar{\chi}(y) + \chi(y)\bar{\chi}(x) \right\} \\ &= 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 \\ &\quad + \chi(x-y)\chi(y)\bar{\chi}(x) - \chi(y)\bar{\chi}(x) + \chi(x)\bar{\chi}(x-y)\bar{\chi}(y) - \chi(x)\bar{\chi}(y) \\ &= 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 + 2\,\mathrm{Re}\left((\chi(x-y) - 1)\chi(y)\bar{\chi}(x)\right) \\ &\le 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 + 2\left|\chi(x-y) - 1\right|. \end{aligned}$$

Hence we have

$$\begin{aligned}|\chi(x)-\chi(y)|^2 &\le 1-|\chi(x-y)|^2+2|\chi(x-y)-1| \\ &= (1-|\chi(x-y)|)(1+|\chi(x-y)|)+2|\chi(x-y)-1| \\ &\le 4|1-\chi(x-y)|\end{aligned}$$

which completes the proof. ■

*Remark 24.44.* The function $f(\lambda)=1_{\{0\}}(\lambda)$ is positive definite since the matrix, $\{f(\lambda_i-\lambda_j)\}_{i,j=1}^n$ is the $n\times n$ identity matrix for all choices of distinct $\{\lambda_i\}_{i=1}^n$ in $\mathbb{R}$. Note however that $f$ is not continuous at $\lambda=0$.

**Lemma 24.45.** *If $\chi\in C(\mathbb{R}^n,\mathbb{C})$ is a positive definite function, then*

1. $\chi(0)\ge 0$.
2. $\chi(-\xi)=\overline{\chi(\xi)}$ *for all* $\xi\in\mathbb{R}^n$.
3. $|\chi(\xi)|\le\chi(0)$ *for all* $\xi\in\mathbb{R}^n$.
4. *If we further assume that $\chi$ is continuous, then*

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta\ge 0 \tag{24.38}$$

*for all* $f\in\mathbb{S}(\mathbb{R}^d)$.

**Proof.** Taking $m=1$ and $\xi_1=0$ we learn $\chi(0)|\lambda|^2\ge 0$ for all $\lambda\in\mathbb{C}$ which proves item 1. Taking $m=2$, $\xi_1=\xi$ and $\xi_2=\eta$, the matrix

$$A:=\begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix}$$

is positive definite from which we conclude $\chi(\xi-\eta)=\overline{\chi(\eta-\xi)}$ (since $A=A^*$ by definition) and

$$0\le\det\begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix}=|\chi(0)|^2-|\chi(\xi-\eta)|^2.$$

and hence $|\chi(\xi)|\le\chi(0)$ for all $\xi$. This proves items 2. and 3. Item 4. follows by approximating the integral in Eq. (24.38) by Riemann sums,

$$\begin{aligned}&\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta \\ &\quad=\lim_{\varepsilon\downarrow 0}\varepsilon^{-2n}\sum_{\xi,\eta\in(\varepsilon\mathbb{Z}^n)\cap[-\varepsilon^{-1},\varepsilon^{-1}]^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}\ge 0.\end{aligned}$$

The details are left to the reader keeping in mind this is where we must use the assumption that $\chi$ is continuous. ■

**Theorem 24.46 (Bochner's Theorem).** *Suppose $\chi\in C(\mathbb{R}^n,\mathbb{C})$ is positive definite function which is continuous at* 0, *then there exists a unique positive measure $\mu$ on $\mathcal{B}_{\mathbb{R}^n}$ such that $\chi=\hat{\mu}$.*

**Proof.** If $\chi(\xi)=\hat{\mu}(\xi)$, then for $f\in\mathcal{S}$ we would have

$$\int_{\mathbb{R}^n}fd\mu=\int_{\mathbb{R}^n}(f^\vee)^\wedge d\mu=\int_{\mathbb{R}^n}f^\vee(\xi)\hat{\mu}(\xi)\mathbf{d}\xi.$$

This suggests that we define

$$I(f):=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi)\mathbf{d}\xi\text{ for all }f\in\mathcal{S}.$$

We will now show $I$ is positive in the sense if $f\in\mathcal{S}$ and $f\ge 0$ then $I(f)\ge 0$. For general $f\in\mathcal{S}$ we have

$$\begin{aligned}I(|f|^2)&=\int_{\mathbb{R}^n}\chi(\xi)\left(|f|^2\right)^\vee(\xi)\mathbf{d}\xi=\int_{\mathbb{R}^n}\chi(\xi)\left(f^\vee\bigstar\bar{f}^\vee\right)(\xi)\mathbf{d}\xi \\ &=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi-\eta)\bar{f}^\vee(\eta)\mathbf{d}\eta\mathbf{d}\xi=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi-\eta)\overline{f^\vee(-\eta)}\mathbf{d}\eta\mathbf{d}\xi \\ &=\int_{\mathbb{R}^n}\chi(\xi-\eta)f^\vee(\xi)\overline{f^\vee(\eta)}\mathbf{d}\eta\mathbf{d}\xi\ge 0.\end{aligned} \tag{24.39}$$

For $t>0$ let $p_t(x):=t^{-n/2}e^{-|x|^2/2t}\in\mathcal{S}$ and define

$$I_t(x):=I\bigstar p_t(x):=I(p_t(x-\cdot))=I(\left|\sqrt{p_t(x-\cdot)}\right|^2)$$

which is non-negative by Eq. (24.39) and the fact that $\sqrt{p_t(x-\cdot)}\in\mathcal{S}$. Using

$$\begin{aligned}[p_t(x-\cdot)]^\vee(\xi)&=\int_{\mathbb{R}^n}p_t(x-y)e^{iy\cdot\xi}\mathbf{d}y=\int_{\mathbb{R}^n}p_t(y)e^{i(y+x)\cdot\xi}\mathbf{d}y \\ &=e^{ix\cdot\xi}p_t^\vee(\xi)=e^{ix\cdot\xi}e^{-t|\xi|^2/2},\end{aligned}$$

$$\begin{aligned}\langle I_t,\psi\rangle&=\int_{\mathbb{R}^n}I(p_t(x-\cdot))\psi(x)\mathbf{d}x \\ &=\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)\left[p_t(x-\cdot)\right]^\vee(\xi)\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &=\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)e^{ix\cdot\xi}e^{-t|\xi|^2/2}\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &=\int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)e^{-t|\xi|^2/2}\mathbf{d}\xi\end{aligned}$$

which coupled with the dominated convergence theorem shows

$$\langle I\bigstar p_t,\psi\rangle\to\int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)\mathbf{d}\xi=I(\psi)\text{ as }t\downarrow 0.$$

Hence if $\psi\ge 0$, then $I(\psi)=\lim_{t\downarrow 0}\langle I_t,\psi\rangle\ge 0$.

Let $K\subset\mathbb{R}$ be a compact set and $\psi\in C_c(\mathbb{R},[0,\infty))$ be a function such that $\psi=1$ on $K$. If $f\in C_c^\infty(\mathbb{R},\mathbb{R})$ is a smooth function with $\operatorname{supp}(f)\subset K$, then $0\le\|f\|_\infty\psi-f\in\mathcal{S}$ and hence

$$0\le\langle I,\|f\|_\infty\psi-f\rangle=\|f\|_\infty\langle I,\psi\rangle-\langle I,f\rangle$$

and therefore $\langle I,f\rangle\le\|f\|_\infty\langle I,\psi\rangle$. Replacing $f$ by $-f$ implies, $-\langle I,f\rangle\le\|f\|_\infty\langle I,\psi\rangle$ and hence we have proved

$$|\langle I,f\rangle|\le C(\operatorname{supp}(f))\,\|f\|_\infty \tag{24.40}$$

for all $f\in\mathcal{D}_{\mathbb{R}^n}:=C_c^\infty(\mathbb{R}^n,\mathbb{R})$ where $C(K)$ is a finite constant for each compact subset of $\mathbb{R}^n$. Because of the estimate in Eq. (24.40), it follows that $I|_{\mathcal{D}_{\mathbb{R}^n}}$ has a unique extension $I$ to $C_c(\mathbb{R}^n,\mathbb{R})$ still satisfying the estimates in Eq. (24.40) and moreover this extension is still positive. So by the Riesz – Markov Theorem **??**, there exists a unique Radon – measure $\mu$ on $\mathbb{R}^n$ such that such that $\langle I,f\rangle=\mu(f)$ for all $f\in C_c(\mathbb{R}^n,\mathbb{R})$.

To finish the proof we must show $\hat{\mu}(\eta)=\chi(\eta)$ for all $\eta\in\mathbb{R}^n$ given

$$\mu(f)=\int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi)\mathbf{d}\xi\text{ for all }f\in C_c^\infty(\mathbb{R}^n,\mathbb{R}). \tag{24.41}$$

Let $f\in C_c^\infty(\mathbb{R}^n,\mathbb{R}_+)$ be a radial function such $f(0)=1$ and $f(x)$ is decreasing as $|x|$ increases. Let $f_\varepsilon(x):=f(\varepsilon x)$, then by Theorem **??**,

$$\mathcal{F}^{-1}\left[e^{-i\eta x}f_\varepsilon(x)\right](\xi)=\varepsilon^{-n}f^\vee(\frac{\xi-\eta}{\varepsilon})$$

and therefore, from Eq. (24.41),

$$\int_{\mathbb{R}^n}e^{-i\eta x}f_\varepsilon(x)d\mu(x)=\int_{\mathbb{R}^n}\chi(\xi)\varepsilon^{-n}f^\vee(\frac{\xi-\eta}{\varepsilon})\mathbf{d}\xi. \tag{24.42}$$

Because $\int_{\mathbb{R}^n}f^\vee(\xi)\mathbf{d}\xi=\mathcal{F}f^\vee(0)=f(0)=1$, we may apply the approximate $\delta$ – function Theorem 24.47 below to Eq. (24.42) to find (using the continuity of $\chi$ here!)

$$\int_{\mathbb{R}^n}e^{-i\eta x}f_\varepsilon(x)d\mu(x)\to\chi(\eta)\text{ as }\varepsilon\downarrow 0. \tag{24.43}$$

On the the other hand, when $\eta=0$, the monotone convergence theorem implies $\mu(f_\varepsilon)\uparrow\mu(1)=\mu(\mathbb{R}^n)$ and therefore $\mu(\mathbb{R}^n)=\mu(1)=\chi(0)<\infty$. Now knowing the $\mu$ is a finite measure we may use the dominated convergence theorem to concluded

$$\mu(e^{-i\eta x}f_\varepsilon(x))\to\mu(e^{-i\eta x})=\hat{\mu}(\eta)\text{ as }\varepsilon\downarrow 0$$

for all $\eta$. Combining this equation with Eq. (24.43) shows $\hat{\mu}(\eta)=\chi(\eta)$ for all $\eta\in\mathbb{R}^n$.

**Better proof** is to use continuity Theorem 23.63 from the probability notes which is based on Helly's selection theorem. To this end, let

$$\rho_t(x):=(\chi\cdot\hat{p}_t)^\vee(x)=\int_{\mathbb{R}^n}\chi(\xi)\,\hat{p}_t(\xi)\,e^{i\xi\cdot x}\mathbf{d}\xi.$$

Notice that

$$\begin{aligned}\hat{p}_t(\xi)\,e^{i\xi\cdot x}&=e^{i\xi\cdot x}\int_{\mathbb{R}^n}p_t(y)\,e^{-iy\cdot\xi}\mathbf{d}y\\&=\int_{\mathbb{R}^n}p_t(y)\,e^{i(x-y)\cdot\xi}\mathbf{d}y\\&=\int_{\mathbb{R}^n}p_t(x+y)\,e^{-iy\cdot\xi}\mathbf{d}y\\&=[p_t(x+\cdot)]^\wedge\end{aligned}$$

so that

$$\rho_t(x)=\left\langle\chi,[p_t(x+\cdot)]^\wedge\right\rangle=\left\langle\chi,\left[\left|\sqrt{p_t(x+\cdot)}\right|^2\right]^\wedge\right\rangle\ge 0.$$

This shows that $\rho_t\ge 0$.

**Claim:** For all $t>0$,

$$\int_{\mathbb{R}^n}\rho_t(x)\,\mathbf{d}x=(\chi\cdot\hat{p}_t)(0)=\chi(0)<\infty. \tag{24.44}$$

Given the claim, we have that $d\mu_t(x)=\rho_t(x)\,\mathbf{d}x$ is a finite measure such that $\hat{\mu}_t=\chi\cdot\hat{p}_t\to\chi$ as $t\downarrow 0$ so by the continuity Theorem 24.20 it follows that $\chi=\hat{\mu}$ for some measure $\mu$. So it only remains to prove the claim.

**Proof of claim.** We start with the identity,

$$\begin{aligned}\int_{\mathbb{R}^n}\rho_t(x)\,\hat{p}_s(x)\,\mathbf{d}x&=\int_{\mathbb{R}^n}(\chi\cdot\hat{p}_t)^\vee(x)\,\hat{p}_s(x)\,\mathbf{d}x\\&=\int_{\mathbb{R}^n}(\chi\cdot\hat{p}_t)(\xi)\,p_s(\xi)\,\mathbf{d}x.\end{aligned}$$

We then make use of Fatou's lemma (using $\rho_t\ge 0$) to show,

$$\begin{aligned}\int_{\mathbb{R}^n} \rho_t(x)\,\mathbf{d}x &= \int_{\mathbb{R}^n} \liminf_{s\downarrow 0} \rho_t(x)\,\hat{p}_s(x)\,\mathbf{d}x \\ &\le \liminf_{s\downarrow 0} \int_{\mathbb{R}^n} \rho_t(x)\,\hat{p}_s(x)\,\mathbf{d}x \\ &= \liminf_{s\downarrow 0} \int_{\mathbb{R}^n} (\chi\cdot\hat{p}_t)(\xi)\,p_s(\xi)\,\mathbf{d}\xi \\ &= (\chi\cdot\hat{p}_t)(0) = \chi(0) < \infty.\end{aligned}$$

This shows $\rho_t \in L^1 \cap C_0$ and in particular $\hat{\rho}_t$ is continuous. Finally, by the $L^2$ - theory of the Fourier transform, $\hat{\rho}_t = \left[(\chi\cdot\hat{p}_t)^\vee\right]^\wedge = (\chi\cdot\hat{p}_t)$ a.e. and as both sides of this equation are continuous we have $\hat{\rho}_t(\xi) = (\chi\cdot\hat{p}_t)(\xi)$ for all $\xi \in \mathbb{R}^n$. Taking $\xi = 0$ then gives Eq. (24.44). ■

**Theorem 24.47 (Approximate $\delta -$ functions).** *Let $p \in [1,\infty]$, $\varphi \in L^1(\mathbb{R}^d)$, $a := \int_{\mathbb{R}^d} \varphi(x)dx$, and for $t > 0$ let $\varphi_t(x) = t^{-d}\varphi(x/t)$. Then*

1. *If $f \in L^p$ with $p < \infty$ then $\varphi_t * f \to af$ in $L^p$ as $t \downarrow 0$.*
2. *If $f \in BC(\mathbb{R}^d)$ and $f$ is uniformly continuous then $\|\varphi_t * f - af\|_\infty \to 0$ as $t \downarrow 0$.*
3. *If $f \in L^\infty$ and $f$ is continuous on $U \subset_o \mathbb{R}^d$ then $\varphi_t * f \to af$ uniformly on compact subsets of $U$ as $t \downarrow 0$.*

**Proof.** Making the change of variables $y = tz$ implies

$$\varphi_t * f(x) = \int_{\mathbb{R}^d} f(x-y)\varphi_t(y)dy = \int_{\mathbb{R}^d} f(x-tz)\varphi(z)dz$$

so that

$$\begin{aligned}\varphi_t * f(x) - af(x) &= \int_{\mathbb{R}^d} [f(x-tz) - f(x)]\,\varphi(z)dz \\ &= \int_{\mathbb{R}^d} [\tau_{tz}f(x) - f(x)]\,\varphi(z)dz. \qquad (24.45)\end{aligned}$$

Hence by Minkowski's inequality for integrals (Theorem **??** of the analysis notes), Proposition **??** and the dominated convergence theorem,

$$\|\varphi_t * f - af\|_p \le \int_{\mathbb{R}^d} \|\tau_{tz}f - f\|_p\,|\varphi(z)|\,dz \to 0 \text{ as } t\downarrow 0.$$

Item 2. is proved similarly. Indeed, form Eq. (24.45)

$$\|\varphi_t * f - af\|_\infty \le \int_{\mathbb{R}^d} \|\tau_{tz}f - f\|_\infty\,|\varphi(z)|\,dz$$

which again tends to zero by the dominated convergence theorem because $\lim_{t\downarrow 0}\|\tau_{tz}f - f\|_\infty = 0$ uniformly in $z$ by the uniform continuity of $f$.

Item 3. Let $B_R = B(0,R)$ be a large ball in $\mathbb{R}^d$ and $K \sqsubset\sqsubset U$, then

$$\begin{aligned}&\sup_{x\in K} |\varphi_t * f(x) - af(x)| \\ &\le \left|\int_{B_R} [f(x-tz) - f(x)]\,\varphi(z)dz\right| + \left|\int_{B_R^c} [f(x-tz) - f(x)]\,\varphi(z)dz\right| \\ &\le \int_{B_R} |\varphi(z)|\,dz \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\|f\|_\infty \int_{B_R^c} |\varphi(z)|\,dz \\ &\le \|\varphi\|_1 \cdot \sup_{x\in K, z\in B_R} |f(x-tz) - f(x)| + 2\|f\|_\infty \int_{|z|>R} |\varphi(z)|\,dz\end{aligned}$$

so that using the uniform continuity of $f$ on compact subsets of $U$,

$$\limsup_{t\downarrow 0}\sup_{x\in K} |\varphi_t * f(x) - af(x)| \le 2\|f\|_\infty \int_{|z|>R} |\varphi(z)|\,dz \to 0 \text{ as } R\to\infty.$$

■

## 24.8 Appendix: Some Calculus Estimates

We end this section by gathering together a number of calculus estimates that we will need in the future.

1. Taylor's theorem with integral remainder states, if $f \in C^k(\mathbb{R})$ and $z, \Delta \in \mathbb{R}$ or $f$ be holomorphic in a neighborhood of $z \in \mathbb{C}$ and $\Delta \in \mathbb{C}$ be sufficiently small so that $f(z + t\Delta)$ is defined for $t \in [0,1]$, then

$$\begin{aligned}f(z+\Delta) &= \sum_{n=0}^{k-1} f^{(n)}(z)\frac{\Delta^n}{n!} + \Delta^k r_k(z,\Delta) \qquad (24.46) \\ &= \sum_{n=0}^{k-1} f^{(n)}(z)\frac{\Delta^n}{n!} + \Delta^k\left[\frac{1}{k!}f^{(k)}(z) + \varepsilon(z,\Delta)\right] \qquad (24.47)\end{aligned}$$

where

$$\begin{aligned}r_k(z,\Delta) &= \frac{1}{(k-1)!}\int_0^1 f^{(k)}(z+t\Delta)(1-t)^{k-1}dt \qquad (24.48) \\ &= \frac{1}{k!}f^{(k)}(z) + \varepsilon(z,\Delta) \qquad (24.49)\end{aligned}$$

and

$$\varepsilon(z,\Delta) = \frac{1}{(k-1)!}\int_0^1 \left[f^{(k)}(z+t\Delta) - f^{(k)}(z)\right](1-t)^{k-1}dt \to 0 \text{ as } \Delta \to 0. \tag{24.50}$$

To prove this, use integration by parts to show,

$$\begin{aligned} r_k(z,\Delta) &= \frac{1}{k!}\int_0^1 f^{(k)}(z+t\Delta)\left(-\frac{d}{dt}\right)(1-t)^k dt \\ &= -\frac{1}{k!}\left[f^{(k)}(z+t\Delta)(1-t)^k\right]_{t=0}^{t=1} + \frac{\Delta}{k!}\int_0^1 f^{(k+1)}(z+t\Delta)(1-t)^k dt \\ &= \frac{1}{k!}f^{(k)}(z) + \Delta r_{k+1}(z,\Delta), \end{aligned}$$

i.e.

$$\Delta^k r_k(z,\Delta) = \frac{1}{k!}f^{(k)}(z)\Delta^k + \Delta^{k+1}r_{k+1}(z,\Delta).$$

The result now follows by induction.

2. For $y \in \mathbb{R}$, $\sin y = y\int_0^1 \cos(ty)\,dt$ and hence

$$|\sin y| \le |y|. \tag{24.51}$$

3. For $y \in \mathbb{R}$ we have

$$\cos y = 1 + y^2\int_0^1 -\cos(ty)(1-t)\,dt \ge 1 + y^2\int_0^1 -(1-t)\,dt = 1 - \frac{y^2}{2}.$$

Equivalently put[3],

$$g(y) := \cos y - 1 + y^2/2 \ge 0 \text{ for all } y \in \mathbb{R}. \tag{24.52}$$

4. Since

$$\begin{aligned} |e^z - 1 - z| &= \left|z^2\int_0^1 e^{tz}(1-t)\,dt\right| \\ &\le |z|^2\int_0^1 e^{t\operatorname{Re} z}(1-t)\,dt \\ &\le |z|^2\int_0^1 e^{0\vee\operatorname{Re} z}(1-t)\,dt \end{aligned}$$

we have shown

$$|e^z - 1 - z| \le e^{0\vee\operatorname{Re} z}\cdot\frac{|z|^2}{2}. \tag{24.53}$$

In particular if $\operatorname{Re} z \le 0$, then

$$|e^z - 1 - z| \le |z|^2/2. \tag{24.54}$$

5. Since $e^{iy} - 1 = iy\int_0^1 e^{ity}dt$, $\left|e^{iy}-1\right| \le |y|$ and hence

$$\left|e^{iy}-1\right| \le 2\wedge|y| \text{ for all } y \in \mathbb{R}. \tag{24.55}$$

**Lemma 24.48.** *For $z = re^{i\theta}$ with $-\pi < \theta < \pi$ and $r > 0$, let $\ln z = \ln r + i\theta$. Then $\ln : \mathbb{C}\setminus(-\infty,0] \to \mathbb{C}$ is a holomorphic function such that $e^{\ln z} = z$*[4] *and if $|z| < 1$ then*

$$|\ln(1+z) - z| \le |z|^2\frac{1}{2(1-|z|)^2} \text{ for } |z| < 1. \tag{24.56}$$

**Proof.** Clearly $e^{\ln z} = z$ and $\ln z$ is continuous. Therefore by the inverse function theorem for holomorphic functions, $\ln z$ is holomorphic and

$$z\frac{d}{dz}\ln z = e^{\ln z}\frac{d}{dz}\ln z = 1.$$

Therefore, $\frac{d}{dz}\ln z = \frac{1}{z}$ and $\frac{d^2}{dz^2}\ln z = -\frac{1}{z^2}$. So by Taylor's theorem,

[3] Alternatively,

$$|\sin y| = \left|\int_0^y \cos x\,dx\right| \le \left|\int_0^y |\cos x|\,dx\right| \le |y|$$

and for $y \ge 0$ we have,

$$\cos y - 1 = \int_0^y -\sin x\,dx \ge \int_0^y -x\,dx = -y^2/2.$$

This last inequality may also be proved as a simple calculus exercise following from; $g(\pm\infty) = \infty$ and $g'(y) = 0$ iff $\sin y = y$ which happens iff $y = 0$.

[4] For the purposes of this lemma it suffices to define $\ln(1+z) = -\sum_{n=1}^\infty (-z)^n/n$ and to then observe: 1)

$$\frac{d}{dz}\ln(1+z) = \sum_{n=0}^\infty (-z)^n = \frac{1}{1+z},$$

and 2) the functions $1+z$ and $e^{\ln(1+z)}$ both solve

$$f'(z) = \frac{1}{1+z}f(z) \text{ with } f(0) = 1$$

and therefore $e^{\ln(1+z)} = 1+z$.

$$\ln(1+z) = z - z^2 \int_0^1 \frac{1}{(1+tz)^2}(1-t)\,dt. \tag{24.57}$$

If $t \geq 0$ and $|z| < 1$, then

$$\left|\frac{1}{(1+tz)}\right| \leq \sum_{n=0}^{\infty} |tz|^n = \frac{1}{1-t|z|} \leq \frac{1}{1-|z|}.$$

and therefore,

$$\left|\int_0^1 \frac{1}{(1+tz)^2}(1-t)\,dt\right| \leq \frac{1}{2(1-|z|)^2}. \tag{24.58}$$

Eq. (24.56) is now a consequence of Eq. (24.57) and Eq. (24.58).

**Alternative formulation and proof.** For the purposes of this lemma we could have defined $\ln(1+z)$ by

$$\ln(1+z) = -\sum_{n=1}^{\infty} (-z)^n/n \text{ for all } |z| < 1.$$

We then observe that $\ln(1+0) = 0$ and

$$\frac{d}{dz}\ln(1+z) = \sum_{n=0}^{\infty} (-z)^n = \frac{1}{1+z}.$$

Using these observations and the chain rule we find that $f(z) := e^{\ln(1+z)}$ solves the differential equation,

$$f'(z) = \frac{1}{1+z} f(z) \text{ with } f(0) = 1. \tag{24.59}$$

Since this equation has a unique solution for $|z| < 1$ and $f(z) = 1+z$ also solves the equation we may conclude that

$$e^{\ln(1+z)} = 1+z.$$

More explicitly if $f$ solves Eq. (24.59) then

$$\frac{d}{dz}\frac{f(z)}{1+z} = \frac{f'(z)(1+z) - f(z)}{(1+z)^2} = 0$$

and hence

$$\frac{f(z)}{1+z} = \frac{f(0)}{1+0} = 1 \implies f(z) = 1+z.$$

We may now conclude that

$$|\ln(1+z) - z| = \left|-\sum_{n=2}^{\infty} (-z)^n/n\right| \leq \frac{1}{2}\sum_{n=2}^{\infty} |z|^n = \frac{1}{2}\frac{|z|^2}{1-|z|}$$

which is the estimate in Eq. (24.56). ∎

**Lemma 24.49.** *For all $y \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$,*

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \leq \frac{|y|^{n+1}}{(n+1)!} \tag{24.60}$$

*and in particular,*

$$\left|e^{iy} - 1\right| \leq |y| \wedge 2 \tag{24.61}$$

*and*

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \leq y^2 \wedge \frac{|y|^3}{3!}. \tag{24.62}$$

*More generally for all $n \in \mathbb{N}$ we have*

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| \leq \frac{|y|^{n+1}}{(n+1)!} \wedge \frac{2|y|^n}{n!}. \tag{24.63}$$

**Proof.** By Taylor's theorem (see Eq. (24.46) with $f(y) = e^{iy}$, $x = 0$ and $\Delta = y$) we have

$$\left|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}\right| = \left|\frac{y^{n+1}}{n!}\int_0^1 i^{n+1} e^{ity}(1-t)^n\,dt\right| \leq \frac{|y|^{n+1}}{n!}\int_0^1 (1-t)^n\,dt = \frac{|y|^{n+1}}{(n+1)!}$$

which is Eq. (24.60). Using Eq. (24.60) with $n = 0$ and the simple estimate; $\left|e^{iy} - 1\right| \leq 2$ gives Eq. (24.61). Similarly, Eq. (24.60) follows from the estimates coming from Eq. (24.60) with $n = 1$ and $n = 2$ respectively;

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \leq \left|e^{iy} - (1+iy)\right| + \left|\frac{y^2}{2}\right| \leq \left|\frac{y^2}{2}\right| + \left|\frac{y^2}{2}\right| = y^2$$

and

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \leq \frac{|y|^3}{3!}.$$

Equation (24.63) is proved similarly and hence will be omitted. ∎

**Lemma 24.50.** *If $X$ is a square integrable random variable, then*

$$\left|f(\lambda) - \left(1 + i\lambda\mathbb{E}X - \frac{\lambda^2}{2!}\mathbb{E}\left[X^2\right]\right)\right| \leq \mathbb{E}\left|e^{i\lambda X} - \left(1 + i\lambda X - \lambda^2\frac{X^2}{2!}\right)\right| \leq \lambda^2\varepsilon(\lambda) \tag{24.64}$$

*where*

$$\varepsilon(\lambda) := \mathbb{E}\left[X^2 \wedge \frac{|\lambda| |X|^3}{3!}\right] \to 0 \text{ as } \lambda \to 0. \tag{24.65}$$

**Proof.** Using Eq. (24.62) with $y = \lambda X$ and taking expectations gives Eq. (24.64). The DCT, with $X^2 \in L^1(P)$ being the dominating function, allows us to conclude that $\lim_{\varepsilon \to 0} \varepsilon(\lambda) = 0$. ■

# Weak Convergence of Random Sums

Throughout this chapter, we will assume the following standing notation unless otherwise stated. For each $n \in \mathbb{N}$, let $\{X_{n,k}\}_{k=1}^n$ be independent random variables and let

$$S_n := \sum_{k=1}^n X_{n,k} \text{ and} \tag{25.1}$$

$$f_{nk}(\lambda) := \mathbb{E}\left[e^{i\lambda X_{n,k}}\right] \text{ (characteristic function of } X_{n,k}). \tag{25.2}$$

The goal of this chapter is to discuss some of the possible weak limits of such $\{S_n\}_{n=1}^{\infty}$. under various conditions. The first question is what sort of limiting distributions can we expect to get. One answer is that the distribution should be infinitely divisible.

**Definition 25.1.** *A probability distribution, $\mu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is* ***infinitely divisible*** *iff for all $n \in \mathbb{N}$ there exists i.i.d. nondegenerate random variables, $\{X_{n,k}\}_{k=1}^n$, such that $X_{n,1} + \dots + X_{n,n} \overset{d}{=} \mu$. This can be formulated in the following two equivalent ways. For all $n \in \mathbb{N}$ there should exists a non-degenerate probability measure, $\mu_n$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_n^{*n} = \mu$. For all $n \in \mathbb{N}$, $\hat{\mu}(\lambda) = [g_n(\lambda)]^n$ for some non-constant characteristic function, $g_n$.*

*Remark 25.2.* If $\mu, \nu \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ are infinitely divisible then $\mu * \nu$ is as well. Stated another way, if $X$ and $Y$ are two independent random variables with infinitely divisible laws then $X + Y$ is infinite divisible as well. Indeed, given $n \in \mathbb{N}$, we may find $\mu_n, \nu_n \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_n^{*n} = \mu$ and $\nu_n^{*n} = \nu$. It then follows that $[\mu_n * \nu_n]^{*n} = \mu * \nu$.

**Theorem 25.3.** *Suppose that $\mu$ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and* $\mathrm{Law}(X) = \mu$. *Then $\mu$ is infinitely divisible iff there exits an array, $\{X_{n,k} : 1 \le k \le m_n\}$ with $\{X_{n,k}\}_{k=1}^{m_n}$ being i.i.d. such that $\sum_{k=1}^{m_n} X_{n,k} \implies X$ and $m_n \uparrow \infty$ as $n \to \infty$.*

**Proof.** The only non-trivial direction is $(\Longleftarrow)$. I will only prove the special case where $m_n = n$. [See Kallenberg [26, Lemma 15.13, p. 294] for the needed result involving the tail bounds needed to cover the full case. ]

Given any $k \in \mathbb{N}$ we decompose $S_{nk}$ into $k$ independent summands, $S_{nk} = \sum_{i=1}^k S_n^i$, where

$$S_n^i = \sum_{j=k(i-1)+1}^{ki} X_{n,j}.$$

Notice that $\{S_n^i\}_{i=1}^k$ are i.i.d. for each $n \in \mathbb{N}$ and since $S_{nk} \implies X$ as $n \to \infty$ we know that $\{S_{nk}\}_{n=1}^{\infty}$ is tight and there exists $\varepsilon(r) \downarrow 0$ as $r \uparrow \infty$ such that

$$P(|S_{nk}| > r) \le \varepsilon(r).$$

Since

$$\begin{aligned} P\left(S_n^1 > r\right)^k &= P\left(S_n^i > r \text{ for } 1 \le i \le k\right) \\ &\le P(S_{nk} > kr) \le P(|S_{nk}| > kr) \le \varepsilon(kr) \end{aligned}$$

and similarly,

$$\begin{aligned} P\left(-S_n^1 > r\right)^k &= P\left(-S_n^i > r \text{ for } 1 \le i \le k\right) \\ &\le P(-S_{nk} > kr) \le P(|S_{nk}| > kr) \le \varepsilon(kr) \end{aligned}$$

we see that $P\left(\left|S_n^1\right| > r\right) \le 2\varepsilon(kr)^{1.k} \to 0$ as $r \uparrow \infty$ which shows that $\{S_n^1\}_{n=1}^{\infty}$ has tight distributions as well.

Thus there exists a subsequence $\{n_l\}$ such that $S_{n_l}^1 \implies Y$ as $l \to \infty$. Let $\{Y_i\}_{i=1}^k$ be i.i.d. random variables with $Y_i \overset{d}{=} Y$. Then by Exercise 23.12 it follows that

$$S_{kn_l} = \sum_{i=1}^k S_{n_l}^i \implies Y_1 + \dots Y_k$$

from which we conclude that $X \overset{d}{=} Y_1 + \dots + Y_k$. Since $k$ was arbitrary we have shown $X$ is infinitely divisible. ■

*Remark 25.4.* [See Theorem 25.25 and Corollary 25.27 below for a similar result where we no longer assume that $\{X_{n,k}\}_{k=1}^{m_n}$ are i.i.d. but we do require additional normalizations (Assumption 2 below) and size restrictions on $\mathrm{Var}(X_{n,k})$ of condition (M) of Definition 25.9 below.

The Lévy Kintchine formula of Theorem 25.7 below asserts that $\mu \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is infinitely divisible iff $\hat{\mu}(\lambda) = e^{\psi(\lambda)}$ where $\psi(\lambda)$ has the form given in Eq. (25.4) below. [A more restrictive class of distributions are the so called **stable distributions**, see Definition 25.29, Lemma 25.33, and Theorem 25.34.] Before stating Lévy Kintchine formula it is worth recording a couple of examples.

*Example 25.5 (Following Theorem 19.30).* Let $\mu \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, $\{Z_n\}_{n=1}^\infty$ be i.i.d. random variables with $Z_n \stackrel{d}{=} \mu$, $N_\alpha \stackrel{d}{=} Poi(\alpha)$, and $Y \stackrel{d}{=} \sigma N(0,1) + c$ all be chosen so that $\{Z_n\}_{n=1}^\infty \cup \{N_\alpha, Y\}$ are independent. Then $S := Y + \sum_{n \le N_\alpha} Z_n$ is infinitely divisible. Indeed we have

$$\begin{aligned} f_S(\lambda) &= \mathbb{E}\left[e^{i\lambda S}\right] = \mathbb{E}\left[e^{i\lambda Y}\right] \cdot \mathbb{E}\left[e^{i\lambda \sum_{k \le N_\alpha} Z_k}\right] \\ &= \mathbb{E}\left[e^{i\lambda Y}\right] \cdot \sum_{k=0}^\infty \mathbb{E}\left[e^{i\lambda \sum_{k \le N_\alpha} Z_k} | N_\alpha = n\right] P(N_\alpha = n) \\ &= \mathbb{E}\left[e^{i\lambda Y}\right] \cdot \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda (Z_1 + \dots + Z_n)}\right] P(N_\alpha = n) \\ &= \exp\left(-\frac{1}{2}\sigma^2\lambda^2 + i\lambda\mu\right) \sum_{n=0}^\infty e^{-\alpha}\frac{\alpha^n}{n!}\hat{\mu}(\lambda)^n \\ &= \exp\left(-\frac{1}{2}\sigma^2\lambda^2 + i\lambda\mu + \alpha(\hat{\mu}(\lambda) - 1)\right) = e^{\psi(\lambda)} \end{aligned}$$

where

$$\psi(\lambda) = -\frac{1}{2}\sigma^2\lambda^2 + ic\lambda + \int_{\mathbb{R}} \left(e^{i\lambda x} - 1\right) d\nu(x)$$

and $d\nu(x) := \alpha d\mu(x)$ is an arbitrary finite measure on $\mathbb{R}$. As $\sigma$, $c$, and $\nu$ are arbitrary it follows that $e^{t\psi(\lambda)}$ is the characteristic function of a probability measure for all $t > 0$ and in particular of $t = 1/n$.

It is interesting to note that if $X_\alpha := \sum_{n \le N_\alpha} Z_n$ and $m \in \mathbb{N}$ then the law of the sum of $m$ independent copies of $X_{\alpha/m}$ is the law of $X_\alpha$. This explicitly shows that $X_\alpha$ is infinitely divisible.

*Example 25.6 (Exercise 24.5).* Recall from Exercise 24.5, if $\nu$ is any finite measure $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, there exists a (necessarily unique) probability measure $\mu$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu} = e^\psi$ where

$$\psi(\lambda) = \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x). \tag{25.3}$$

Again $\hat{\mu}_t(\lambda) = e^{t\psi(\lambda)}$ is again of this form for all $t > 0$ and therefore $\mu$ is infinite divisible. If we let $d\tilde{\nu}(x) := \frac{1}{x^2} d\nu(x)$, we may rewrite Eq. (25.3) as

$$\psi(\lambda) = i\lambda b + \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x 1_{|x| \le 1}\right) d\tilde{\nu}(x)$$

where

$$b = -\int_{\mathbb{R}} x 1_{|x|>1} d\tilde{\nu}(x) = -\int_{|x|>1} \frac{1}{|x|} d\nu(x) \in \mathbb{R}$$

and $\tilde{\nu}$ is a positive measure (perhaps infinite measure) such that

$$\int_{\mathbb{R}} x^2 d\tilde{\nu}(x) < \infty.$$

Keeping these two examples in mind should make the following important theorem plausible.

**Theorem 25.7 (Lévy Kintchine formula).** *A probability measure $\mu$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is infinitely divisible iff $\hat{\mu}(\lambda) = e^{\psi(\lambda)}$ where*

$$\psi(\lambda) = i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x| \le 1}\right) d\nu(x) \tag{25.4}$$

*for some $b \in \mathbb{R}$, $a \ge 0$, and some measure $\nu$ on $\mathbb{R} \setminus \{0\}$ such that*

$$\int_{\mathbb{R}\setminus\{0\}} \left(x^2 \wedge 1\right) d\nu(x) < \infty. \tag{25.5}$$

*[Note that the term $-\frac{1}{2}a\lambda^2$ in Eq. (25.4) the logarithm of the characteristic function of $N(0, \sqrt{a})$.]*

**Proof.** We will give the easy direction of this proof, namely the implication $(\Longleftarrow)$. To summarize we want to show if $\psi$ is of the form in Eq. (25.4), then there exists a unique probability measure $\mu$ such that $\hat{\mu} = e^\psi$. As $a$, $b$, and $\nu$ are arbitrary it follows, for $t > 0$, that $t\psi$ is still of the form Eq. (25.4) and therefore $\mu$ is infinite divisible.

If the measure $\nu$ appearing in is a finite measure then

$$\psi(\lambda) = i\lambda b' - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}} \left(e^{i\lambda x} - 1\right) d\nu(x)$$

where

$$b' = b - \int_{|x| \le 1} x d\nu(x).$$

Thus we may use Example 25.5 in order to construct a random variable with distribution given by $\mu$.

For general $\nu$ satisfying Eq. (25.5) let, for $\varepsilon > 0$, $d\nu_\varepsilon(x) := 1_{|x| \ge \varepsilon} d\nu(x)$ – a finite measure on $\mathbb{R}$. Thus by Example 25.5 there exits a probability measure, $\mu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu}_\varepsilon = e^{\psi_\varepsilon}$ where

$$\psi_\varepsilon(\lambda) := i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\le 1}\right) 1_{|x|\ge\varepsilon} d\nu(x).$$

As $e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\le 1}$ is bounded and less than $C(\lambda)x^2$ for $|x| \le 1$, we may use DCT to show $\psi_\varepsilon(\lambda) \to \psi(\lambda)$. Furthermore the DCT also shows $\psi(\lambda)$ is continuous and therefore $e^{\psi(\lambda)}$ is continuous. Thus we have shown that $\mu_\varepsilon \to e^\psi$ where the limit is continuous and therefore by the continuity Theorem 24.20, there exists a probability measure $\mu$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\hat{\mu} = e^\psi$.

The proof of the other implication ( $\Longrightarrow$ ) will be discussed in Appendix 25.4 below. For more information about Poisson processes and Lévy processes; see Protter [36, Chapter I], [6, Chapter 9.5], and [15, Chapter XVII.2, p. 558-] for analytic proofs. Also see http://www.math.uconn.edu/~bass/scdp.pdf, Kallenberg [26, Theorem 15.13, p. 294], and [1]. ■

We are now going to drop the assumption that the $\{X_{n,k}\}_{k=1}^n$ are identically distributed but we will add some normalization conditions (see Assumption 2 below) and also impose some additional conditions on the $X_{n,k}$ to assure that each term is "small" and roughly "comparable" in size. The point is that we do not want any one of the $\{X_{n,k}\}_{k=1}^n$ to dominate the sum $S_n$ in Eq. (25.1) in the limit as $n \to \infty$.

**Assumption 2** *Assume* $\mathbb{E}[X_{n,k}] = 0$, $\sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] < \infty$, *and* $\mathrm{Var}(S_n) = \sum_{k=1}^n \sigma_{n,k}^2 = 1$.

*Example 25.8.* Suppose $\{X_n\}_{n=1}^\infty$ are mean zero square integrable random variables with $\sigma_k^2 = \mathrm{Var}(X_k)$. If we let $s_n^2 := \sum_{k=1}^n \mathrm{Var}(X_k) = \sum_{k=1}^n \sigma_k^2$, $\sigma_{n,k}^2 := \sigma_k^2/s_n^2$, and $X_{n,k} := X_k/s_n$, then $\{X_{n,k}\}_{k=1}^n$ satisfy the above hypothesis and $S_n = \frac{1}{s_n}\sum_{k=1}^n X_k$.

The next definition records some possible meanings of small and comparable in size.

**Definition 25.9.** *Let* $\{X_{n,k}\}$ *be as above.*

1. $\{X_{n,k}\}$ *satisfies the* ***Lindeberg Condition*** $(LC)$ *iff*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] = 0 \text{ for all } t > 0. \tag{25.6}$$

*[Since* $\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right]$ *is a decreasing function of* $t$ *it suffices to check* $(LC)$ *along any sequence of* $\{t_n\}$ *with* $t_n \downarrow 0$.*]*

2. $\{X_{n,k}\}$ *satisfies* ***condition*** $(M)$ *if*

$$D_n := \max\left\{\sigma_{n,k}^2 : k \le n\right\} \to 0 \text{ as } n \to \infty. \tag{25.7}$$

3. $\{X_{n,k}\}$ *is* ***uniformly asymptotic negligibility (UAN)*** *if for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \max_{k\le n} P(|X_{n,k}| > \varepsilon) = 0. \tag{25.8}$$

Each of these conditions imposes constraints on the size of the tails of the $\{X_{n,k}\}$, see Lemma 25.13 below where it is shown $(LC) \implies (M) \implies (UAN)$. Condition $(M)$ asserts that all of the terms in the sum $\sum_{k=1}^n \sigma_{n,k}^2 = \mathrm{Var}(S_n) = 1$ are small so that no one term is contributing by itself.

*Remark 25.10.* The reader should observe that in order for condition $(M)$ to hold in the setup in Example 25.8 it is necessary that $\lim_{n\to\infty} s_n^2 = \infty$.

*Example 25.11.* Suppose $\{X_n\}_{n=1}^\infty$ are i.i.d. with $\mathbb{E}X_n = 0$ and $\mathrm{Var}(X_n) = \sigma^2$. Then $\left\{X_{n,k} := \frac{1}{\sqrt{n}\sigma}X_k\right\}_{k=1}^n$ satisfy $(LC)$. Indeed,

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] &= \frac{1}{n\sigma^2}\sum_{k=1}^n \mathbb{E}\left[X_k^2 : \left|\frac{X_k}{\sqrt{n}\sigma}\right| > t\right] \\ &= \frac{1}{\sigma^2}\mathbb{E}\left[X_1^2 : |X_1| > \sqrt{n}\sigma t\right]\end{aligned}$$

which tends to zero as $n \to \infty$ by DCT.

The last question we would like to address is when is the sequence $\{S_n\}_{n=1}^\infty$ **asymptotically normal,** i.e. under what conditions does it happen that $S_n \implies N(0,1)$. Under the normalization Assumption 2, the asymptotic normality results (proved in the next section) may be summarized as follows.

1. The Lindeberg-Feller CLT Theorem 25.15 asserts that if $(LC)$ holds then $S_n \implies N(0,1)$, i.e. $\{S_n\}_{n=1}^\infty$ is **asymptotically normal.**
2. Conversely, if the weaker condition $(M)$ holds and $S_n \implies N(0,1)$, then $(LC)$ holds, see Theorem 25.20.

## 25.1 Lindeberg-Feller CLT

In this section we will assume that Assumption 2 is in place, i.e. that

$$\begin{aligned}&\mathbb{E}[X_{n,k}] = 0,\ \sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] < \infty, \text{ and} \\ &\mathrm{Var}(S_n) = \sum_{k=1}^n \sigma_{n,k}^2 = 1.\end{aligned}$$

**Lemma 25.12.** *Let $\{X_{n,k}\}_{k=1}^n$ for $n \in \mathbb{N}$ be as in Assumption 2. If $\{X_{n,k}\}_{k=1}^n$ satisfy the* ***Liapunov condition*** *(LiapC);*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left|X_{n,k}\right|^\alpha = 0 \text{ for some } \alpha > 2 \tag{25.9}$$

*then (LC) holds. More generally, if $\{X_{n,k}\}$ satisfies the Liapunov condition,*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \varphi\left(|X_{n,k}|\right)\right] = 0$$

*where $\varphi : [0,\infty) \to [0,\infty)$ is a non-decreasing function such that $\varphi(t) > 0$ for all $t > 0$, then $\{X_{n,k}\}$ satisfies (LC).*

**Proof.** Assuming Eq. (25.9), then for any $t > 0$,

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] &\le \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \left|\frac{X_{n,k}}{t}\right|^{\alpha-2} : |X_{n,k}| > t\right] \\ &\le \frac{1}{t^{\alpha-2}} \sum_{k=1}^n \mathbb{E}\left[|X_{n,k}|^\alpha\right] \\ &= \frac{1}{t^{\alpha-2}} \sum_{k=1}^n \mathbb{E}\left|X_{n,k}\right|^\alpha \to 0 \text{ as } n \to \infty.\end{aligned}$$

The generalization is proved similarly;

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] &\le \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \frac{\varphi(|X_{n,k}|)}{\varphi(t)} : |X_{n,k}| > t\right] \\ &\le \frac{1}{\varphi(t)} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \varphi(|X_{n,k}|)\right] \to 0 \text{ as } n \to \infty.\end{aligned}$$

■

**Lemma 25.13.** *Let $\{X_{n,k} : 1 \le k \le n < \infty\}$ be as above, then (LiapC) $\implies$ (LC) $\implies$ (M) $\implies$ (UAN). Moreover the Lindeberg Condition (LC) implies the following strong form of (UAN),*

$$\sum_{k=1}^n P\left(|X_{n,k}| > \varepsilon\right) \le \frac{1}{\varepsilon^2} \sum_{k=1}^n \mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| > \varepsilon\right] \to 0. \tag{25.10}$$

**Proof.** The assertion, (LiapC) $\implies$ (LC), was proved in Lemma 25.12. For $k \le n$,

$$\begin{aligned}\sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] &= \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|\le t}\right] + \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|>t}\right] \\ &\le t^2 + \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|>t}\right] \le t^2 + \sum_{m=1}^n \mathbb{E}\left[X_{n,m}^2 1_{|X_{n,m}|>t}\right]\end{aligned}$$

and therefore using (LC) we find

$$\lim_{n\to\infty} D_n := \lim_{n\to\infty} \max_{k\le n} \sigma_{n,k}^2 \le t^2 \text{ for all } t > 0.$$

This clearly implies (M) holds. For $\varepsilon > 0$ we have by Chebyschev's inequality that

$$P\left(|X_{n,k}| > \varepsilon\right) \le \frac{1}{\varepsilon^2} \mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| > \varepsilon\right] \le \frac{1}{\varepsilon^2} \sigma_{n,k}^2 \tag{25.11}$$

and therefore,

$$\max_{k\le n} P\left(|X_{n,k}| > \varepsilon\right) \le \frac{1}{\varepsilon^2} \max_{k\le n} \sigma_{n,k}^2 = \frac{1}{\varepsilon^2} D_n \to 0 \text{ as } n \to \infty$$

which shows (M) $\implies$ (UAN). Summing Eq. (25.11) on $k$ gives Eq. (25.10) and the right member of this equation tends to zero as $n \to \infty$ if (LC) holds. ■

We will need the following lemma for our subsequent applications of the continuity theorem.

**Lemma 25.14.** *Suppose that $a_i, b_i \in \mathbb{C}$ with $|a_i|, |b_i| \le 1$ for $i = 1, 2, \dots, n$. Then*

$$\left|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\right| \le \sum_{i=1}^n |a_i - b_i|.$$

**Proof.** Let $a := \prod_{i=1}^{n-1} a_i$ and $b := \prod_{i=1}^{n-1} b_i$ and observe that $|a|, |b| \le 1$ and that

$$\begin{aligned}|a_n a - b_n b| &\le |a_n a - a_n b| + |a_n b - b_n b| \\ &= |a_n|\,|a - b| + |a_n - b_n|\,|b| \\ &\le |a - b| + |a_n - b_n|.\end{aligned}$$

The proof is now easily completed by induction on $n$. ■

**Theorem 25.15 (Lindeberg-Feller CLT (I)).** *Suppose $\{X_{n,k}\}$ satisfies (LC) and the hypothesis in Assumption 2, then*

$$S_n \implies N(0,1). \tag{25.12}$$

*(See Theorem 25.20 for a converse to this theorem.)*

To prove this theorem we must show

$$\mathbb{E}\left[e^{i\lambda S_n}\right] \to e^{-\lambda^2/2} \text{ as } n \to \infty. \tag{25.13}$$

Before starting the formal proof, let me give an informal explanation for Eq. (25.13). Using

$$f_{nk}(\lambda) \sim 1 - \frac{\lambda^2}{2}\sigma_{nk}^2,$$

we might expect

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda S_n}\right] &= \prod_{k=1}^{n} f_{nk}(\lambda) = e^{\sum_{k=1}^{n} \ln f_{nk}(\lambda)} \\ &= e^{\sum_{k=1}^{n} \ln(1+f_{nk}(\lambda)-1)} \\ &\overset{(A)}{\sim} e^{\sum_{k=1}^{n}(f_{nk}(\lambda)-1)} \left(= \prod_{k=1}^{n} e^{(f_{nk}(\lambda)-1)}\right) \\ &\overset{(B)}{\sim} e^{\sum_{k=1}^{n} -\frac{\lambda^2}{2}\sigma_{nk}^2} = e^{-\frac{\lambda^2}{2}}.\end{aligned}$$

The question then becomes under what conditions are these approximations valid. It turns out that approximation $(A)$, namely that

$$\lim_{n\to\infty}\left|\prod_{k=1}^{n} f_{nk}(\lambda) - \exp\left(\sum_{k=1}^{n}(f_{nk}(\lambda)-1)\right)\right| = 0, \tag{25.14}$$

is valid if condition $(M)$ holds, see Lemma 25.18 below and the approximation $(B)$ is valid, i.e.

$$\lim_{n\to\infty}\sum_{k=1}^{n}(f_{nk}(\lambda)-1) = -\frac{1}{2}\lambda^2,$$

if $(LC)$ is satisfied, see Lemma 25.16. These observations would then constitute a proof of Theorem 25.15. The proof we give below of Theorem 25.15 will not quite follow this route and will not use Lemma 25.18 directly. However, this lemma will be used in the proofs of Theorems 25.20 and 25.25.

**Proof.** (Proof of Theorem 25.15) Since

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n} f_{nk}(\lambda) \text{ and } e^{-\lambda^2/2} = \prod_{k=1}^{n} e^{-\lambda^2\sigma_{n,k}^2/2},$$

we may use Lemma 25.14 to conclude,

$$\left|\mathbb{E}\left[e^{i\lambda S_n}\right] - e^{-\lambda^2/2}\right| \le \sum_{k=1}^{n}\left|f_{nk}(\lambda) - e^{-\lambda^2\sigma_{n,k}^2/2}\right| = \sum_{k=1}^{n}(A_{n,k}+B_{n,k})$$

where

$$A_{n,k} := \left|f_{nk}(\lambda) - 1 + \frac{\lambda^2\sigma_{n,k}^2}{2}\right| \text{ and} \tag{25.15}$$

$$B_{n,k} := \left|\left[1 - \frac{\lambda^2\sigma_{n,k}^2}{2}\right] - e^{-\lambda^2\sigma_{n,k}^2/2}\right|. \tag{25.16}$$

Because of Lemma 25.16 below, to finish the proof it suffices to show $\lim_{n\to\infty}\sum_{k=1}^{n} B_{n,k} = 0$. To estimate $\sum_{k=1}^{n} B_{n,k}$, we use the estimate, $|e^{-u} - 1 + u| \le u^2/2$ valid for $u \ge 0$ (see Eq. (24.54) with $z = -u$). With this estimate we find,

$$\begin{aligned}\sum_{k=1}^{n} B_{n,k} &= \sum_{k=1}^{n}\left|\left[1 - \frac{\lambda^2\sigma_{n,k}^2}{2}\right] - e^{-\lambda^2\sigma_{n,k}^2/2}\right| \\ &\le \sum_{k=1}^{n}\frac{1}{2}\left[\frac{\lambda^2\sigma_{n,k}^2}{2}\right]^2 = \frac{\lambda^4}{8}\sum_{k=1}^{n}\sigma_{n,k}^4 \\ &\le \frac{\lambda^4}{8}\max_{k\le n}\sigma_{n,k}^2\sum_{k=1}^{n}\sigma_{n,k}^2 = \frac{\lambda^4}{8}\max_{k\le n}\sigma_{n,k}^2 \to 0,\end{aligned}$$

wherein we have used $(M)$ (which is implied by $(LC)$) in taking the limit as $n \to \infty$. ∎

**Lemma 25.16.** *Let $A_{n,k}$ be as in Eq. (25.15). If $\{X_{n,k}\}_{k=1}^{n}$ satisfies $(LC)$, then*

$$\limsup_{n\to\infty}\sum_{k=1}^{n} A_{n,k} = 0 \textit{ and}$$

$$\lim_{n\to\infty}\sum_{k=1}^{n}(f_{nk}(\lambda)-1) = -\lambda^2/2 \textit{ for all } \lambda \in \mathbb{R}.$$

**Proof.** Rewriting $A_{n,k}$ using $\mathbb{E}X_{n,k} = 0$ and then using Lemma 24.50 implies for every $\varepsilon > 0$ that,

$$\begin{aligned} A_{n,k} &= \left|\mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k} + \frac{\lambda^2}{2}X_{n,k}^2\right]\right| \\ &\le \mathbb{E}\left|e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k} + \frac{\lambda^2}{2}X_{n,k}^2\right| \\ &\le \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\left|X_{n,k}\right|^3}{3!}\right] \\ &\le \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\left|X_{n,k}\right|^3}{3!} : |X_{n,k}| \le \varepsilon\right] \\ &\qquad + \lambda^2\mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda|\left|X_{n,k}\right|^3}{3!} : |X_{n,k}| > \varepsilon\right] \\ &\le \frac{|\lambda|^3}{3!}\varepsilon \cdot \mathbb{E}\left[\left|X_{n,k}\right|^2 : |X_{n,k}| \le \varepsilon\right] + \lambda^2\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right] \\ &= \frac{|\lambda|^3\varepsilon}{6}\sigma_{n,k}^2 + \lambda^2\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right]. \end{aligned}$$

Summing this equation on $k$ and making use of $(LC)$ gives;

$$\limsup_{n\to\infty}\sum_{k=1}^n A_{n,k} \le \frac{\lambda^3\varepsilon}{6} \to 0 \text{ as } \varepsilon \downarrow 0. \tag{25.17}$$

The second limit follows from the first and the simple estimate;

$$\left|\sum_{k=1}^n (f_{nk}(\lambda) - 1) + \lambda^2/2\right| = \left|\sum_{k=1}^n \left(f_{nk}(\lambda) - 1 + \frac{\lambda^2\sigma_{n,k}^2}{2}\right)\right| \le \sum_{k=1}^n A_{n,k}.$$

■

As an application of Theorem 25.15 we can give half of the proof of Theorem 22.42.

**Theorem 25.17 (Converse assertion in Theorem 22.42).** *If $\{X_n\}_{n=1}^\infty$ are independent random variables and the random series, $\sum_{n=1}^\infty X_n$, is almost surely convergent, then for all $c > 0$ the following three series converge;*

1. $\sum_{n=1}^\infty P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^\infty \mathrm{Var}\left(X_n 1_{|X_n|\le c}\right) < \infty$, *and*
3. $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ *converges.*

**Proof.** Since $\sum_{n=1}^\infty X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and hence for every $c > 0$, $P(\{|X_n| \ge c \text{ i.o.}\}) = 0$. According the Borel zero one law (Lemma 12.44) this implies for every $c > 0$ that $\sum_{n=1}^\infty P(|X_n| > c) < \infty$. Since $X_n \to 0$ a.s., $\{X_n\}$ and $\left\{X_n^c := X_n 1_{|X_n|\le c}\right\}$ are tail equivalent for all $c > 0$. In particular $\sum_{n=1}^\infty X_n^c$ is almost surely convergent for all $c > 0$.

Fix $c > 0$, let $Y_n := X_n^c - \mathbb{E}[X_n^c]$ and let

$$s_n^2 = \mathrm{Var}(Y_1 + \dots + Y_n) = \sum_{k=1}^n \mathrm{Var}(Y_k) = \sum_{k=1}^n \mathrm{Var}(X_k^c) = \sum_{k=1}^n \mathrm{Var}\left(X_k 1_{|X_k|\le c}\right).$$

For the sake of contradictions, suppose $s_n^2 \to \infty$ as $n \to \infty$. Since $|Y_k| \le 2c$, it follows that $\sum_{k=1}^n \mathbb{E}\left[Y_k^2 1_{|Y_k|>s_n t}\right] = 0$ for all sufficiently large $n$ and hence

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[Y_k^2 1_{|Y_k|>s_n t}\right] = 0,$$

i.e. $\{Y_{n,k} := Y_k/s_n\}_{n=1}^\infty$ satisfies $(LC)$ – see Examples 25.8 and Remark 25.10. So by the central limit Theorem 25.15, it follows that

$$\frac{1}{s_n^2}\sum_{k=1}^n (X_n^c - \mathbb{E}[X_n^c]) = \frac{1}{s_n^2}\sum_{k=1}^n Y_k \implies N(0,1).$$

On the other hand we know

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^n X_n^c = \frac{\sum_{k=1}^\infty X_k^c}{\lim_{n\to\infty} s_n^2} = 0 \text{ a.s.}$$

and so by Slutsky's theorem,

$$\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[X_n^c] = \frac{1}{s_n^2}\sum_{k=1}^n X_n^c - \frac{1}{s_n^2}\sum_{k=1}^n Y_k \implies N(0,1).$$

But it is not possible for constant (i.e. **non-random)** variables, $c_n := \frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[X_n^c]$, to converge to a non-degenerate limit. (Think about this either in terms of characteristic functions or in terms of distribution functions.) Thus we must conclude that

$$\sum_{n=1}^\infty \mathrm{Var}\left(X_n 1_{|X_n|\le c}\right) = \sum_{n=1}^\infty \mathrm{Var}(X_n^c) = \lim_{n\to\infty} s_n^2 < \infty.$$

An application of Kolmogorov's convergence criteria (Theorem 22.11) implies that

$$\sum_{n=1}^\infty (X_n^c - \mathbb{E}[X_n^c]) \text{ is convergent } a.s.$$

Since we already know that $\sum_{n=1}^\infty X_n^c$ is convergent almost surely we may now conclude $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ is convergent. ■

Let us now turn to the converse of Theorem 25.15, see Theorem 25.20 below.

**Lemma 25.18.** *Suppose that $\{X_{n,k}\}$ satisfies property $(M)$, i.e. $D_n := \max_{k\le n}\sigma_{n,k}^2 \to 0$. If we define,*

$$\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1 = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right],$$

*then;*

1. $\lim_{n\to\infty}\max_{k\le n}|\varphi_{n,k}(\lambda)| = 0$ *and*
2. $f_{S_n}(\lambda) - \prod_{k=1}^n e^{\varphi_{n,k}(\lambda)} \to 0$ *as* $n\to\infty$, *where*

$$f_{S_n}(\lambda) = \mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^n f_{n,k}(\lambda).$$

**Proof.** For any $\varepsilon > 0$ we have, making use of Eq. (24.61) and Chebyschev's inequality, that

$$\begin{aligned}|\varphi_{n,k}(\lambda)| &= |f_{n,k}(\lambda) - 1| \le \mathbb{E}\left|e^{i\lambda X_{n,k}} - 1\right| \le \mathbb{E}\left[2\wedge|\lambda X_{n,k}|\right]\\ &\le \mathbb{E}\left[2\wedge|\lambda X_{n,k}| : |X_{n,k}|\ge\varepsilon\right] + \mathbb{E}\left[2\wedge|\lambda X_{n,k}| : |X_{n,k}|<\varepsilon\right]\\ &\le 2P\left[|X_{n,k}|\ge\varepsilon\right] + |\lambda|\varepsilon \le \frac{2\sigma_{n,k}^2}{\varepsilon^2} + |\lambda|\varepsilon.\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty}\max_{k\le n}|\varphi_{n,k}(\lambda)| \le \limsup_{n\to\infty}\left[\frac{2D_n}{\varepsilon^2} + |\lambda|\varepsilon\right] = |\lambda|\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

For the second item, observe that $\operatorname{Re}\varphi_{n,k}(\lambda) = \operatorname{Re} f_{n,k}(\lambda) - 1 \le 0$ and hence $\left|e^{\varphi_{n,k}(\lambda)}\right| = e^{\operatorname{Re}\varphi_{n,k}(\lambda)} \le 1$. Therefore by Lemma 25.14 and the estimate (24.54) we find;

$$\begin{aligned}\left|\prod_{k=1}^n e^{\varphi_{n,k}(\lambda)} - \prod_{k=1}^n f_{n,k}(\lambda)\right| &\le \sum_{k=1}^n\left|e^{\varphi_{n,k}(\lambda)} - f_{n,k}(\lambda)\right|\\ &= \sum_{k=1}^n\left|e^{\varphi_{n,k}(\lambda)} - (1+\varphi_{n,k}(\lambda))\right|\\ &\le \frac{1}{2}\sum_{k=1}^n|\varphi_{n,k}(\lambda)|^2\\ &\le \frac{1}{2}\max_{k\le n}|\varphi_{n,k}(\lambda)|\cdot\sum_{k=1}^n|\varphi_{n,k}(\lambda)|.\end{aligned}$$

Since $\mathbb{E}X_{n,k} = 0$ we may write express $\varphi_{n,k}$ as

$$\varphi_{n,k}(\lambda) = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right]$$

and then using estimate in Eq. (24.54) again shows

$$\begin{aligned}\sum_{k=1}^n|\varphi_{n,k}(\lambda)| &= \sum_{k=1}^n\left|\mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right]\right|\\ &\le \sum_{k=1}^n\left|\mathbb{E}\left[\frac{1}{2}|\lambda X_{n,k}|^2\right]\right| \le \frac{\lambda^2}{2}\sum_{k=1}^n\sigma_{n,k}^2 = \frac{\lambda^2}{2}.\end{aligned}$$

Thus we have shown,

$$\left|\prod_{k=1}^n f_{n,k}(\lambda) - \prod_{k=1}^n e^{\varphi_{n,k}(\lambda)}\right| \le \frac{\lambda^2}{4}\max_{k\le n}|\varphi_{n,k}(\lambda)|$$

and the latter expression tends to zero by item 1. ■

**Lemma 25.19.** *Let $X$ be a random variable such that $\mathbb{E}X^2<\infty$ and $\mathbb{E}X = 0$. Further let $f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ and $u(\lambda) := \operatorname{Re}(f(\lambda)-1)$. Then for all $c>0$,*

$$u(\lambda) + \frac{\lambda^2}{2}\mathbb{E}\left[X^2\right] \ge \mathbb{E}\left[X^2\left[\frac{\lambda^2}{2} - \frac{2}{c^2}\right] : |X|>c\right] \tag{25.18}$$

*or equivalently*

$$\mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \ge \mathbb{E}\left[X^2\left[\frac{\lambda^2}{2} - \frac{2}{c^2}\right] : |X|>c\right]. \tag{25.19}$$

*In particular if we choose $|\lambda| \ge \sqrt{6}/|c|$, then*

$$\mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \ge \frac{1}{c^2}\mathbb{E}\left[X^2 : |X|>c\right]. \tag{25.20}$$

**Proof.** For all $\lambda\in\mathbb{R}$, we have (see Eq. (24.52)) $\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2 \ge 0$ and $\cos\lambda X - 1 \ge -2$. Therefore,

$$\begin{aligned}u(\lambda) + \frac{\lambda^2}{2}\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right]\\ &\ge \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2 : |X|>c\right]\\ &\ge \mathbb{E}\left[-2 + \frac{\lambda^2}{2}X^2 : |X|>c\right]\\ &\ge \mathbb{E}\left[-2\frac{|X|^2}{c^2} + \frac{\lambda^2}{2}X^2 : |X|>c\right]\end{aligned}$$

which gives Eq. (25.18). ■

**Theorem 25.20 (Lindeberg-Feller CLT (II)).** *Suppose $\{X_{n,k}\}$ satisfies $(M)$ and $S_n \implies N(0,1)$ (i.e. the central limit theorem in Eq. (25.12) holds), then $\{X_{n,k}\}$ satisfies $(LC)$. So under condition $(M)$, $S_n$ converges to a normal random variable iff $(LC)$ holds.*

**Proof.** By assumption we have

$$\lim_{n\to\infty}\max_{k\le n}\sigma_{n,k}^2 = 0 \text{ and } \lim_{n\to\infty}\prod_{k=1}^n f_{n,k}(\lambda) = e^{-\lambda^2/2}.$$

The second inequality combined with Lemma 25.18 implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^n \varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\prod_{k=1}^n e^{\varphi_{n,k}(\lambda)} = e^{-\lambda^2/2}.$$

Taking the modulus of this equation then implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^n \operatorname{Re}\varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\left|e^{\sum_{k=1}^n \varphi_{n,k}(\lambda)}\right| = e^{-\lambda^2/2}$$

from which we may conclude

$$\lim_{n\to\infty}\sum_{k=1}^n \operatorname{Re}\varphi_{n,k}(\lambda) = -\lambda^2/2.$$

We may write this last limit as

$$\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\left[\cos(\lambda X_{n,k}) - 1 + \frac{\lambda^2}{2}X_{n,k}^2\right] = 0$$

which by Lemma 25.19 implies

$$\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > c\right] = 0$$

for all $c > 0$ which is $(LC)$. ■

As an application of Theorem 25.15 let us see what it has to say about Brownian motion. In what follows we say that $\{B_t\}_{t\ge0}$ is a Gaussian process if for all finite subsets, $\Lambda \subset [0,\infty)$ the random variables $\{B_t\}_{t\in\Lambda}$ are **jointly** Gaussian. We will discuss Gaussian processes in more generality in Chapter 26.

**Proposition 25.21.** *Suppose that $\{B_t\}_{t\ge0}$ is a stochastic process on some probability space, $(\Omega,\mathcal{B},P)$ such that;*

1. *$B_0 = 0$ a.s., $\mathbb{E}B_t = 0$ for all $t \ge 0$,*
2. *$\mathbb{E}(B_t - B_s)^2 = t - s$ for all $0 \le s \le t < \infty$,*
3. *$B$ has independent increments, i.e. if $0 = t_0 < t_1 < \cdots < t_n < \infty$, then $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^n$ are independent random variables.*
4. *(Moment Condition) There exists $p > 2$, $q > 1$ and $c < \infty$ such that $\mathbb{E}|B_t - B_s|^p \le c|t-s|^q$ for all $s,t \in \mathbb{R}_+$.*

*Then $B_t - B_s \overset{d}{=} N(0, t-s)$ for all $0 \le s < t < \infty$. We call such a process satisfying these conditions a* ***pre-Brownian motion.***

**Proof.** Let $0 \le s < t$ and for each $n \in \mathbb{N}$ and $1 \le k \le n$ let $X_{n,k} := B_{t_k} - B_{t_{k-1}}$ where $\{s = t_0 < t_1 < \cdots < t_n = t\}$ is the uniform partition of $[s,t]$. Under the moment condition hypothesis we find,

$$\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^p\right] \le c\sum_{k=1}^n\left(\frac{t-s}{n}\right)^q = c(t-s)^q\frac{n}{n^q} \to 0 \text{ as } n\to\infty.$$

Thus we have shown that $\{X_{n,k}\}$ satisfies a Liapunov condition which by Lemma 25.12 implies that $\{X_{n,k}\}$ satisfies $(LC)$. Therefore, $B_t - B_s = \sum_{k=1}^n X_{n,k} \to N(0,t-s)$ as $n\to\infty$ by the Lindeberg-Feller central limit Theorem 25.15. ■

*Remark 25.22 (Poisson Process).* There certainly are other processes satisfying items 1.-3. other than a pre-Brownian motion. Indeed, if $\{N_t\}_{t\ge0}$ is a Poisson process with intensity $\lambda$ (see Example 27.9), then $B_t := \lambda^{-1}N_t - t$ satisfies the items 1.–3. above. Recall that $\operatorname{Var}(N_t - N_s) = (\lambda t - \lambda s)^2$ and $\mathbb{E}(N_t - N_s) = \lambda(t-s)$ so $\mathbb{E}B_t = 0$ and

$$\begin{aligned}\mathbb{E}(B_t - B_s)^2 &= \operatorname{Var}(B_t - B_s) = \operatorname{Var}\left(\lambda^{-1}(N_t - N_s)\right)\\ &= \lambda^{-2}\operatorname{Var}(N_t - N_s) = \lambda^{-2}(\lambda t - \lambda s)^2 = t - s.\end{aligned}$$

In this case one can show that $\mathbb{E}\left[|B_t - B_s|^p\right] \sim |t-s|$ for all $1 \le p < \infty$.

This last remark leads us to our next topic.

## 25.2 More on Infinitely Divisible Distributions

In the this section we are going to investigate the possible limiting distributions of the $\{S_n\}_{n=1}^\infty$ when we relax the Lindeberg condition. Let us begin with a simple example of the Poisson limit theorem.

**Theorem 25.23 (A Poisson Limit Theorem).** *For each $n \in \mathbb{N}$, let $\{Y_{n,k}\}_{k=1}^n$ be independent Bernoulli random variables with $P(Y_{n,k} = 1) = p_{n,k}$ and $P(Y_{n,k} = 0) = q_{n,k} := 1 - p_{n,k}$. Suppose;*

*1.* $\lim_{n\to\infty}\sum_{k=1}^{n} p_{n,k} = a \in (0,\infty)$ *and*
*2.* $\lim_{n\to\infty}\max_{1\le k\le n} p_{n,k} = 0$. *(So no one term is dominating the sums in item 1.)*

*Then* $S_n = \sum_{k=1}^{n} Y_{n,k} \implies Z$ *where* $Z$ *is a Poisson random variable with mean* $a$. *(See [12, Section 2.6] for more on this theorem.)*

**Proof.** We will give two proofs of this theorem. The first proof relies on the law of rare events in Theorem 23.10 while the second uses Fourier transform methods.

**First proof.** Let $Z_n \overset{d}{=} Poi\left(\sum_{k=1}^{n} p_{n,k}\right)$, then by Theorem 23.10, we know that

$$d_{TV}(Z_n, S_n) \le \sum_{k=1}^{n} p_{n,k}^2 \le \max_{1\le k\le n} p_{n,k} \cdot \sum_{k=1}^{n} p_{n,k}.$$

From the assumptions it follows that $\lim_{n\to\infty} d_{TV}(Z_n, S_n) = 0$ and from part 3. of Exercise 23.6 we know that $\lim_{n\to\infty} d_{TV}(Z_n, Z) = 0$. Therefore, $\lim_{n\to\infty} d_{TV}(Z, S_n) = 0$.

**Second proof.** Recall from Example 24.11 that for any $a > 0$,

$$\mathbb{E}\left[e^{i\lambda Z}\right] = \exp\left(a\left(e^{i\lambda} - 1\right)\right).$$

Since

$$\mathbb{E}\left[e^{i\lambda Y_{n,k}}\right] = e^{i\lambda} p_{n,k} + (1 - p_{n,k}) = 1 + p_{n,k}\left(e^{i\lambda} - 1\right),$$

it follows that

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n}\left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right].$$

Since $1 + p_{n,k}\left(e^{i\lambda} - 1\right)$ lies on the line segment joining 1 to $e^{i\lambda}$, it follows (see Figure 25.1) that

$$\left|1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right| \le 1.$$

Hence we may apply Lemma 25.14 to find

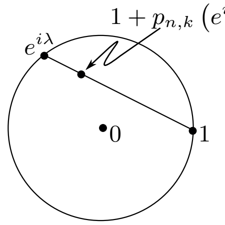



**Fig. 25.1.** Simple circle geometry reflecting the convexity of the disk.

$$\begin{aligned}
&\left|\prod_{k=1}^{n} \exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) - \prod_{k=1}^{n}\left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right]\right| \\
&\le \sum_{k=1}^{n}\left|\exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) - \left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right]\right| \\
&= \sum_{k=1}^{n}\left|\exp(z_{n,k}) - [1 + z_{n,k}]\right|
\end{aligned}$$

where

$$z_{n,k} = p_{n,k}\left(e^{i\lambda} - 1\right).$$

Since $\operatorname{Re} z_{n,k} = p_{n,k}(\cos\lambda - 1) \le 0$, we may use the calculus estimate in Eq. (24.54) to conclude,

$$\begin{aligned}
&\left|\prod_{k=1}^{n} \exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) - \prod_{k=1}^{n}\left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right]\right| \\
&\le \frac{1}{2}\sum_{k=1}^{n}|z_{n,k}|^2 \le \frac{1}{2}\max_{1\le k\le n}|z_{n,k}|\sum_{k=1}^{n}|z_{n,k}| \\
&\le 2\max_{1\le k\le n} p_{n,k}\sum_{k=1}^{n} p_{n,k}.
\end{aligned}$$

Using the assumptions, we may conclude

$$\left|\prod_{k=1}^{n} \exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) - \prod_{k=1}^{n}\left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right]\right| \to 0 \text{ as } n \to \infty.$$

Since

$$\prod_{k=1}^{n} \exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) = \exp\left(\sum_{k=1}^{n} p_{n,k}\left(e^{i\lambda} - 1\right)\right) \to \exp\left(a\left(e^{i\lambda} - 1\right)\right),$$

we have shown

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{E}\left[e^{i\lambda S_n}\right] &= \lim_{n\to\infty}\prod_{k=1}^{n}\left[1 + p_{n,k}\left(e^{i\lambda} - 1\right)\right] \\
&= \lim_{n\to\infty}\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda} - 1\right)\right) = \exp\left(a\left(e^{i\lambda} - 1\right)\right).
\end{aligned}$$

The result now follows by an application of the continuity Theorem 24.20. ■

*Remark 25.24.* Keeping the notation in Theorem 25.23, we have

$$\mathbb{E}\left[Y_{n,k}\right] = p_{n,k} \text{ and } \operatorname{Var}\left(Y_{n,k}\right) = p_{n,k}\left(1-p_{n,k}\right)$$

and

$$s_n^2 := \sum_{k=1}^n \operatorname{Var}\left(Y_{n,k}\right) = \sum_{k=1}^n p_{n,k}\left(1-p_{n,k}\right).$$

Under the assumptions of Theorem 25.23, we see that $s_n^2 \to a$ as $n \to \infty$. Let us now center and normalize the $Y_{n,k}$ by setting;

$$X_{n,k} := \frac{Y_{n,k} - p_{n,k}}{s_n}$$

so that

$$\sigma_{n,k}^2 := \operatorname{Var}\left(X_{n,k}\right) = \frac{1}{s_n^2}\operatorname{Var}\left(Y_{n,k}\right) = \frac{1}{s_n^2} p_{n,k}\left(1-p_{n,k}\right),$$

$\mathbb{E}\left[X_{n,k}\right] = 0$, $\operatorname{Var}\left(\sum_{k=1}^n X_{n,k}\right) = 1$, and the $\{X_{n,k}\}$ satisfy condition $(M)$. On the other hand for small $t$ and large $n$ we have

$$\begin{aligned}
\mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] &= \mathbb{E}\left[X_{n,k}^2 : \left|\frac{Y_{n,k} - p_{n,k}}{s_n}\right| > t\right] \\
&= \mathbb{E}\left[X_{n,k}^2 : |Y_{n,k} - p_{n,k}| > s_n t\right] \\
&\geq \mathbb{E}\left[X_{n,k}^2 : |Y_{n,k} - p_{n,k}| > 2at\right] \\
&= \mathbb{E}\left[X_{n,k}^2 : Y_{n,k} = 1\right] = p_{n,k}\left(\frac{1-p_{n,k}}{s_n}\right)^2
\end{aligned}$$

from which it follows that

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] = \lim_{n\to\infty} \sum_{k=1}^n p_{n,k}\left(\frac{1-p_{n,k}}{s_n}\right)^2 = a.$$

Therefore $\{X_{n,k}\}$ do **not** satisfy $(LC)$. Nevertheless we have by Theorem 25.23 along with Slutzky's Theorem 23.43 that

$$\sum_{k=1}^n X_{n,k} = \frac{\sum_{k=1}^n Y_{n,k} - \sum_{k=1}^n p_{n,k}}{s_n} \implies \frac{Z-a}{a}$$

where $Z$ is a Poisson random variable with mean $a$. Notice that the limit is **not** a normal random variable in agreement with Theorem 25.20.

We are now going see that we may often drop the identically distributed assumption of the $\{X_{n,k}\}_{k=1}^n$ and yet **still have** that the weak limit of the sums of the form, $\sum_{k=1}^n X_{n,k}$, are infinitely divisible distributions. In the next theorem we are going to see this is the case for weak limits under condition $(M)$.

**Theorem 25.25 (Limits under** $(M)$**).** *Suppose* $\{X_{n,k}\}_{k=1}^n$ *satisfy property* $(M)$ *and the normalizations in Assumption 2. If* $S_n := \sum_{k=1}^n X_{n,k} \implies L$ *for some random variable* $L$*, then*

$$f_L(\lambda) := \mathbb{E}\left[e^{i\lambda L}\right] = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

*for some finite positive measure,* $\nu$*, on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *with* $\nu(\mathbb{R}) \leq 1$.

**Proof.** As before, let $f_{n,k}(\lambda) = \mathbb{E}\left[e^{i\lambda X_{n,k}}\right]$ and $\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1$. By the continuity theorem we are assuming

$$\lim_{n\to\infty} f_{S_n}(\lambda) = \lim_{n\to\infty} \prod_{k=1}^n f_{n,k}(\lambda) = f(\lambda)$$

where $f(\lambda)$ is continuous at $\lambda = 0$. We are also assuming property $(M)$, i.e.

$$\lim_{n\to\infty} \max_{k\leq n} \sigma_{n,k}^2 = 0.$$

Under condition $(M)$, we expect $f_{n,k}(\lambda) \cong 1$ for $n$ large. Therefore we expect

$$f_{n,k}(\lambda) = e^{\ln f_{n,k}(\lambda)} = e^{\ln[1 + (f_{n,k}(\lambda) - 1)]} \cong e^{(f_{n,k}(\lambda) - 1)}$$

and hence that

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^n f_{n,k}(\lambda) \cong \prod_{k=1}^n e^{(f_{n,k}(\lambda)-1)} = \exp\left(\sum_{k=1}^n \left(f_{n,k}(\lambda) - 1\right)\right). \tag{25.21}$$

This is in fact correct, since Lemma 25.18 indeed implies

$$\lim_{n\to\infty}\left[\mathbb{E}\left[e^{i\lambda S_n}\right] - \exp\left(\sum_{k=1}^n \left(f_{n,k}(\lambda) - 1\right)\right)\right] = 0. \tag{25.22}$$

Since $\mathbb{E}\left[X_{n,k}\right] = 0$,

$$\begin{aligned}
f_{n,k}(\lambda) - 1 = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right] &= \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right] \\
&= \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x)
\end{aligned}$$

where $\mu_{n,k} := P \circ X_{n,k}^{-1}$ is the law of $X_{n,k}$. Therefore we have

$$\begin{aligned}
\exp\left(\sum_{k=1}^n \left(f_{n,k}(\lambda) - 1\right)\right) &= \exp\left(\sum_{k=1}^n \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x)\right) \\
&= \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) \sum_{k=1}^n d\mu_{n,k}(x)\right) \\
&= \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_n^*(x)\right) \qquad (25.23)
\end{aligned}$$

where $\nu_n^* := \sum_{k=1}^n \mu_{n,k}$. Let us further observe that

$$\int_{\mathbb{R}} x^2 d\nu_n^*(x) = \sum_{k=1}^n \int_{\mathbb{R}} x^2 d\mu_{n,k}(x) = \sum_{k=1}^n \sigma_{n,k}^2 = 1.$$

Hence if we define $d\nu_n(x) := x^2 d\nu_n^*(x)$, then $\nu_n$ is a probability measure and we have from Eqs. (25.22) and Eq. (25.23) that

$$\left| f_{S_n}(\lambda) - \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_n(x)\right)\right| \to 0. \tag{25.24}$$

Let

$$\varphi(\lambda, x) := \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} = -\frac{\lambda^2}{2}\int_0^1 e^{it\lambda x} 2(1-t)\,dt \tag{25.25}$$

(the second equality is from Taylor's theorem) and extend $\varphi(\lambda,\cdot)$ to $\bar{\mathbb{R}}$ by setting $\varphi(\lambda, \pm\infty) = 0$. Then $\{\varphi(\lambda,\cdot)\}_{\lambda\in\mathbb{R}} \subset C(\bar{\mathbb{R}})$ and therefore by Helly's selection Theorem 23.61 there is a probability measure $\bar{\nu}$ on $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and a subsequence, $\{n_l\}$ of $\{n\}$ such that $\nu_{n_l}(\varphi(\lambda,\cdot)) \to \bar{\nu}(\varphi(\lambda,\cdot))$ for all $\lambda \in \mathbb{R}$ (in fact $\nu_{n_l}(h) \to \bar{\nu}(h)$ for all $h \in C(\bar{\mathbb{R}})$). Combining this with Eq. (25.24) allows us to conclude,

$$\begin{aligned} f_L(\lambda) &= \lim_{l\to\infty} \mathbb{E}\left[e^{i\lambda S_{n_l}}\right] = \lim_{l\to\infty} \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_{n_l}^*(x)\right) \\ &= \lim_{l\to\infty} \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu_{n_l}(x)\right) = \exp\left(\int_{\bar{\mathbb{R}}} \varphi(\lambda, x)\, d\bar{\nu}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right) \end{aligned}$$

where $\nu := \bar{\nu}|_{\mathcal{B}_{\mathbb{R}}}$. The last equality follows from the fact that $\varphi(\lambda, \pm\infty) = 0$. The measure $\nu$ now satisfies, $\nu(\mathbb{R}) = \bar{\nu}(\mathbb{R}) \le \bar{\nu}(\bar{\mathbb{R}}) = 1$. ∎

We are now going to drop the assumption that $\mathrm{Var}(S_n) = 1$ for all $n$ and replace it with the following property.

**Definition 25.26.** *We say that $\{X_{n,k}\}_{k=1}^n$ has **bounded variation** $(BV)$ iff*

$$\sup_n \mathrm{Var}(S_n) = \sup_n \sum_{k=1}^n \sigma_{n,k}^2 < \infty. \tag{25.26}$$

**Corollary 25.27 (Limits under $(BV)$).** *Suppose $\{X_{n,k}\}_{k=1}^n$ are independent mean zero random variables for each $n$ which satisfy properties $(M)$ and $(BV)$. If $S_n := \sum_{k=1}^n X_{n,k} \implies L$ for some random variable $L$, then*

$$f_L(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right) \tag{25.27}$$

*where $\nu$ – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.*

**Proof.** Let $s_n^2 := \mathrm{Var}(S_n)$. If $\lim_{n\to\infty} s_n = 0$, then $S_n \to 0$ in $L^2$ and hence weakly, therefore Eq. (25.27) holds with $\nu \equiv 0$. So let us now suppose $\lim_{n\to\infty} s_n \neq 0$. Since $\{s_n\}_{n=1}^\infty$ is bounded, we may by passing to a subsequence if necessary, assume $\lim_{n\to\infty} s_n = s > 0$. By replacing $X_{n,k}$ by $X_{n,k}/s_n$ and hence $S_n$ by $S_n/s_n$, we then know by Slutzky's Theorem 23.43 that $S_n/s_n \implies L/s$. Hence by an application of Theorem 25.25, we may conclude

$$f_L(\lambda/s) = f_{L/s}(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

where $\nu$ – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu(\mathbb{R}) \le 1$. Letting $\lambda \to s\lambda$ in this expression then implies

$$\begin{aligned} f_L(\lambda) &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda s x} - 1 - i\lambda s x}{x^2} d\nu(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda s x} - 1 - i\lambda s x}{(sx)^2} s^2 d\nu(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_s(x)\right) \end{aligned}$$

where $\nu_s$ is the finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ defined by

$$\nu_s(A) := s^2 \nu\left(s^{-1}A\right) \text{ for all } A \in \mathcal{B}_{\mathbb{R}}.$$

∎

From Eq. (25.25) we see that $\varphi(\lambda, x) := \left(e^{i\lambda x} - 1 - i\lambda x\right)/x^2$ is a smooth function of $(\lambda, x)$. Moreover,

$$\frac{d}{d\lambda}\varphi(\lambda, x) = \frac{ixe^{i\lambda x} - ix}{x^2} = i\frac{e^{i\lambda x} - 1}{x}$$

and

$$\frac{d^2}{d\lambda^2}\varphi(\lambda, x) = i\frac{ixe^{i\lambda x}}{x} = -e^{i\lambda x}.$$

Using these remarks and the fact that $\nu(\mathbb{R}) < \infty$, it is easy to see that

$$f_L'(\lambda) = \left(\int_{\mathbb{R}} i\frac{e^{i\lambda x} - 1}{x} d\nu_s(x)\right) f_L(\lambda)$$

and

$$f_L''(\lambda) = \left(\int_{\mathbb{R}} -e^{i\lambda x} d\nu_s(x) + \left[\left(\int_{\mathbb{R}} i\frac{e^{i\lambda x} - 1}{x} d\nu_s(x)\right)^2\right]\right) f_L(\lambda)$$

and in particular, $f_L'(0) = 0$ and $f_L''(0) = -\nu_s(\mathbb{R})$. Therefore by Theorem 24.8 the probability measure $\mu$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\hat{\mu}(\lambda) = f_L(\lambda)$ has mean zero and variance, $\nu_s(\mathbb{R}) < \infty$. This later condition reflects the $(BV)$ assumption that we made.

**Theorem 25.28.** *The following class of* ***symmetric*** *distributions on* $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ *are equal;*

1. $C_1$ *– all possible limiting distributions under properties* $(M)$ *and* $(BV)$.
2. $C_2$ *– all distributions with characteristic functions of the form given in Corollary 25.27.*
3. $C_3$ *– all infinitely divisible distributions with mean zero and finite variance.*

**Proof.** The inclusion, $C_1 \subset C_2$, is the content of Corollary 25.27. For $C_2 \subset C_3$, observe that if

$$\hat{\mu}(\lambda) = \exp\left(\int_\mathbb{R} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

then $\hat{\mu}(\lambda) = [\hat{\mu}_n(\lambda)]^n$ where $\mu_n$ is the unique probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that

$$\hat{\mu}_n(\lambda) = \exp\left(\int_\mathbb{R} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} \frac{1}{n} d\nu(x)\right).$$

For $C_3 \subset C_1$, simply define $\{X_{n,k}\}_{k=1}^n$ to be i.i.d with $\mathbb{E}\left[e^{i\lambda X_{n,k}}\right] = \hat{\mu}_n(\lambda)$. In this case $S_n = \sum_{k=1}^n X_{n,k} \overset{d}{=} \mu$. ■

## 25.3 Stable Distributions

**Definition 25.29.** *A non-degenerate distribution* $\mu = \mathrm{Law}(X)$ *on* $\mathbb{R}$ *is* ***stable*** *if whenever* $X_1$ *and* $X_2$ *are independent copies of* $X$*, then for all* $a, b \in \mathbb{R}$ *there exists* $c, d \in \mathbb{R}$ *such that* $aX_1 + bX_2 \overset{d}{=} cX + d$ *with some constants* $c$ *and* $d$.

*Example 25.30.* Any Gaussian random variable is stable. Indeed if $X \overset{d}{=} \sigma N + \mu$ where $\sigma > 0$ and $\mu \in \mathbb{R}$ and $N = N(0,1)$, then $X_i = \sigma N_i + \mu$ where $N_1$ and $N_2$ are independent with $N_i \overset{d}{=} N$ we will have $aX_1 + bX_2$ is Gaussian mean $(a+b)\mu$ and variance $\left(a^2 + b^2\right)\sigma^2$ so that

$$\begin{aligned} aX_1 + bX_2 &\overset{d}{=} \sqrt{(a^2+b^2)}\sigma N + (a+b)\mu \\ &\overset{d}{=} \sqrt{(a^2+b^2)}(X - \mu) + (a+b)\mu \\ &= \sqrt{(a^2+b^2)}X + \left(a + b - \sqrt{(a^2+b^2)}\right)\mu. \end{aligned}$$

*Example 25.31.* Poisson random variables are not stable. For suppose that $Z = Pois(\rho)$ and $Z_1 \overset{d}{=} Z_2 \overset{d}{=} Z$, then $Z_1 + Z_2 \overset{d}{=} Pois(2\rho)$. If we could find $a, b$ such that

$$Pois(2\rho) \overset{d}{=} Z_1 + Z_2 \overset{d}{=} aZ + b$$

we would have

$$e^{-2\rho}\frac{(2\rho)^n}{n!} = P(aZ + b = n) = P\left(Z = \frac{n-b}{a}\right) \text{ for all } n.$$

In particular this implies that $\frac{n-b}{a} = k_n \in \mathbb{N}_0$ for all $n \in \mathbb{N}_0$ and the map $n \to k_n$ must be invertible so as probabilities are conserved. This can only be the case if $a = 1$ and $b = 0$ and we would conclude that $Z \overset{d}{=} Pois(2\rho)$ which is absurd.

**Lemma 25.32.** *Suppose that* $\{X_i\}_{i=1}^n$ *are i.i.d. random variables such that* $X_1 + \cdots + X_n = c$ *a.s., then* $X_i = c/n$ *a.s.*

**Proof.** Let $f(\lambda) := \mathbb{E}e^{i\lambda X_1}$, then

$$e^{i\lambda c} = \mathbb{E}\left[e^{i\lambda(X_1 + \cdots + X_n)}|X_1\right] = e^{i\lambda X_1} f(\lambda) \text{ a.s.}$$

from which it follows that $f(\lambda) = e^{i\lambda(c - X_1)}$ a.s. and in particular for an $\omega$ where this equality holds we find, $f(\lambda) = e^{i\lambda(c - X_1(\omega))} = e^{i\lambda c'}$. By uniqueness of the Fourier transform it follows that $X_1 = c'$ a.s. and therefore $c = X_1 + \cdots + X_n = nc'$ a.s., i.e. $c' = c/n$. ■

**Lemma 25.33.** *If* $\mu$ *is a stable distribution then it is infinitely divisible.*

**Proof.** Let $\{X_n\}_{n=1}^N$ be i.i.d. random variables with $\mathrm{Law}(X_n) = \mu = \mathrm{Law}(X)$. As $\mu$ is stable we know that

$$X_1 + \cdots + X_N \overset{d}{=} aX + b. \tag{25.28}$$

As $\mu$ is non-degenerate, it follows from Lemma 25.32 that $a \neq 0$, Therefore form Eq. (25.28) we find,

$$X \overset{d}{=} \sum_{i=1}^N \frac{1}{a}(X_i - b/N)$$

and this shows that $X$ is infinitely divisible. ■

The converse of this lemma is not true as is seen by considering Poisson random variables, see Example 25.31. The following characterization of the stable law may be found in [6, Chapter 9.9]. For a whole book about stable laws and their properties see Samorodnitsky and Taqqu [40].

**Theorem 25.34.** *A probability measure $\mu$ on $\mathbb{R}$ is a stable distribution iff $\mu$ is Gaussian or $\hat{\mu}(\lambda) = e^{\psi(\lambda)}$ where*

$$\psi(\lambda) = i\lambda b + \int_{\mathbb{R}} \left( e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2} \right) \frac{m_1 1_{x>0} - m_2 1_{x<0}}{|x|^{1+\alpha}} dx$$

*for some constants, $0 < \alpha < 2$, $m_i \geq 0$ and $b \in \mathbb{R}$.*

To get some feeling for this theorem. Let us consider the case of a stable random variable $X$ which is **also** assumed to be symmetric. In this case if $X_1, X_2$ are independent copies of $X$ and $a, b \in \mathbb{R}$ and $c = c(a_1, a_2)$ and $d = d(a_1, a_2)$ are then chosen so that $aX_1 + bX_2 \overset{d}{=} cX + d$, we must have that $d = 0$ and may take $c > 0$ by the symmetry assumption. Letting $f(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$ we may now conclude that

$$f(a\lambda) \cdot f(b\lambda) = \mathbb{E}\left[e^{i\lambda(aX_1 + bX_2)}\right] = \mathbb{E}\left[e^{i\lambda cX}\right] = f(c\lambda).$$

It turns out the solution to these functional equation are of the form $f(\lambda) = e^{-k|\lambda|^\alpha}$. If $f(\lambda)$ is of this form then

$$f(a\lambda) \cdot f(b\lambda) = \exp\left(-k\left(|a|^\alpha + |b|^\alpha\right)|\lambda|^\alpha\right) = f(c\lambda)$$

where $c = \left(|a|^\alpha + |b|^\alpha\right)^{1/\alpha}$. Moreover it turns out the $f$ is a characteristic function when $0 < \alpha \leq 2$. The case $\alpha = 2$ is the Gaussian case, then case $\alpha = 1$ is the Cauchy distribution, for example if

$$d\mu(x) = \frac{1}{\pi(1+x^2)} dx \text{ then } \hat{\mu}(\lambda) = e^{-|\lambda|}.$$

For $\alpha \leq 1$ we find that we have

$$f'(\lambda) = -k|\lambda|^{\alpha-1} f(\lambda) \leq 0 \text{ and}$$
$$f''(\lambda) = \left[k^2 |\lambda|^{2\alpha-2} - k(\alpha-1)|\lambda|^{\alpha-2}\right] f(\lambda) \geq 0$$

so that $f$ is a decreasing convex symmetric function for $\lambda \geq 0$. Therefore by Polya's criteria of Exercise 24.7 it follows that $e^{-k|\lambda|^\alpha}$ is the characteristic function of a probability measure for $0 \leq \alpha \leq 1$. The full proof is not definitely not given here.

## 25.4 *Appendix: Lévy exponent and Lévy Process facts – Very Preliminary!!

We would like to characterize all processes with independent stationary increments with values in $\mathbb{R}$ or more generally $\mathbb{R}^d$. We begin with some more examples.

**Proposition 25.35.** *For every finite measure $\nu$, the function*

$$f(\lambda) := \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

*is the characteristic function of a probability measure, $\mu = \mu_\nu$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The convention here is that*

$$\frac{e^{i\lambda x} - 1 - i\lambda x}{x^2}|_{x=0} := \lim_{x \to 0} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} = -\frac{1}{2}\lambda^2.$$

**Proof.** This is the content of Exercise 24.5 ■

1. If $\{X_t\}_{t\geq 0}$ is a right continuous process with stationary and independent increments, then let $f_t(\lambda) := \mathbb{E}\left[e^{i\lambda(X_{t+\sigma} - X_\sigma)}\right]$ for any $\sigma \geq 0$. It then follows that
$$\begin{aligned} f_{t+s}(\lambda) &= \mathbb{E}\left[e^{i\lambda(X_{t+s} - X_0)}\right] = \mathbb{E}\left[e^{i\lambda(X_{t+s} - X_t + X_t - X_0)}\right] \\ &= \mathbb{E}\left[e^{i\lambda(X_{t+s} - X_t)}\right] \cdot \mathbb{E}\left[e^{i\lambda(X_t - X_0)}\right] \\ &= f_s(\lambda) \cdot f_t(\lambda). \end{aligned}$$
The right continuity of $X_t$ now insures that $f_t$ is also right continuous. The only solution to the above functional equation is therefore of the form, $f_t(\lambda) = e^{t\psi(\lambda)}$ for some function $\psi(\lambda)$. Since
$$e^{t \operatorname{Re} \psi(\lambda)} = |f_t(\lambda)| \leq 1$$
it follows that $\operatorname{Re}\psi(\lambda) \leq 0$. Let $\lambda \in \mathbb{R}$ be fixed and define $h(t) := f_t(\lambda)$, then $h$ is right continuous, $h(0) = 1$, and $h(t+s) = h(t)h(s)$. Let ln be a branch of the logarithm defined near 1 such that $\ln 1 = 0$. Then there exists $\varepsilon$ such that for all $t \leq \varepsilon$ we have $g(t) := \ln h(t)$ is well defined and $g(t)$ satisfies, $g(t+s) = g(t) + g(s)$ for all $0 \leq s, t \leq \varepsilon$. We now set $g_\varepsilon(t) := g(\varepsilon t)$ and then $g_\varepsilon(s+t) = g_\varepsilon(s) + g_\varepsilon(t)$ for all $0 \leq s, t \leq 1$ and is still right continuous. As usual it now follows that $g_\varepsilon(1) = g_\varepsilon(n \cdot 1/n) = n \cdot g_\varepsilon(1/n)$ for all $n$ and therefore for all $0 \leq k \leq n$, we have $g_\varepsilon(k/n) = \frac{k}{n} g_\varepsilon(1)$. Using the right continuity of $g_\varepsilon$ it now follows that $g_\varepsilon(t) = t g_\varepsilon(1)$ for all $0 \leq t < 1$. Thus we have shown $g(\varepsilon t) = t g(\varepsilon)$ for $0 \leq t < 1$ and therefore if we set $\theta := g(\varepsilon)/\varepsilon$ we have shown $g(t) = t\theta$ for $t \in [0, \varepsilon)$ that is ,
$$h(t) = e^{t\theta} \text{ for } 0 \leq t < \varepsilon.$$
This formula is now seen to be correct for all $t \geq 0$. Indeed if $t = k\varepsilon/2 + \tau$ with $0 \leq \tau < \varepsilon/2$, then

$$h(t) = h(\varepsilon/2)^k h(\tau) = \left[e^{\theta\varepsilon/2}\right]^k e^{\tau\theta} = e^{\theta[k\varepsilon/2+\tau]} = e^{t\theta}.$$

Thus we have shown that $f_t(\lambda) = e^{t\psi(\lambda)}$ for some function $\psi(\lambda)$. Let us further observe that

$$\psi(\lambda) = \lim_{t\downarrow 0} \frac{f_t(\lambda) - 1}{t}$$

from which it follows that $\psi$ must be measurable. Furthermore,

$$\psi(-\lambda) = \lim_{t\downarrow 0} \frac{f_t(-\lambda) - 1}{t} = \lim_{t\downarrow 0} \frac{\overline{f_t(\lambda)} - 1}{t} = \overline{\psi(\lambda)}.$$

We are going to show more.

2. Let $\{z_i\}_{i=1}^n \subset \mathbb{C}$ such that $\sum_{i=1}^n z_i = 1$ and $\{\lambda_i\}_{i=1}^n \subset \mathbb{R}$, then

$$\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j = \lim_{t\downarrow 0} \sum_{i,j=1}^n \frac{f_t(\lambda_i - \lambda_j) - 1}{t} z_i \bar{z}_j = \lim_{t\downarrow 0} \frac{1}{t} \sum_{i,j=1}^n f_t(\lambda_i - \lambda_j) z_i \bar{z}_j$$

while for any $\{z_i\}_{i=1}^n \subset \mathbb{C}$ we have

$$\begin{aligned}\sum_{i,j=1}^n f_t(\lambda_i - \lambda_j) z_i \bar{z}_j &= \sum_{i,j=1}^n \mathbb{E}\left[e^{i(\lambda_i - \lambda_j)X_t}\right] z_i \bar{z}_j \\ &= \sum_{i,j=1}^n \mathbb{E}\left[e^{i\lambda_i X_t} z_i \cdot e^{-i\lambda_j X_t} \bar{z}_j\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n e^{i\lambda_i X_t} z_i \cdot \sum_{j=1}^n e^{-i\lambda_j X_t} \bar{z}_j\right] \\ &= \mathbb{E}\left[\left|\sum_{i=1}^n e^{i\lambda_i X_t} z_i\right|^2\right] \geq 0.\end{aligned}$$

Therefore it follows that when $\sum_{i=1}^n z_i = 1$ then $\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j \geq 0$. We say the $\psi$ is conditionally positive definite in this case.

3. The Schoenberg correspondence says (see [1, Theorem 1.1.13]) that if $\psi$ is continuous at zero, $\psi(-\lambda) = \overline{\psi(\lambda)}$ and $\psi$ is conditionally positive definite, then $e^{t\psi(\lambda)}$ is a characteristic function. We will prove this below using Bochner's Theorem 24.46.
4. But first some examples;

a) Let $\psi(\lambda) = i\lambda a - b\lambda^2$ with $a \in \mathbb{R}$ and $b \geq 0$.Then $\psi(-\lambda) = -i\lambda a - b\lambda^2 = \overline{\psi(\lambda)}$ and for $\sum_{i=1}^n z_i = 1$ we have

$$\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j = \sum_{i,j=1}^n \left[-i(\lambda_i - \lambda_j) a - b(\lambda_i - \lambda_j)^2\right] z_i \bar{z}_j.$$

Noting that

$$\sum_{i,j=1}^n \lambda_i z_i \bar{z}_j = \sum_{i=1}^n \lambda_i z_i \sum_{j=1}^n \bar{z}_j = \sum_{i=1}^n \lambda_i z_i \cdot 0 = 0$$

and similarly that $\sum_{i,j=1}^n \left[\lambda_i^2\right] z_i \bar{z}_j = 0$, it follows that

$$\begin{aligned}\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j &= \sum_{i,j=1}^n \left[-b(-2\lambda_i\lambda_j)^2\right] z_i \bar{z}_j \\ &= 2b\left|\sum_{i,j=1}^n \lambda_i z_i\right|^2 \geq 0.\end{aligned}$$

b) Suppose that $\{Z_i\}_{i=1}^\infty$ are i.i.d. random variables and $\{N\}$ is an independent Poisson process with intensity $\lambda$. Let $X := Z_1 + \cdots + Z_N$, then

$$\begin{aligned}f_X(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right] &= \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda X} : N = n\right] \\ &= \sum_{n=0}^\infty \mathbb{E}\left[e^{i\lambda[Z_1+\cdots+Z_n]} : N = n\right] \\ &= e^{-\lambda} \sum_{n=0}^\infty \frac{\lambda^n}{n!} \left[f_{Z_1}(\lambda)\right]^n = \exp\left(\lambda\left(f_{Z_1}(\lambda) - 1\right)\right).\end{aligned}$$

So in this case $\psi(\lambda) = f_{Z_1}(\lambda) - 1$ and we know by the theory above that $\psi(\lambda)$ is conditionally positive definite.

**Lemma 25.36.** *Suppose that $\{A_{ij}\}_{i,j=1}^d \subset \mathbb{C}$ is a matrix such that $A^* = A$ and $A \geq 0$. Then for all $n \in \mathbb{N}_0$, the matrix with entries $\left(A_{ij}^n\right)_{i,j=1}^n$ is positive semi-definite.*

**Proof.** Since $A_{ij} = (Ae_j, e_i)$ where $(v, w) := \sum_{j=1}^d v_j \bar{w}_j$ is the standard inner product on $\mathbb{C}^d$, it follows that

$$A_{ij}^n = \left(A^{\otimes n} e_j^{\otimes n}, e_i^{\otimes n}\right)$$

and therefore,

$$\sum_{i,j=1}^{d} A_{ij}^{n} \bar{z}_i z_j = \sum_{i,j=1}^{d} \left(A^{\otimes n} e_j^{\otimes n}, e_i^{\otimes n}\right) \bar{z}_i z_j = \left(A^{\otimes n} \psi, \psi\right)$$

where $\psi := \sum_{j=1}^{d} z_j e_j^{\otimes n} \in \left(\mathbb{C}^d\right)^{\otimes n}$. So it suffices to show $A^{\otimes n} \geq 0$. To do this let $\{u_i\}_{i=1}^{d}$ be an O.N. basis for $\mathbb{C}^d$ such that $Au_i = \lambda_i u_i$ for all $i$. Since $A \geq 0$ we know that $\lambda_i \geq 0$ and therefore

$$A^{\otimes n} \left(u_{i_1} \otimes \cdots \otimes u_{i_d}\right) = \left(\lambda_{i_1} \dots \lambda_{i_d}\right) \left(u_{i_1} \otimes \cdots \otimes u_{i_d}\right)$$

where $\left(\lambda_{i_1} \dots \lambda_{i_d}\right) \geq 0$. This shows that $A^{\otimes n}$ is unitarily equivalent to a diagonal matrix with non-negative entries and hence is positive semi-definite. ∎

**Proposition 25.37.** *Suppose that* $\{A_{ij}\}_{i,j=1}^{d} \subset \mathbb{C}$ *is a matrix such that* $A^* = A$ *and* $A$ *is conditionally positive definite, for example* $A_{ij} := \psi\left(\lambda_i - \lambda_j\right)$ *as above. Then the matrix with entries,* $\left(e^{A_{ij}}\right)_{i,j=1}^{d}$ *is positive definite.*

**Proof.** Let $u := (1, \dots, 1)^{\text{tr}} \in \mathbb{C}^d$. Let $\xi \in \mathbb{C}^d$ and write $\xi = z + \alpha u$ where $(z, u) = 0$ and $\alpha := (\xi, u)/d$. Letting $B := \sqrt{A}$ on $u^{\perp}$ and 0 on $\mathbb{C} \cdot u$, we have

$$\begin{aligned}
(A\xi, \xi) &= \left(A\left(z + \alpha u\right), z + \alpha u\right) \\
&= (Az, z) + 2\operatorname{Re}\left[\bar{\alpha}\left(Az, u\right)\right] + |\alpha|^2 \left(Au, u\right) \\
&= (Az, z) + 2\operatorname{Re}\left[\bar{\alpha}\left(B^2 z, u\right)\right] + |\alpha|^2 \left(Au, u\right) \\
&= (Az, z) + 2\operatorname{Re}\left[\bar{\alpha}\left(Bz, B^* u\right)\right] + |\alpha|^2 \left(Au, u\right) \\
&\geq (Az, z) - 2\left\|Bz\right\| \cdot |\alpha| \left\|B^* u\right\| + |\alpha|^2 \left(Au, u\right) \\
&\geq (Az, z) - \left[\left\|Bz\right\|^2 + |\alpha|^2 \left\|B^* u\right\|^2\right] + |\alpha|^2 \left(Au, u\right) \\
&= |\alpha|^2 \left[\left(Au, u\right) - \left\|B^* u\right\|^2\right].
\end{aligned}$$

Since

$$\left(u\ u^{\text{tr}} \xi, \xi\right) = |\alpha|^2 \left(u\ u^{\text{tr}} u, u\right) = |\alpha|^2 d^2,$$

it follows that

$$\left(\left(A + \lambda u\ u^{\text{tr}}\right) \xi, \xi\right) \geq |\alpha|^2 \left[\left(Au, u\right) - \left\|B^* u\right\|^2 + \lambda d^2\right] \geq 0$$

provided $\lambda d^2 \geq \left\|B^* u\right\|^2 - \left(Au, u\right)$.

We now fix such a $\lambda \in \mathbb{R}$ so that $\left(A + \lambda u\ u^{\text{tr}}\right) \geq 0$. It then follows from Lemma 25.36 that



$$e^{\lambda} e^{A_{ij}} = e^{A_{ij} + \lambda} = e^{\left(A + \lambda u\ u^{\text{tr}}\right)_{ij}} = \sum_{n=0}^{\infty} \frac{\left(A + \lambda u\ u^{\text{tr}}\right)_{ij}^{n}}{n!}$$

are the matrix entries of a positive definite matrix. Scaling this matrix by $e^{-\lambda} > 0$ then gives the result that $\left(e^{A_{ij}}\right)_{i,j} \geq 0$. ∎

As a consequence it follows that $e^{t\psi(\lambda)}$ is a positive definite function whenever $\psi$ is conditionally positive definite.

**Proposition 25.38.** *Suppose that* $\{Z_i\}_{i=1}^{\infty}$ *are i.i.d. random vectors in* $\mathbb{R}^d$ *with* $\text{Law}\left(Z_i\right) = \mu$ *and* $\{N_t\}_{t \geq 0}$ *be an independent Poisson process with intensity* $\lambda$. *Then* $\{X_t := S_{N_t}\}_{t \geq 0}$ *is a Lévy process with* $\mathbb{E}\left[e^{ik \cdot X_t}\right] = e^{t\psi(k)}$ *where*

$$\psi\left(k\right) = \lambda \int_{\mathbb{R}^n} \left(e^{ik \cdot x} - 1\right) d\mu\left(x\right) = \mathbb{E}\left[e^{ik \cdot Z_1}\right].$$

**Proof.** It has already been shown in Theorem 19.30 that $\{X_t\}_{t \geq 0}$ has stationary independent increments and being right continuous it is a Lévy process. It only remains to compute the Fourier transform,

$$\begin{aligned}
\mathbb{E}\left[e^{ik \cdot X_t}\right] &= \left[Q_t\left(x \to e^{ik \cdot x}\right)\right](0) \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \mathbb{E}\left[e^{ik \cdot \left(Z_1 + \dots + Z_n\right)}\right] = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \hat{\mu}\left(k\right)^n \\
&= e^{t\lambda\left(\hat{\mu}(k) - 1\right)} = \exp\left(t\lambda \int_{\mathbb{R}^n} \left(e^{ik \cdot x} - 1\right) d\mu\left(x\right)\right).
\end{aligned}$$

∎

More generally, if we let $B_t$ be Brownian motion in $\mathbb{R}^n$ with $\operatorname{Cov}\left(B_t^i, B_t^j\right) = A_{ij} t$ and $b \in \mathbb{R}^n$, then assuming $B$ and $X$ above are independent, then $X_t = bt + B_t + X_t$ is again a Levy process whose Fourier transform is given by,

$$\mathbb{E}\left[e^{ik \cdot X_t}\right] = \exp\left(ibt + Ak \cdot k + \lambda \int_{\mathbb{R}^n} \left(e^{ik \cdot x} - 1\right) d\mu\left(x\right)\right).$$

Thus

$$\psi\left(\lambda\right) = ibt + Ak \cdot k + \lambda \int_{\mathbb{R}^n} \left(e^{ik \cdot x} - 1\right) d\mu\left(x\right)$$

is a Lévy exponent for all choice of $b \in \mathbb{R}^n$, all $\lambda > 0$, probability measures $\mu$ on $\mathbb{R}^n$, and $A \geq 0$.

Lévy proved that in general $\psi\left(k\right)$ will be a Lévy exponent iff $\psi$ has the form given in Eq. (25.29) below.

**Theorem 25.39 (Lévy Kintchine formula).** *If* $\psi$ *is continuous at zero and conditionally positive definite, then*

$$\psi(\lambda) = i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\leq 1}\right) d\nu(x) \qquad (25.29)$$

*for some* $b \in \mathbb{R}$, $a \geq 0$, *and some measure* $\nu$ *such that*

$$\int_{\mathbb{R}\setminus\{0\}} \left(x^2 \wedge 1\right) d\nu(x) < \infty.$$

# Part V

## Stochastic Processes II

We are now going to discuss continuous time stochastic processes in more detail. We will be using Poisson processes (Definition 13.8) and Brownian motion (Definition 19.24) as our model cases. Up to now we have not proved the existence of Brownian motion. This lapse will be remedied in the next couple of chapters. We are going to begin by constructing a process $\{B_t\}_{t\geq 0}$ satisfying all of the properties of a Brownian motion in Definition 19.24 except for the continuity of the sample paths. We will then use Kolmogorov's continuity criteria (Theorem 27.8) to show we can "modify" this process in such a way so as to produce an example of Brownian motion. We start with a class of random fields which are relatively easy to understand. (BRUCE – metion the free Euclidean field and its connections to SLE.)

# Gaussian Random Fields

Recall from Section 11.9 (which the reader should review if necessary) that a random variable, $Y : \Omega \to \mathbb{R}$ is said to be **Gaussian** if

$$\mathbb{E}e^{i\lambda Y} = \exp\left(-\frac{1}{2}\lambda^2 \operatorname{Var}(Y) + i\lambda \mathbb{E}Y\right) \ \forall \lambda \in \mathbb{R}.$$

More generally a random vector, $X : \Omega \to \mathbb{R}^N$, is said to be **Gaussian** if $\lambda \cdot X$ is a Gaussian random variable for all $\lambda \in \mathbb{R}^N$. Equivalently put, $X : \Omega \to \mathbb{R}^N$ is **Gaussian** provided

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right) \ \forall \lambda \in \mathbb{R}^N. \tag{26.1}$$

*Remark 26.1.* To conclude that a random vector, $X : \Omega \to \mathbb{R}^N$, is Gaussian it is **not** enough to check that each of its components are Gaussian random variables. The following simple counter example was provided by Nate Eldredge. Let $X \overset{d}{=} N(0,1)$ and $Y$ be an independent Bernoulli random variable with $P(Y=1) = P(Y=-1) = 1/2$. Then the random vector, $(X, X \cdot Y)^{\mathrm{tr}}$ has Gaussian components but is not Gaussian.

**Exercise 26.1 (Same as Exercise 11.10.).** Prove the assertion made in Remark 26.1 by computing $\mathbb{E}\left[e^{i(\lambda_1 X + \lambda_2 XY)}\right]$. (Another proof that $(X, X\cdot Y)^{\mathrm{tr}}$ is not Gaussian follows from the fact that $X$ and $XY$ are uncorrelated but not independent[1] which would then contradict Lemma 12.25.)

## 26.1 Gaussian Integrals

The following theorem gives a useful way of computing Gaussian integrals of polynomials and exponential functions.

[1] To formally see that they are not independent, observe that $|X| \le \frac{1}{2}$ iff $|XY| \le \frac{1}{2}$ and therefore,

$$P\left(|X| \le \frac{1}{2} \text{ and } |XY| \le \frac{1}{2}\right) = P\left(|X| \le \frac{1}{2}\right) =: \alpha$$

while

$$P\left(|X| \le \frac{1}{2}\right) P\left(|XY| \le \frac{1}{2}\right) = \alpha^2 \ne \alpha.$$

**Theorem 26.2.** *Suppose $X \overset{d}{=} N(Q,0)$ where $Q$ is a $N \times N$ symmetric positive definite matrix. Let $L = L^Q := Q_{ij}\partial_i\partial_j$ (sum on repeated indices) where $\partial_i := \partial/\partial x_i$. Then for any polynomial function, $q : \mathbb{R}^N \to \mathbb{R}$,*

$$\mathbb{E}[q(X)] = \left(e^{\frac{1}{2}L}q\right)(0) := \sum_{n=0}^{\infty} \frac{1}{n!}\left(\left(\frac{L}{2}\right)^n q\right)(0) \ \text{(a finite sum)}. \tag{26.2}$$

**Proof. First Proof.** The first proof is conceptually clear but technically a bit more difficult. In this proof we will begin by proving Eq. (26.2) when $q(x) = e^{i\lambda \cdot x}$ where $\lambda \in \mathbb{R}^N$. The function $q$ is not a polynomial, but never mind. In this case,

$$\mathbb{E}[q(X)] = \mathbb{E}\left[e^{i\lambda\cdot X}\right] = e^{-\frac{1}{2}Q\lambda\cdot\lambda}.$$

On the other hand,

$$\left(\frac{1}{2}Lq\right)(x) = \frac{1}{2}Q_{ij}\partial_i\partial_j e^{i\lambda\cdot x} = \frac{1}{2}(Q\lambda\cdot\lambda)e^{i\lambda\cdot x} = -\frac{1}{2}(Q\lambda\cdot\lambda)q(x).$$

Therefore,

$$e^{\frac{1}{2}L}q = \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n q = e^{-\frac{1}{2}Q\lambda\cdot\lambda}q$$

and hence

$$\left(e^{\frac{1}{2}L}q\right)(0) = e^{-\frac{1}{2}Q\lambda\cdot\lambda}.$$

Thus we have shown

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = e^{\frac{1}{2}L}e^{i\lambda\cdot x}|_{x=0}.$$

The result now formally follows by differentiating this equation in $\lambda$ and then setting $\lambda = 0$. Indeed observe that

$$\begin{aligned}\mathbb{E}\left[(iX)^\alpha\right] &= \partial_\lambda^\alpha \mathbb{E}\left[e^{i\lambda\cdot X}\right]|_{\lambda=0} = \partial_\lambda^\alpha e^{\frac{1}{2}L}e^{i\lambda\cdot x}|_{x=0,\lambda=0} \\ &= e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x}|_{x=0,\lambda=0} = e^{\frac{1}{2}L}(ix)^\alpha|_{x=0}.\end{aligned}$$

To justify this last equation we must show,

$$\partial_\lambda^\alpha e^{\frac{1}{2}L}e^{i\lambda\cdot x} = e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x}$$

which is formally true since mixed partial derivatives commute. However there is also an infinite sum involved so we have to be a bit more careful. To see what is involved, on one hand

$$\partial_\lambda^\alpha e^{\frac{1}{2}L} e^{i\lambda\cdot x} = \partial_\lambda^\alpha \sum_{n=0}^\infty \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}$$

while on the other,

$$\begin{aligned} e^{\frac{1}{2}L}\partial_\lambda^\alpha e^{i\lambda\cdot x} &= \sum_{n=0}^\infty \frac{1}{n!}\left(\left(\frac{L}{2}\right)^n \partial_\lambda^\alpha e^{i\lambda\cdot x}\right) = \sum_{n=0}^\infty \frac{1}{n!}\partial_\lambda^\alpha\left(\left(\frac{L}{2}\right)^n e^{i\lambda\cdot x}\right) \\ &= \sum_{n=0}^\infty \partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right). \end{aligned}$$

Thus to complete the proof we must show,

$$\partial_\lambda^\alpha \sum_{n=0}^\infty \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x} = \sum_{n=0}^\infty \partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right).$$

Perhaps the easiest way to do this would be to use the Cauchy estimates[2] which allow one to show that if $\{f_n(\lambda)\}_{n=0}^\infty$ is a sequence of analytic functions such that $\sum_{n=0}^\infty f_n(\lambda)$ is uniformly convergent on compact subsets, then $\sum_{n=0}^\infty \partial_\lambda^\alpha f_n(\lambda)$ is also uniformly convergent on compact subsets and therefore,

$$\partial_\lambda^\alpha \sum_{n=0}^\infty f_n(\lambda) = \sum_{n=0}^\infty \partial_\lambda^\alpha f_n(\lambda).$$

Now apply this result with $f_n(\lambda) := \frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}$ to get the result. The details are left to the reader.

This proof actually shows more than what is claimed. Namely, 1. $Q$ may be only non-negative definite and 2. Eq. (26.2) holds for $q(x) = p(x)e^{i\lambda\cdot x}$ where $\lambda\in\mathbb{R}^N$ and $p$ is a polynomial.

**Second Proof.** Let

$$u(t,y) := \mathbb{E}\left[q\left(y+\sqrt{t}X\right)\right] = Z^{-1}\int_{\mathbb{R}^N} q\left(y+\sqrt{t}x\right)e^{-Q^{-1}x\cdot x/2}dx \tag{26.3}$$

$$= Z^{-1}\int_{\mathbb{R}^N} q(y+x)\frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}}dx. \tag{26.4}$$

[2] If you want to avoid the Cauchy estimates it would suffice to show by hand that

$$\sum_{n=0}^\infty \sup_{|\lambda|\le R}\left|\partial_\lambda^\alpha\left(\frac{1}{n!}\left(-\frac{1}{2}Q\lambda\cdot\lambda\right)^n e^{i\lambda\cdot x}\right)\right| < \infty$$

for all multi- indices, $\alpha$.

One now verifies that

$$\partial_t \frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}} = \frac{1}{2}L\frac{e^{-Q^{-1}x\cdot x/2t}}{t^{n/2}}.$$

Using this result and differentiating under the integral in Eq. (26.4) then shows,

$$\partial_t u(t,y) = \frac{1}{2}L_y u(t,y) \text{ with } u(0,y) = q(y).$$

Moreover, from Eq. (26.3), one easily sees that $u(t,y)$ is a polynomial in $(t,y)$ and the degree in $y$ is the same as the degree of $q$. On the other hand,

$$v(t,y) := \sum_{n=0}^\infty \frac{t^n}{n!}\left(\left(\frac{L}{2}\right)^n q\right)(y) = \left(e^{tL/2}q\right)(y)$$

satisfies the same equation as $u$ in the same finite dimensional space of polynomials of degree less than or equal to $\deg(q)$. Therefore by uniqueness of solutions to ODE we must have $u(t,y) = v(t,y)$. The result now follows by taking $t=1$ and $y=0$ and observing that

$$\begin{aligned} u(1,0) &= \mathbb{E}\left[q\left(0+\sqrt{1}X\right)\right] = \mathbb{E}[q(X)] \text{ and} \\ v(1,0) &= \left(e^{L/2}q\right)(0). \end{aligned}$$

**Third Proof.** Let $u\in\mathbb{R}^N$. Since

$$\partial_u \exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right) = -\left(Q^{-1}x\cdot u\right)\exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)$$

it follow by integration by parts that

$$\begin{aligned} \mathbb{E}\left[\left(Q^{-1}X\cdot u\right)p(X)\right] &= -\frac{1}{Z}\int_{\mathbb{R}^N} p(x)\,\partial_u \exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)dx \\ &= \frac{1}{Z}\int_{\mathbb{R}^N}(\partial_u p)(x)\exp\left(-\frac{1}{2}Q^{-1}x\cdot x\right)dx \\ &= \mathbb{E}\left[(\partial_u p)(X)\right]. \end{aligned}$$

Replacing $u$ by $Qu$ in this equation leads to important identity,

$$\mathbb{E}\left[(X\cdot u)\,p(X)\right] = \mathbb{E}\left[(\partial_{Qu}p)(X)\right]. \tag{26.5}$$

It is clear that using this identity and induction it would be possible to compute $\mathbb{E}[p(X)]$ for any polynomial $p$. So to finish the proof it suffices to show

$$e^{L/2}\left((x\cdot u)\,p(x)\right)|_{x=0} = e^{L/2}\left((\partial_{Qu}p)(x)\right)|_x = 0.$$

This is correct notice that

$$e^{L/2}\left((x\cdot u)\,p(x)\right)=e^{L/2}\left((x\cdot u)\,e^{-L/2}e^{L/2}p(x)\right)=e^{L/2}M_{(x\cdot u)}e^{-L/2}e^{L/2}p(x)$$

Letting $q$ be any polynomial and

$$F_t:=e^{tL/2}M_{(x\cdot u)}e^{-tL/2},$$

we have

$$\frac{d}{dt}F_t=e^{tL/2}\left[\frac{L}{2},M_{(x\cdot u)}\right]e^{-tL/2}=e^{tL/2}\partial_{Qu}e^{-tL/2}=\partial_{Qu}$$

and therefore,

$$e^{L/2}M_{(x\cdot u)}e^{-L/2}=F_1=M_{(x\cdot u)}+\partial_{Qu}.$$

Hence it follows that

$$\begin{aligned}e^{L/2}\left((x\cdot u)\,p(x)\right)|_{x=0}&=\left[\left(M_{(x\cdot u)}+\partial_{Qu}\right)e^{L/2}p(x)\right]_{x=0}=\left[\partial_{Qu}e^{L/2}p(x)\right]_{x=0}\\&=\left[e^{L/2}\partial_{Qu}p(x)\right]_{x=0}\end{aligned}\tag{26.6}$$

which is the same identity as in Eq. (26.5). ■

*Example 26.3.* Suppose $X\overset{d}{=}N(1,0)\in\mathbb{R}$, then

$$\mathbb{E}\left[X^{2n}\right]=\left[e^{\Delta/2}x^{2n}\right]_{x=0}=\frac{1}{n!\cdot 2^n}\Delta^n\|x\|^{2n}=\frac{(2n)!}{2^n\cdot n!}.$$

## 26.2 Existence of Gaussian Fields

**Definition 26.4.** *Let $T$ be a set. A Gaussian random field indexed by $T$ is a collection of random variables, $\{X_t\}_{t\in T}$ on some probability space $(\Omega,\mathcal{B},P)$ such that for any finite subset, $\Lambda\subset_f T$, $\{X_t:t\in\Lambda\}$ is a Gaussian random vector.*

Associated to a Gaussian random field, $\{X_t\}_{t\in T}$, are the two functions,

$$c:T\to\mathbb{R}\text{ and }Q:T\times T\to\mathbb{R}$$

defined by $c(t):=\mathbb{E}X_t$ and $Q(s,t):=\operatorname{Cov}(X_s,X_t)$. By the previous results, the functions $(Q,c)$ uniquely determine the finite dimensional distributions $\{X_t:t\in T\}$, i.e. the joint distribution of the random variables, $\{X_t:t\in\Lambda\}$, for all $\Lambda\subset_f T$.

**Definition 26.5.** *Suppose $T$ is a set and $\{X_t:t\in T\}$ is a random field. For any $\Lambda\subset T$, let $\mathcal{B}_\Lambda:=\sigma(X_t:t\in\Lambda)$.*

**Proposition 26.6.** *Suppose $T$ is a set and $c:T\to\mathbb{R}$ and $Q:T\times T\to\mathbb{R}$ are given functions such that $Q(s,t)=Q(t,s)$ for all $s,t\in T$ and for each $\Lambda\subset_f T$*

$$\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)\geq 0\text{ for all }\lambda:\Lambda\to\mathbb{R}.$$

*Then there exists a probability space, $(\Omega,\mathcal{B},P)$, and random variables, $X_t:\Omega\to\mathbb{R}$ for each $t\in T$ such that $\{X_t\}_{t\in T}$ is a Gaussian random process with*

$$\mathbb{E}[X_s]=c(s)\text{ and }\operatorname{Cov}(X_s,X_t)=Q(s,t)\tag{26.7}$$

*for all $s,t\in T$.*

**Proof.** Since we will construct $(\Omega,\mathcal{B},P)$ by Kolmogorov's extension Theorem 19.68, let $\Omega:=\mathbb{R}^T$, $\mathcal{B}=\mathcal{B}_{\mathbb{R}^T}$, and $X_t(\omega)=\omega_t$ for all $t\in T$ and $\omega\in\Omega$. Given $\Lambda\subset_f T$, let $\mu_\Lambda$ be the unique Gaussian measure on $\left(\mathbb{R}^\Lambda,\mathcal{B}_\Lambda:=\mathcal{B}_{\mathbb{R}^\Lambda}\right)$ such that

$$\begin{aligned}&\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\mu_\Lambda(x)\\&=\exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)+i\sum_{s\in\Lambda}c(s)\,\lambda(s)\right).\end{aligned}$$

The main point now is to show $\left\{\left(\mathbb{R}^\Lambda,\mathcal{B}_\Lambda,\mu_\Lambda\right)\right\}_{\Lambda\subset_f T}$ is a consistent family of measures. For this, suppose $\Lambda\subset\Gamma\subset_f T$ and $\pi:\mathbb{R}^\Gamma\to\mathbb{R}^\Lambda$ is the projection map, $\pi(x)=x|_\Lambda$. For any $\lambda\in\mathbb{R}^\Lambda$, let $\tilde\lambda\in\mathbb{R}^\Gamma$ be defined so that $\tilde\lambda=\lambda$ on $\Lambda$ and $\tilde\lambda=0$ on $\Gamma\setminus\Lambda$. We then have,

$$\begin{aligned}&\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\left(\mu_\Gamma\circ\pi^{-1}\right)(x)\\&\quad=\int_{\mathbb{R}^\Gamma}e^{i\sum_{t\in\Lambda}\lambda(t)\pi(x)(t)}d\mu_\Gamma(x)\\&\quad=\int_{\mathbb{R}^\Gamma}e^{i\sum_{t\in\Gamma}\tilde\lambda(t)x(t)}d\mu_\Gamma(x)\\&\quad=\exp\left(-\frac{1}{2}\sum_{s,t\in\Gamma}Q(s,t)\,\tilde\lambda(s)\,\tilde\lambda(t)+i\sum_{s\in\Gamma}c(s)\,\tilde\lambda(s)\right)\\&\quad=\exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda}Q(s,t)\,\lambda(s)\,\lambda(t)+i\sum_{s\in\Lambda}c(s)\,\lambda(s)\right)\\&\quad=\int_{\mathbb{R}^\Lambda}e^{i\sum_{t\in\Lambda}\lambda(t)x(t)}d\mu_\Lambda(x).\end{aligned}$$

Since this is valid for all $\lambda \in \mathbb{R}^{\Lambda}$, it follows that $\mu_{\Gamma} \circ \pi^{-1} = \mu_{\Lambda}$ as desired. Hence by Kolmogorov's theorem, there exists a unique probability measure, $P$ on $(\Omega, \mathcal{B})$ such that

$$\int_{\Omega} f(\omega|_{\Lambda})\, dP(\omega) = \int_{\mathbb{R}^{\Lambda}} f(x)\, d\mu_{\Lambda}(x)$$

for all $\Lambda \subset_f T$ and all bounded measurable functions, $f : \mathbb{R}^{\Lambda} \to \mathbb{R}$. In particular, it follows that

$$\begin{aligned}\mathbb{E}\left[e^{i\sum_{t\in\Lambda}\lambda(t)X_t}\right] &= \int_{\Omega} e^{i\sum_{t\in\Lambda}\lambda(t)\omega(t)} dP(\omega)\\ &= \exp\left(-\frac{1}{2}\sum_{s,t\in\Lambda} Q(s,t)\,\lambda(s)\,\lambda(t) + i\sum_{s\in\Lambda} c(s)\,\lambda(s)\right)\end{aligned}$$

for all $\lambda \in \mathbb{R}^{\Lambda}$. From this it follows that $\{X_t\}_{t\in T}$ is a Gaussian random field satisfying Eq. (26.7). ■

**Exercise 26.2.** Suppose $T = [0,\infty)$ and $\{X_t : t \in T\}$ is a mean zero Gaussian random field (process). Show that $\mathcal{B}_{[0,\sigma]} \overset{X_\sigma}{\perp\!\!\!\perp} \mathcal{B}_{[\sigma,\infty)}$ for all $0 \le \sigma < \infty$ iff

$$Q(s,\sigma)\,Q(\sigma,t) = Q(\sigma,\sigma)\,Q(s,t) \;\; \forall\; 0 \le s \le \sigma \le t < \infty. \tag{26.8}$$

**Hint:** see use Exercises 16.12 and 16.11.

## 26.3 Gaussian Field Interpretation of Pre-Brownian Motion

**Lemma 26.7.** *Suppose that $\{B_t\}_{t\ge0}$ is a pre-Brownian motion as described in Proposition 25.21, also see Corollary 19.25. Then $\{B_t\}_{t\ge0}$ is a mean zero Gaussian random process with $\mathbb{E}[B_tB_s] = s\wedge t$ for all $s,t \ge 0$.*

**Proof.** Suppose we are given $0 = t_0 < t_1 < \cdots < t_n < \infty$ and recall from Proposition 25.21 that $B_0 = 0$ a.s. and $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^{n}$ are independent mean zero Gaussian random variables. Hence it follows from Corollary 12.26 that $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^{n}$ is a Gaussian random vector. Since the random vector $\left\{B_{t_j}\right\}_{j=0}^{n}$ is a linear transformation $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^{n}$ it follows from Lemma 11.42 that $\left\{B_{t_j}\right\}_{j=0}^{n}$ is a Gaussian random vector. Since $0 = t_0 < t_1 < \cdots < t_n < \infty$ was arbitrary, it follows that $\{B_t\}_{t\ge0}$ is a Gaussian process. Since $B_t = B_t - B_0 \overset{d}{=} N(0,t)$ we see that $\mathbb{E}B_t = 0$ for all $t$. Moreover we have for $0 \le s < t < \infty$ that

$$\begin{aligned}\mathbb{E}[B_tB_s] &= \mathbb{E}[(B_t - B_s + B_s - B_0)(B_s - B_0)]\\ &= \mathbb{E}[(B_t - B_s)(B_s - B_0)] + \mathbb{E}\left[(B_s - B_0)^2\right]\\ &= \mathbb{E}[B_t - B_s]\cdot\mathbb{E}[B_s - B_0] + s = 0\cdot 0 + s\end{aligned}$$

which completes the proof. ■

**Theorem 26.8.** *The function $Q(s,t) := s\wedge t$ defined on $s,t \ge 0$ is positive definite.*

**Proof.** We are going to give a six proofs of this theorem.

1. Choose any independent square integrable random variables, $\{X_j\}_{j=1}^{n}$, such that $\mathbb{E}X_j = 0$ and $\operatorname{Var}(X_j) = t_j - t_{j-1}$. Let $Y_j := X_1 + \cdots + X_j$ for $j = 1,2,\ldots,n$. We then have, for $j \le k$ that
$$\begin{aligned}\operatorname{Cov}(Y_j,Y_k) &= \sum_{m\le j,\; n\le k}\operatorname{Cov}(X_m,X_n) = \sum_{m\le j,\; n\le k}\delta_{m,n}(t_m - t_{m-1})\\ &= \sum_{m\le j}(t_m - t_{m-1}) = t_j,\end{aligned}$$
i.e. $t_j\wedge t_k = \operatorname{Cov}(Y_j,Y_k)$. But such covariance matrices are always positive definite. Indeed,
$$\begin{aligned}\sum_{j,k\le n} t_j\wedge t_k\lambda_j\lambda_k &= \sum_{j,k\le n}\lambda_j\lambda_k\operatorname{Cov}(Y_j,Y_k)\\ &= \operatorname{Var}(\lambda_1Y_1 + \cdots + \lambda_nY_n) \ge 0\end{aligned}$$
with equality holding iff $\lambda_1Y_1 + \cdots + \lambda_nY_n = 0$ from which it follows that
$$0 = \mathbb{E}[Y_j(\lambda_1Y_1 + \cdots + \lambda_nY_n)] = \lambda_j(t_j - t_{j-1}),$$
i.e. $\lambda_j = 0$.
2. According to Exercise 23.13 we can find stochastic processes $\left\{B_n(t) = \sqrt{n}S_{[nt]}\right\}_{n=1}^{\infty}$ such that $\mathbb{E}[B_n(t)\,B_n(s)] \to s\wedge t$ as $n\to\infty$ and therefore
$$\begin{aligned}\sum_{s,t\in\Lambda}(s\wedge t)\,\lambda_s\lambda_t &= \lim_{n\to\infty}\sum_{s,t\in\Lambda}\mathbb{E}[B_n(t)\,B_n(s)]\,\lambda_s\lambda_t\\ &= \lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{t\in\Lambda}B_n(t)\,\lambda_t\right)^2\right] \ge 0.\end{aligned}$$

3. Appealing to Corollary 19.25, there exists a time homogeneous Markov processes $\{B_t\}_{t\geq 0}$ with Markov transition kernels given by
$$Q_t(x,dy) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2} dy. \tag{26.9}$$
It is now easy to see that $s \wedge t = \mathrm{Cov}(B_s, B_t)$ which is automatically non-negative as we saw in the proof of item 2.
4. Let $\Lambda = \{0 < t_1 < \cdots < t_n < \infty\}$ and $\{\lambda_i\}_{i=1}^n \subset \mathbb{R}$ be given. Further let $\alpha_j := \lambda_i + \lambda_{i+1} + \cdots + \lambda_n$ with the convention that $\alpha_{n+1} = 0$. We then have,
$$\begin{aligned}\sum_{i=1}^n t_i \wedge t_j \lambda_i &= \sum_{i=1}^n t_i \wedge t_j (\alpha_i - \alpha_{i+1}) \\ &= \sum_{i=1}^n [t_i \wedge t_j - t_{i-1} \wedge t_j]\,\alpha_i = \sum_{1\leq i\leq j} [t_i - t_{i-1}]\,\alpha_i\end{aligned}$$
where $t_0 := 0$. Hence it follows that
$$\begin{aligned}\sum_{i,j=1}^n t_i \wedge t_j \lambda_i\lambda_j &= \sum_{j=1}^n \sum_{1\leq i\leq j} [t_i - t_{i-1}]\,\alpha_i\lambda_j = \sum_{1\leq i\leq j\leq n} [t_i - t_{i-1}]\,\alpha_i\lambda_j \\ &= \sum_{1\leq i\leq n} [t_i - t_{i-1}]\,\alpha_i^2 \geq 0\end{aligned}$$
with equality iff $\alpha_i = 0$ for all $i$ which is equivalent to $\lambda_i = 0$ for all $i$.
5. Let $h_t(\tau) := t \wedge \tau$ be as after Theorem **??** below and using the results and notation proved there we find,
$$\sum_{s,t\in\Lambda} (s\wedge t)\,\lambda_s\lambda_t = \sum_{s,t\in\Lambda} \langle h_t, h_s\rangle_T\,\lambda_s\lambda_t = \left\|\sum_{t\in\Lambda} \lambda_t h_t\right\|_T^2 \geq 0.$$
This shows $Q$ is positive semi-definite and equality holds iff $\sum_{t\in\Lambda} \lambda_t h_t = 0$. After taking the derivative of this identity, it is not hard to see that $\lambda_t = 0$ for all $t$ so that $Q$ is positive definite.
6. The function $Q(s,t) = s\wedge t$ restricted to $s,t \in [0,T]$ for some $T < \infty$ is the Green's function for the positive definite second order differential operator $-\frac{d^2}{dt^2}$ which is equipped with Dirichlet boundary condition at $t = 0$ and Neumann boundary conditions at $t = T$.

■

We have already given a Markov process proof of the existence of Pre-Brownian motion in Corollary 19.25. Given Theorem 26.8 we can also give a Gaussian process proof of the existence of pre-Brownian motion which we summarize in the next proposition.

**Proposition 26.9 (Pre-Brownian motion).** *Let $\{B_t\}_{t\geq 0}$ be a mean zero Gaussian process such that $\mathrm{Cov}(B_s, B_t) = s\wedge t$ for all $s,t \geq 0$ and let $\mathcal{B}_t := \sigma(B_s : s \leq t)$ and $\mathcal{B}_{t+} := \cap_{\sigma>t}\mathcal{B}_\sigma$ Then;*

1. *$B_0 = 0$ a.s.*
2. *$\{B_t\}_{t\geq 0}$ has independent increments with $B_t - B_s \overset{d}{=} N(0,(t-s))$ for all $0 \leq s < t < \infty$.*
3. *For all $t \geq s \geq 0$, $B_t - B_s$ is independent of $\mathcal{B}_{s+}$.*
4. *$\{B_t\}_{t\geq 0}$ is a time homogeneous Markov process with transition kernels $\{Q_t(x,dy)\}_{t\geq 0}$ given as in Eq. (26.9).*

**Proof.** See Exercise 26.3 – 26.5. ■

**Exercise 26.3 (Independent increments).** Let
$$\mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\}$$
be a partition of $[0,T]$, $\Delta_i B := B_{t_i} - B_{t_{i-1}}$ and $\Delta_i t := t_i - t_{i-1}$. Show $\{\Delta_i B\}_{i=1}^n$ are independent mean zero normal random variables with $\mathrm{Var}(\Delta_i B) = \Delta_i t$.

**Exercise 26.4 (Increments independent of the past).** Let $\mathcal{B}_t := \sigma(B_s : s \leq t)$. For each $s \in (0,\infty)$ and $t > s$, show;

1. $B_t - B_s$ is independent of $\mathcal{B}_s$ and
2. more generally show, $B_t - B_s$ is independent of $\mathcal{B}_{s+} := \cap_{\sigma>s}\mathcal{B}_\sigma$.

**Exercise 26.5 (The simple Markov property).** Show $B_t - B_s$ is independent of $\mathcal{B}_s$ for all $t \geq s$. Use this to show, for any bounded measurable function, $f : \mathbb{R} \to \mathbb{R}$ that
$$\begin{aligned}\mathbb{E}[f(B_t)|\mathcal{B}_{s+}] &= \mathbb{E}[f(B_t)|\mathcal{B}_s] = \mathbb{E}[f(B_t)|B_s] \\ &= (p_{t-s} * f)(B_s) =: \left(e^{(t-s)\Delta/2} f\right)(B_s) \text{ a.s.,}\end{aligned}$$
where
$$p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}$$
so that $p_t * f = Q_t(\cdot, f)$. This problem verifies that $\{B_t\}_{t\geq 0}$ is a "**Markov process**" with transition kernels $\{Q_t\}_{t\geq 0}$ which have $\frac{1}{2}\Delta = \frac{1}{2}\frac{d^2}{dx^2}$ as there "**infinitesimal generator.**"

**Exercise 26.6.** Let
$$\mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\}$$
and $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded measurable function. Show

$$\mathbb{E}\left[f\left(B_{t_1},\dots,B_{t_n}\right)\right]=\int_{\mathbb{R}^n} f\left(x_1,\dots,x_n\right) q_{\mathcal{P}}\left(x\right) dx$$

where

$$q_{\mathcal{P}}\left(x\right):=p_{t_1}\left(x_1\right) p_{t_2-t_1}\left(x_2-x_1\right)\dots p_{t_n-t_{n-1}}\left(x_n-x_{n-1}\right).$$

**Hint:** Either use Exercise 26.3 by writing

$$f\left(x_1,\dots,x_n\right)=g\left(x_1,x_2-x_1,x_3-x_2,\dots,x_n-x_{n-1}\right)$$

for some function, $g$ or use Exercise 26.5 first for functions, $f$ of the form,

$$f\left(x_1,\dots,x_n\right)=\prod_{j=1}^{n}\varphi_j\left(x_j\right).$$

**Better yet**, do it by both methods!

# 27

# Versions and Modifications

We need to introduce a bit of terminology which we will use throughout this part of the book. As before we will let $T$ be an index space which will typically be $\mathbb{R}_+$ or $[0,1]$ in this part of the book. We further sill suppose that $(\Omega, \mathcal{B}, P)$ is a given probability space, $(S, \rho)$ is a separable (for simplicity) metric state space, and $X_t : \Omega \to S$ is a measurable stochastic processes.

**Definition 27.1 (Versions).** *Suppose, $X_t : \Omega \to S$ and $\tilde{X}_t : \Omega \to S$ are two processes defined on $T$. We say that $\tilde{X}$ is a* version *or a* modification *of $X$ provided, for each $t \in T$, $X_t = \tilde{X}_t$ a.s.. (Notice that the null set may depend on the parameter $t$ in the uncountable set, $T$.)*

**Definition 27.2.** *We say two processes are **indistinguishable** iff $P^*(Y_\cdot \neq X_\cdot) = 0$, i.e. iff there is a measurable set, $E \subset \Omega$, such that $P(E) = 0$ and $\{Y_\cdot \neq X_\cdot\} \subset E$ where*

$$\begin{aligned} \{Y_\cdot \neq X_\cdot\} &= \{\omega \in \Omega : Y_t(\omega) \neq X_t(\omega) \text{ for some } t \in [0,\infty)\} \\ &= \cup_{t\in[0,\infty)} \{\omega \in \Omega : Y_t(\omega) \neq X_t(\omega)\}. \end{aligned} \tag{27.1}$$

So $Y$ is a modification of $X$ iff

$$0 = \sup_{t\in T} P\left(X_t \neq \tilde{X}_t\right) = \sup_{t\in T} P\left(\left\{\rho\left(X_t, \tilde{X}_t\right) > 0\right\}\right)$$

while $Y$ is indistinguishable for $X$ iff

$$0 = P^*\left(X_t \neq \tilde{X}_t \ \forall\, t\right) = P^*\left(\left\{\sup_{t\in T} \rho\left(X_t, \tilde{X}_t\right) > 0\right\}\right).$$

Thus the formal difference between the two notions is simply whether the supremum is taken outside or inside the probabilities. See Exercise 30.1 for an example of two processes which are modifications of each other but are not indistinguishable.

**Exercise 27.1.** Suppose $\{Y_t\}_{t\geq 0}$ is a version of a process, $\{X_t\}_{t\geq 0}$. Further suppose that $t \to Y_t(\omega)$ and $t \to X_t(\omega)$ are both right continuous everywhere. Show $E := \{Y_\cdot \neq X_\cdot\}$ is a measurable set such that $P(E) = 0$ and hence $X$ and $Y$ are indistinguishable. **Hint:** replace the union in Eq. (27.1) by an appropriate countable union.

**Exercise 27.2.** Suppose that $\{X_t : \Omega \to S\}_{t\geq 0}$ is a process such that for each $N \in \mathbb{N}$ there is a right continuous modification, $\left\{\tilde{X}_t^{(N)}\right\}_{0\leq t<N}$ of $\{X_t\}_{0\leq t<N}$. Show that $X$ admits a right continuous modifications, $\tilde{X}$, defined for all $t \geq 0$.

## 27.1 Kolmolgorov's Continuity Criteria

Let $\mathbb{D}_n := \left\{\frac{i}{2^n} : i \in \mathbb{Z}\right\}$ and $\mathbb{D} := \cup_{n=0}^\infty \mathbb{D}_n$ be the dyadic rational numbers.

**Lemma 27.3.** *Let $\mathbb{D}_+ = \mathbb{D} \cap [0,\infty)$ and $s \in \mathbb{D}_+ = \mathbb{D} \cap [0,\infty)$ and $n \in \mathbb{N}_0$ be given, then;*

1. *there exists a unique $i = i(n,s) \in \mathbb{N}_0$ such that $i2^{-n} \leq s < (i+1)\,2^{-n}$ and*
2. *$s$ may be uniquely written as*

$$s = \frac{i}{2^n} + \sum_{k=1}^\infty \frac{a_k}{2^{n+k}},$$

*where $a_k = a_k(n,s) \in \{0,1\}$ with $a_k = 0$ for all sufficiently large $k$.*

*Example 27.4.* Suppose that $s = 85/32 = 85/2^5 \in \mathbb{D}$ and $2 \in \mathbb{N}_0$ are given, then $2^2 s = 85/8 = 10 + 5/8$, i.e.

$$s = \frac{10}{2^2} + \frac{5}{2^5}.$$

Similarly, $2^3 \cdot 5/2^5 = 5/4 = 1 + 1/4$ so that $5/2^5 = 1/2^3 + 1/2^5$ and we have expressed $s$ as

$$s = \frac{10}{2^2} + \frac{1}{2^3} + \frac{0}{2^4} + \frac{1}{2^5}.$$

**Proof.** The first assertion follows from the fact that $\mathbb{D}_+$ is partitioned by $[\{i2^{-n}, (i+1)\,2^{-n}) \cap \mathbb{D}\}_{i\in\mathbb{N}_0}$. For the second assertion define the $a_1 = 1$ if $\frac{i}{2^n} + \frac{a_1}{2^n} \leq s$ and 0 otherwise, then choose $a_2 = 1$ if $\frac{i}{2^n} + \frac{a_1}{2^{n+1}} + \frac{a_2}{2^{n+2}} \leq s$ and 0 otherwise, etc. It is easy to check that $s_m := \frac{i}{2^n} + \sum_{k=1}^m \frac{a_k}{2^{n+k}}$ so constructed satisfies, $s - \frac{1}{2^{n+m}} < s_m \leq s$ for all $m \in \mathbb{N}$. As $s_m \in \mathbb{D}_{m+n}$ and $s \in \mathbb{D}_N$ for some $N$, if $m + n \geq N$ then we must have $s_m = s$ because $s - s_m < \frac{1}{2^{n+m}}$ and $s, s_m \in \mathbb{D}_{m+n}$.

■

Suppose now that $(S,\rho)$ is a metric space and $x : Q := \mathbb{D} \cap [0,1] \to S$. For $n \in \mathbb{N}_0$ let

$$\begin{aligned}\Delta_n(x) &= \max\left\{\rho\left(x\left(i2^{-n}\right), x\left((i-1)2^{-n}\right)\right) : 1 \le i \le 2^n\right\} \\ &= \max\left\{\rho(x(t), x(s)) : s,t \in Q \cap \mathbb{D}_n \text{ with } |s-t| \le \frac{1}{2^n}\right\}.\end{aligned}$$

If $\gamma \in (0,1)$ and $x : \mathbb{D}_1 := \mathbb{D} \cap [0,1] \to S$ is a $\gamma -$ **Hölder continuous** function, i.e.

$$\rho(x(t), x(s)) \le K|t-s|^\gamma$$

for some and $K < \infty$, then $\Delta_n(x) \le K2^{-n\gamma}$ for all $n$ and in particular for all $\alpha \in (0,\gamma)$ we have

$$\sum_{n=0}^\infty 2^{n\alpha}\Delta_n(x) \le K\sum_{n=0}^\infty 2^{n\alpha}2^{-n\gamma} = K\left(1-\frac{1}{2^{\gamma-\alpha}}\right)^{-1} < \infty.$$

Our next goal is to produce the following "converse" to this statement.

**Lemma 27.5.** *Suppose $\alpha > 0$ and $x : Q := \mathbb{D} \cap [0,1] \to S$ is a map such that $\sum_{n=0}^\infty 2^{n\alpha}\Delta_n(x) < \infty$, then*

$$\rho(x(t), x(s)) \le 2^{1+\alpha} \cdot \left[\sum_{k=0}^\infty 2^{\alpha k}\Delta_k(x)\right] \cdot |t-s|^\alpha \text{ for all } s,t \in Q. \tag{27.2}$$

*Moreover there exists a unique continuous function $\tilde{x} : [0,1] \to S$ extending $x$ and this extension is still $\alpha$ – Hölder continuous. (When $\alpha > 1$ it follows by Exercise 27.3 that $x(t)$ is constant.)*

**Proof.** Let $s,t \in Q$ with $s < t$ and choose $n$ so that

$$\frac{1}{2^{n+1}} < t-s \le \frac{1}{2^n} \tag{27.3}$$

and observe that if $s \in [i2^{-n}, (i+1)2^{-n})$, then $t \in [i2^{-n}, (i+2)2^{-n})$ and therefore

$$s = \frac{i}{2^n} + \sum_{k=1}^\infty \frac{a_k}{2^{n+k}} \text{ and } t = \frac{j}{2^n} + \sum_{k=1}^\infty \frac{b_k}{2^{n+k}}$$

where $j \in \{i, i+1\}$ and $a_k, b_k \in \{0,1\}$ with $a_k = b_k = 0$ for a.a. $k$. Letting

$$s_m := \frac{i}{2^n} + \sum_{k=1}^m \frac{a_k}{2^{n+k}}$$

as above we have $s_N = s$ for large $N$. Since $s_m, s_{m+1} \in Q \cap \mathbb{D}_{n+m+1}$ with $|s_m - s_{m+1}| \le 2^{-(n+m+1)}$, it follows from the definition of $\Delta_{n+m+1}$ that $\rho(x(s_{m+1}), x(s_m)) \le \Delta_{n+m+1}(x)$ which combined with the triangle inequality shows

$$\rho(x(s), x(s_0)) \le \sum_{m=1}^N \rho(x(s_m), x(s_{m-1})) \le \sum_{m=1}^\infty \Delta_{n+m}(x).$$

Similarly $\rho(x(t), x(t_0)) \le \sum_{m=1}^\infty \Delta_{n+m}(x)$ while $\rho(x(s_0), x(t_0)) \le \Delta_n(x)$. One more application of the triangle inequality now shows,

$$\begin{aligned}\rho(x(t), x(s)) &\le \Delta_n(x) + 2\cdot\sum_{m=1}^\infty \Delta_{n+m}(x) \\ &\le 2\cdot\sum_{k=n}^\infty \Delta_k(x) = 2\cdot\sum_{k=n}^\infty 2^{-\alpha k}\cdot 2^{\alpha k}\Delta_k(x) \\ &\le 2\cdot\left(2^{-n}\right)^\alpha \sum_{k=n}^\infty 2^{\alpha k}\Delta_k(x).\end{aligned}$$

Combining this with the lower bound in Eq. (27.3) in the form

$$\left(2^{-n}\right)^\alpha = 2^\alpha\left(2^{-(n+1)}\right)^\alpha < 2^\alpha(t-s)^\alpha,$$

gives the estimate in Eq. (27.2).

For the last assertion we define $\tilde{x}(t) := \lim_{Q \ni s \to t} x(s)$. This limit exists since for any sequence $\{s_n\}_{n=1}^\infty \subset Q$ with $s_n \to t \in [0,1]$, the sequence $\{x(s_n)\}_{n=1}^\infty$ is Cauchy in $S$ because of Eq. (27.2) and hence convergent in $S$. It is easy to check that $\lim_{n\to\infty} x(s_n)$ is independent of the choice of the sequence $\{s_n\}_{n=1}^\infty$. A simple limiting argument now shows that

$$\rho(\tilde{x}(t), \tilde{x}(s)) \le 2^{1+\alpha}\cdot\sum_{k=0}^\infty 2^{\alpha k}\Delta_k(x)\cdot|t-s|^\alpha \text{ for all } s,t \in [0,1]$$

which shows that $\tilde{x}$ is Hölder continuous. As we had no choice but to define $\tilde{x}$ the way we did if $\tilde{x}$ is to be continuous, the extension is unique. ■

**Exercise 27.3.** Show; if $x : Q \to S$ is $\alpha -$ Hölder continuous for some $\alpha > 1$, then $x$ is constant.

**Notation 27.6** *For $0 < \alpha < 1$ and $x : Q \to S$, let*

$$K_\alpha(x) := 2^{(1+\alpha)}\sum_{k=0}^\infty 2^{\alpha k}\Delta_k(x). \tag{27.4}$$

**Theorem 27.7 (Kolmogorov's Continuity Criteria).** *Let $\{X_t\}_{t\in Q}$ be an $S$ – valued stochastic process and suppose there exists $\gamma, \varepsilon > 0$ such that*

$$\mathbb{E}\left[\rho\left(X_t, X_s\right)^\gamma\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in Q. \tag{27.5}$$

*Then for all $\alpha \in (0, \varepsilon/\gamma)$,*

$$\rho\left(X_t, X_s\right) \leq K_\alpha\left(X\right)\left|t-s\right|^\alpha \text{ for all } s,t \in Q \tag{27.6}$$

*where $K_\alpha(X)$ is as in Eq. (??) satisfies $K_\alpha(X) \in L^\gamma(P)$ and in fact,*

$$\left\|K_\alpha\left(X\right)\right\|_\gamma \leq \frac{C^{1/\gamma}\cdot 2^{(1+\alpha)}}{1-2^{\alpha-\varepsilon/\gamma}} < \infty. \tag{27.7}$$

**Proof.** According to Exercise 27.4 below when $\gamma < 1$ or more generally when $\gamma < 1+\varepsilon$ it actually follows that $X_t = X_0$ a.s. and therefore Eq. (27.6) holds for some $K_\alpha$ which is equal to zero almost surely.

So we may now suppose that $\gamma \geq 1+\varepsilon \geq 1$ and let $\alpha \in (0, \varepsilon/\gamma)$ and $K_\alpha(\cdot)$ is defined as in Eq. (27.4). The estimate in Eq. (27.6) is now a consequence of Lemma 27.5. So it only remains to verify Eq. (27.7). From the following simple estimate,

$$\Delta_k\left(X\right)^\gamma = \max_{1\leq i\leq 2^k} \rho\left(X_{i2^{-k}}, X_{(i-1)2^{-k}}\right)^\gamma \leq \sum_{i=1}^{2^k} \rho\left(X_{i2^{-k}}, X_{(i-1)2^{-k}}\right)^\gamma,$$

we find

$$\mathbb{E}\left[\Delta_k\left(X\right)^\gamma\right] \leq \sum_{i=1}^{2^k} \mathbb{E}\left[\rho\left(X_{i2^{-k}}, X_{(i-1)2^{-k}}\right)^\gamma\right] \leq 2^k \cdot C\left(2^{-k}\right)^{1+\varepsilon} = C2^{-k\varepsilon} \tag{27.8}$$

and therefore $\left\|\Delta_k(X)\right\|_\gamma \leq C^{1/\gamma}2^{-k\varepsilon/\gamma}$. Combining this inequality with Minkowski's inequality shows

$$\begin{aligned}\left\|K_\alpha\left(X\right)\right\|_\gamma &= 2^{(1+\alpha)}\left\|\sum_{k=0}^\infty\left[2^{\alpha k}\Delta_k\left(X\right)\right]\right\|_\gamma \leq 2^{(1+\alpha)}\sum_{k=0}^\infty 2^{\alpha k}\left\|\Delta_k\left(X\right)\right\|_\gamma \\ &\leq 2^{(1+\alpha)}C^{1/\gamma}\sum_{k=0}^\infty\left(2^{\alpha-\varepsilon/\gamma}\right)^k = \frac{C^{1/\gamma}2^{(1+\alpha)}}{1-2^{\alpha-\varepsilon/\gamma}} < \infty\end{aligned}$$

provided $\alpha < \varepsilon/\gamma$. ■

**Theorem 27.8 (Kolmogorov's Continuity Criteria).** *Let $T \in \mathbb{N}$, $D = [0,T] \subset \mathbb{R}$, $(S, \rho)$ be a complete separable metric space and suppose that $X_t : \Omega \to S$ is a process for $t \in D$. Assume there exists $\gamma, C, \varepsilon > 0$ such that such that*

$$\mathbb{E}\left[\rho\left(X_t, X_s\right)^\gamma\right] \leq C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in D. \tag{27.9}$$

*Then there is a modification, $\tilde{X}$, of $X$ which is $\alpha$–Hölder continuous for all $\alpha \in (0, \varepsilon/\gamma)$ and for each such $\alpha$ there is a random variable $K_\alpha(X) \in L^\gamma(P)$ such that*

$$\rho(\tilde{X}_t, \tilde{X}_s) \leq K_\alpha\left(X\right)\left|t-s\right|^\alpha \text{ for all } s,t \in D. \tag{27.10}$$

*(Again according to Exercise 27.4, we will have $X_t = X_0$ a.s. for all $t \in D$ unless $\gamma \geq 1+\varepsilon$.)*

**Proof.** From Theorem 27.7 we know for all $\alpha \in (0, \varepsilon/\gamma)$ there is a random variable $K_\alpha(X) \in L^\gamma(P)$ such that

$$\rho(X_t, X_s) \leq K_\alpha\left(X\right)\left|t-s\right|^\alpha \text{ for all } s,t \in D\cap\mathbb{D}.$$

On the set $\{K_\alpha(X) < \infty\}$, $\{X_t\}_{t\in D\cap\mathbb{D}}$ has a unique continuous extension to $D$ which we denote by $\left\{\tilde{X}_t\right\}_{t\in D}$. Moroever this extenson is easily seen to satisfy Eq. (27.10). Lastly we have for $s \in D\cap\mathbb{D}$ and $t \in D$ that

$$\begin{aligned}\rho\left(X_t, \tilde{X}_t\right)^\gamma &\leq \liminf_{D\cap\mathbb{D}\ni s\to t}\left[\rho\left(X_t, X_s\right) + \rho\left(X_s, \tilde{X}_t\right)\right]^\gamma \\ &= \liminf_{D\cap\mathbb{D}\ni s\to t}\rho\left(X_t, X_s\right)^\gamma\end{aligned}$$

and so by Fatou's lemma,

$$\mathbb{E}\left[\rho\left(X_t, \tilde{X}_t\right)^\gamma\right] \leq \liminf_{D\cap\mathbb{D}\ni s\to t}\mathbb{E}\left[\rho\left(X_t, X_s\right)^\gamma\right] \leq \liminf_{D\cap\mathbb{D}\ni s\to t} C\left|t-s\right|^{1+\varepsilon} = 0.$$

This certainly implies that $\rho\left(X_t, \tilde{X}_t\right) = 0$ a.s. for every $t \in D$ and therefore that $\tilde{X}$ is a modification of $X$. ■

Our construction of Brownian motion in Theorem 28.3 below will give us an opportunity to apply Theorem 27.8. At this time let us observe that it is important that $\varepsilon$ is greater than 0 in the previous two theorems.

*Example 27.9.* Recall that a Poisson process, $\{N_t\}_{t\geq 0}$, with parameter $\lambda$ satisfies (by definition): (i) $N$ has *independent increments*, and (ii) if $0 \leq u < v$ then $N_v - N_u$ has the Poisson distribution with parameter $\lambda(v-u)$. Using the generating function (or the Laplace or Fourier transform, see Example 24.11), one can show that for any $k \in \mathbb{N}$, that

$$\mathbb{E}\left|N_t - N_s\right|^k \sim \lambda\left|t-s\right| \text{ for } \left|t-s\right| \text{ small.} \tag{27.11}$$

Notice that we can not use Eq. (27.11) for any $k \in \mathbb{N}$ to satisfy the hypothesis of Theorem 27.8 which is good since $\{N_t\}_{t\geq 0}$ is integer value and does not have a continuous modification. However, see Example 30.27 below where it is shown that $\{N_t\}_{t\geq 0}$ has a right continuous modification.

**Exercise 27.4.** Let $T \in \mathbb{N}$, $D = [0,T] \subset \mathbb{R}$, $(S,\rho)$ be a complete separable metric space and suppose that $X_t : \Omega \to S$ is a process for $t \in D$. Assume there exists $\gamma > 0$, $C > 0$, and $\varepsilon > 0$ such that $(1+\varepsilon)/\gamma > 1$ and

$$\mathbb{E}\left[\rho\left(X_t, X_s\right)^\gamma\right] \le C\left|t-s\right|^{1+\varepsilon} \text{ for all } s,t \in D.$$

Show $X_t = X_0$ a.s. for each $t \in D$. **Hint:** for $\gamma \in (0,1)$ use the inequality[1] $(a+b)^\gamma \le a^\gamma + b^\gamma$ for all $a,b \ge 0$ while for $\gamma \ge 1$ use Minkowski's inequality. (This Exercise was inspired by questions posed by Dennis Leung.)

## 27.2 Kolmolgorov's Tightness Criteria

Before leaving this chapter let us record a couple of results pertaining to the weak convergence of continuous processes. In this section let us suppose that $(S,\rho)$ is a complete metric space satisfying the Heine–Borel property[2] (for example $S = \mathbb{R}^d$ for some $d < \infty$). As usual we let $C([0,1],S)$ denote the continuous functions from $[0,1]$ into $S$. We make $C([0,1],S)$ into a metric space by defininig

$$\rho_\infty(x,y) := \max_{0\le t\le 1} \rho\left(x(t), y(t)\right) \ \forall\ x,y \in C([0,1],S). \tag{27.12}$$

It is standard and not to hard to verify that $(C([0,1],S),\rho_\infty)$ is also a complete separable metric space. Moreover the compact subset, $\mathcal{K}$, of $(C([0,1],S),\rho_\infty)$ are precisely those sets which are closed, uniformly bounded, and equicontinous, see Section **??** for more information in this regard. For our purposes here the main thing to notice is that for each $0 < \alpha < 1$ and $C < \infty$, the set

$$\mathcal{K}(\alpha,C) := \left\{x \in C([0,1],S) : \rho\left(x(t), x(s)\right) \le K\left|t-s\right|^\alpha \ \forall\ s,t \in [0,1]\right\}$$

is a closed, uniformly bounded, and equicontinous, subset of $C([0,1],S)$ and hence is compact.

**Exercise 27.5.** Show $(C([0,1],S),\rho_\infty)$ is separable. **Hints:**

1. Choose a countable dense subset, $\Lambda$, of $S$ and then choose finite subset $\Lambda_n \subset \Lambda$ such that $\Lambda_n \uparrow \Lambda$.
2. Let $\mathbb{D}_n := \left\{\frac{k}{2^n} : 0 \le k \le 2^n\right\}$ and $\mathbb{D} = \cup_{n=0}^\infty \mathbb{D}_n$. Further let $\mathbb{F}_n := \left\{x : [0,1] \to \Lambda_n\right\}$ such that $x|_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]}$ is constant for all $1 \le k \le 2^n$ and further suppose that $x|_{[0,2^{-n}]}$ is constant.
3. Given $y \in C([0,1],S)$ and $\varepsilon > 0$, show there exists $n \in \mathbb{N}$ and an $x \in \mathbb{F}_n$ such that $\rho_\infty(y,x) \le \varepsilon$.
4. For $k,n \in \mathbb{N}$ let
$$\mathcal{F}_n^k := \left\{y \in C([0,1],S) : \min_{x\in\mathbb{F}_n} \rho_\infty(y,x) \le \frac{1}{k}\right\}$$
and let $\Gamma := \left\{(k,n) \in \mathbb{N}^2 : \mathcal{F}_n^k \ne \emptyset\right\}$. For each $(k,n) \in \Gamma$, choose a function, $y_{k,n} \in \mathcal{F}_n^k$.
5. Now show that $\{y_{k,n} : (k,n) \in \Gamma\}$ is a countable dense subset of $(C([0,1],S),\rho_\infty)$.

**Theorem 27.10 (Tightness Criteria).** *Let $S$ be a complete metric space satisfying the Heine–Borel property. Suppose that $\{B_n(t) : 0 \le t \le 1\}_{n=1}^\infty$ is a sequence of $S$ – valued continuous stochastic processes and suppose there exists $\gamma, \varepsilon > 0$ and $C < \infty$ such that*

$$\sup_n \mathbb{E}\left[\rho\left(B_n(t), B_n(s)\right)^\gamma\right] \le C\left|t-s\right|^{1+\varepsilon} \text{ for all } 0 \le s,t \le 1 \tag{27.13}$$

*and for some point $s_0 \in S$ we have*

$$\lim_{N\uparrow\infty} \sup_n P\left[\rho\left(B_n(0), s_0\right) > N\right] = 0. \tag{27.14}$$

*Then the collection of measures, $\{\mu_n := \mathrm{Law}_P(B_n)\}_{n=1}^\infty$ on $C([0,1],S)$ are tight.*

**Proof.** Let $\alpha \in (0,\varepsilon/\gamma)$ and for $\omega \in C([0,1],S)$ let

$$K_\alpha(\omega) = 2^{-(1+\alpha)\gamma} \sum_{k=0}^\infty 2^{\alpha k} \Delta_k(\omega)$$

where

$$\Delta_k(\omega) := \max\left\{\rho\left(\omega\left(\frac{j-1}{2^k}\right), \omega\left(\frac{j}{2^k}\right)\right) : 1 \le j \le 2^k\right\}.$$

The assumptions of this theorem allows us to apply Theorem 27.7 in order to learn;

$$\sup_n \mathbb{E}\left[K_\alpha(B_n)^\gamma\right] \le M(C,\gamma,\varepsilon,\alpha) < \infty.$$

---

[1] If $f(x) = x^\gamma$ for $\gamma \in (0,1)$, then $f$ is an increasing function which is concave down, i.e. $f'$ is decreasing. For $a > 0$ let $g(x) := f(x+a) - f(x)$. By looking at a picture or just noting that $g'(x) \le 0$ since $f'$ is decreasing, it follows that $g$ is a decreasing function of $x$. In particular it follows that

$$f(b+a) - f(b) = g(b) \le g(0) = f(a) - f(0) = f(a).$$

[2] The Heine–Borel property means that closed and bounded sets are compact.

Now let $\Omega_N$ denote those $\omega \in C([0,1], S)$ such that $\rho(\omega(0), s_0) \leq N$ and $K_\alpha(\omega) \leq N$. We then have that

$$\begin{aligned}
\mu_n(\Omega_N^c) &= P(B_n \notin \Omega_N) \\
&= P(\rho(B_n(0), s_0) > N \text{ or } K_\alpha(B_n) > N) \\
&\leq P(\rho(B_n(0), s_0) > N) + P(K_\alpha(B_n) > N) \\
&\leq P[\rho(B_n(0), s_0) > N] + \frac{1}{N^\gamma}\mathbb{E}K_\alpha(B_n)^\gamma \\
&\leq P[\rho(B_n(0), s_0) > N] + \frac{1}{N^\gamma}M(C, \gamma, \varepsilon, \alpha).
\end{aligned}$$

From this inequality and the hypothesis of the theorem it follows that $\lim_{N\to\infty} \sup_n \mu_n(\Omega_N^c) = 0$. To complete the proof it suffices to observe that for $\omega \in \Omega_N$ we have $\rho(\omega(0), s_0) \leq N$ and

$$\rho(\omega(t), \omega(s)) \leq K_\alpha(\omega)|t-s|^\alpha \leq N|t-s|^\alpha \ \forall\ 0 \leq s, t \leq 1.$$

Therefore by the Arzela - Ascoli Theorem **??** and Remark **??**, it follows that $\Omega_N$ is precompact inside of the complete separable metric space, $C([0,1], S)$. ■

This theorem is often useful for checking the tightness hypothesis of the next theorem.

**Theorem 27.11 (Weak Convergence Theorem).** *Keeping the notation above and further assume $\{B_0(t) : 0 \leq t \leq 1\}$ is another $S$ – valued continuous process. Then $B_n \implies B_0$ iff $\{B_n\}_{n=1}^\infty$ is tight and $B_n \overset{f.d.}{\implies} B_0$, i.e. $B_n$ converges to $B_0$ in the sense of finite dimensional distributions.*

**Proof.** If $B_n \implies B_0$ then $\{B_n\}_{n=1}^\infty$ is tight by Prokhorov's Theorem 23.89. Moreover if $f : S^k \to \mathbb{R}$ is a bounded continuous function and $0 \leq t_1 < t_2 < \dots < t_k \leq 1$, then

$$F(x) := f(x(t_1), \dots, x(t_k)) \text{ for } x \in C([0,1], S) \tag{27.15}$$

defines a bounded continuous (**cylinder**) function on $C([0,1], S)$. Therefore by the definition of weak convergence it follows that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[f(B_n(t_1), \dots, B_n(t_k))] &= \lim_{n\to\infty} \mathbb{E}[F(B_n(\cdot))] = \mathbb{E}[F(B_0(\cdot))] \\
&= \mathbb{E}[f(B_0(t_1), \dots, B_0(t_k))]
\end{aligned}$$

and we have shown $B_n \overset{\text{f.d.}}{\implies} B_0$.

For the converse we now suppose that $\{B_n\}_{n=1}^\infty$ is tight and $B_n \overset{\text{f.d.}}{\implies} B_0$ and for the sake of contradiction assume that $B_n$ does not converge weakly to $B_0$. This means there exists an $\varepsilon > 0$ and a bounded continuous function $F : C([0,1], S) \to \mathbb{R}$ such that $|\mathbb{E}[F(B_n)] - \mathbb{E}F(B_0)| \geq \varepsilon$ for infinitely many $n$. Therefore there is a subsequence $B_k' = B_{n_k}$ such that

$$|\mathbb{E}F(B_k') - \mathbb{E}F(B_0)| \geq \varepsilon > 0 \text{ for all } k \in \mathbb{N}. \tag{27.16}$$

Moreover by the assumed tightness, Prokhorov's Theorem 23.89 allows us to pass to a further subsequence (still denoted by $B_k'$) if necessary so that $B_k' \implies X$ for some continuous process in $S$, i.e. for a $C([0,1], S)$ – random variable $X$. Passing to the limit as $k \to \infty$ in Eq. (27.16) then implies that

$$|\mathbb{E}F(X) - \mathbb{E}F(B_0)| \geq \varepsilon > 0. \tag{27.17}$$

On the other hand $B_n \overset{\text{f.d.}}{\implies} B_0$ as $n \to \infty$ and therefore $X$ and $B_0$ are continuous processes on $[0,1]$ with the same finite dimensional distributions and hence are indistinguishable by Exercise 27.1. However, this then implies $|\mathbb{E}F(X) - \mathbb{E}F(B_0)| = 0$ which contradicts Eq. (27.17).

■

# Brownian Motion I

Our next goal is to prove existence of Brownian motion and then describe some of its basic path properties.

**Definition 28.1 (Brownian Motion).** *A* ***Brownian motion*** $\{B_t\}_{t\geq 0}$ *is an adapted mean zero Gaussian random process on some filtered probability space,* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, *satisfying; 1) for each* $\omega \in \Omega$, $t \to B_t(\omega)$ *is continuous, and 2)*

$$\mathbb{E}[B_t B_s] = t \wedge s \text{ for all } s, t \geq 0. \tag{28.1}$$

*So a Brownian motion is a pre-Brownian motion with continuous sample paths.*

*Remark 28.2.* If no filtration is given, we can use the process to construct one. Namely, let $\mathcal{B}_t^0 := \sigma(B_s : s \leq t)$ and replace $\mathcal{B}$ by $\vee \mathcal{B}_t^0$ if necessary. We call $\{\mathcal{B}_t^0\}$ the raw filtration associated to $\{B_t\}$.

**Theorem 28.3 (Wiener 1923).** *Brownian motions exists. Moreover for any* $\alpha \in (0, 1/2)$, $t \to B_t$ *is locally* $\alpha$ *– Hölder continuous almost surely.*

**Proof.** For $0 \leq s < t < \infty$, $\tilde{B}_t - \tilde{B}_s$ is a mean zero Gaussian random variable with

$$\mathbb{E}\left[\left(\tilde{B}_t - \tilde{B}_s\right)^2\right] = \mathbb{E}\left[\tilde{B}_t^2 + \tilde{B}_s^2 - 2B_s\tilde{B}_t\right] = t + s - 2s = t - s.$$

Hence if $N$ is a standard normal random variable, then $\tilde{B}_t - \tilde{B}_s \overset{d}{=} \sqrt{t-s}N$ and therefore, for any $p \in [1, \infty)$,

$$\mathbb{E}\left|\tilde{B}_t - \tilde{B}_s\right|^p = (t-s)^{p/2}\,\mathbb{E}|N|^p. \tag{28.2}$$

Hence an application of Theorem 27.8 shows, with $\varepsilon = p > 2$, $\beta = p/2 - 1$, $\alpha \in \left(0, \frac{p/2-1}{p}\right) = \left(0, \frac{1}{2} - 1/p\right)$, there exists a modification, $B$ of $\tilde{B}$ such that

$$|B_t - B_s| \leq C_{\alpha,T}\,|t-s|^\alpha \text{ for } s, t \in [0, T).$$

By applying this result with $T = N \in \mathbb{N}$, we find there exists a continuous version, $B$, of $\tilde{B}$ for all $t \in [0, \infty)$ and this version is locally Hölder continuous with Hölder constant $\alpha < 1/2$. ■

For the rest of this chapter we will assume that $\{B_t\}_{t\geq 0}$ is a Brownian motion on some probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ and $\mathcal{B}_t := \sigma(B_s : s \leq t)$.

## 28.1 Donsker's Invariance Principle

In this section we will see that Brownian motion may be thought of as a limit of random walks – this is the content of Donsker's invariance principle or the so called functional central limit theorem. The setup is to start with a random walk, $S_n := X_1 + \cdots + X_n$ where $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with zero mean and variance one. We then define for each $n \in \mathbb{N}$ the following continuous process,

$$B_n(t) := \frac{1}{\sqrt{n}}\left(S_{[nt]} + (nt - [nt])\,X_{[nt]+1}\right) \tag{28.3}$$

where for $\tau \in \mathbb{R}_+$, $[\tau]$ is the integer part of $\tau$, i.e. the nearest integer to $\tau$ which is no greater than $\tau$. The first step in this program is to prove convergence in the sense of finite dimensional distributions.

**Proposition 28.4.** *Let* $B$ *be a standard Brownian motion, then* $B_n \overset{f.d.}{\Longrightarrow} B$.

**Proof.** In Exercise 23.13 you showed that $\left\{\frac{1}{\sqrt{n}}S_{[nt]}\right\}_{t\geq 0} \overset{\text{f.d.}}{\Longrightarrow} B$ as $n \to \infty$. Suppose that $0 < t_1 < t_2 < \cdots < t_k < \infty$ are given and we let

$$W_n := \frac{1}{\sqrt{n}}\left(S_{[nt_1]}, \dots, S_{[nt_k]}\right),$$
$$Y_n := (B_n(t_1), \dots, B_n(t_k))$$

and $\varepsilon_n := Y_n - W_n \in \mathbb{R}^k$. From Eq. (28.3) and Chebyshev's inequality (for all $\delta > 0$),

$$P\left(\left|(\varepsilon_n)_i\right| > \delta\right) \leq \frac{1}{\delta}\mathbb{E}\left[\frac{1}{\sqrt{n}}(nt_i - [nt_i])\left|X_{[nt]+1}\right|\right] \leq \frac{1}{\sqrt{n}}\frac{\mathbb{E}|X_1|}{\delta} \to 0$$

as $n \to \infty$. This then easily implies that $\varepsilon_n \overset{P}{\to} 0$ as $n \to \infty$ and therefore by Slutzky's Theorem 23.43 it follows that $Y_n = X_n + \varepsilon_n \implies (B(t_1), \dots, B(t_k))$. ■

Let $\Omega := C([0, \infty), \mathbb{R})$ which becomes a complete metric space in the metric defined by,

$$\rho(\omega_1, \omega_2) := \sum_{n=1}^\infty \frac{1}{2^n}\max_{0\leq t\leq n}|\omega_1(t) - \omega_2(t)| \wedge 1.$$

**Theorem 28.5 (Donsker's invariance principle).** *Let $\Omega$, $B_n$, and $B$ be as above. Then for $B_n \implies B$, i.e.*

$$\lim_{n\to\infty} \mathbb{E}\left[F\left(B_n\right)\right] = \mathbb{E}\left[F\left(B\right)\right]$$

*for all bounded continuous functions $F : \Omega \to \mathbb{R}$.*

One method of proof, see [27] and [42], goes by the following two steps. 1) Show that the finite dimensional distributions of $B_n$ converge to a $B$ as was done in Proposition 28.4. 2) Show the distributions of $B_n$ are tight. A proof of the tightness may be based on Ottaviani's maximal inequality in Corollary 22.52, [26, Corollary 16.7]. Another possible proof or this theorem is based on "Skorokhod's representation," see (for example) [26, Theorem 14.9 on p. 275] or [35]. Rather than give the full proof here I will give the proof of a slightly weaker version of the theorem under the more stringent restriction that the $X_n$ posses fourth moments.

**Proposition 28.6 (Random walk approximation bounds).** *Suppose that $\{X_n\}_{n=1}^\infty \subset L^4(P)$ are i.i.d. random variables with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$ and $\gamma := \mathbb{E}X_n^4 < \infty$. Then there exists $C < \infty$ such that*

$$\mathbb{E}\left|B_n(t) - B_n(s)\right|^4 \le C\left|t-s\right|^2 \text{ for all } s,t \in \mathbb{R}_+. \tag{28.4}$$

**Exercise 28.1.** Provide a proof of Proposition 28.6. **Hints:** Use the results of Exercise 12.6, namely that

$$\mathbb{E}\left|S_l\right|^4 = l\gamma + 3l(l-1), \tag{28.5}$$

to verify that Eq. (28.4) holds for $s,t \in D_n := \frac{1}{n}\mathbb{N}_0$. Take care of the case where $s,t \ge 0$ with $|t-s| < 1/n$ by hand and finish up using these results along with Minkowski's inequality.

**Theorem 28.7 (Baby Donsker Theorem).** *Continuing the notation used in Proposition 28.6 and let $T < \infty$ and $B$ be a Brownian motion. Then*

$$B_n|_{[0,T]} \implies B|_{[0,T]}, \tag{28.6}$$

*i.e.* $\text{Law}\left(B_n|_{[0,T]}\right) \implies \text{Law}\left(B|_{[0,T]}\right)$ *as distributions on $\Omega_T = C([0,T],\mathbb{R})$ – a complete separable metric space. (See the simulation file **Random Walks to BM.xls.)***

**Proof.** If Eq. (28.6) fails to hold there would exists $g \in BC(\Omega_T)$ and a subsequence $B_k' = B_{n_k}$ such that

$$\varepsilon := \inf_k \left|\mathbb{E}\left[g\left(B_k'|_{[0,T]}\right)\right] - \mathbb{E}g\left(B|_{[0,T]}\right)\right| > 0. \tag{28.7}$$

Since $B_n(0) = 0$ for all $n$ and the estimate in Eq. (28.4) of Proposition 28.6 holds, it follows from Theorem 27.10 that $\left\{B_n|_{[0,T]}\right\}_{n=1}^\infty$ is tight. So by Prokhorov's Theorem 23.89, there is a further subsequence $B_l'' = B_{k_l}'$ which is weakly convergent to some $\Omega_T$ – valued process $X$. Replacing $B_k'$ bit $B_l''$ in Eq. (28.7) and then letting $l \to \infty$ in the resulting equation shows

$$\left|\mathbb{E}\left[g\left(X|_{[0,T]}\right)\right] - \mathbb{E}g\left(B|_{[0,T]}\right)\right| \ge \varepsilon > 0. \tag{28.8}$$

On the other hand by Proposition 28.4 we know that $B_n \overset{\text{f.d.}}{\implies} B$ as $n \to \infty$ and therefore $X$ and $B$ are continuous processes on $[0,T]$ with the same finite dimensional distributions and hence are indistinguishable by Exercise 27.1. However this is in contradiction to Eq. (28.8). ∎

## 28.2 Path Regularity Properties of BM

**Definition 28.8.** *Let $(V, \|\cdot\|)$ be a normed space and $Z \in C([0,T],V)$. For $1 \le p < \infty$, the **p - variation** of $Z$ is;*

$$v_p(Z) := \sup_\Pi \left(\sum_{j=1}^n \left\|Z_{t_j} - Z_{t_{j-1}}\right\|^p\right)^{1/p}$$

*where the supremum is taken over all partitions, $\Pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$, of $[0,T]$.*

**Lemma 28.9.** *The function $v_p(Z)$ is a decreasing function of $p$.*

**Proof.** Let $a := \{a_j\}_{j=1}^n$ be a sequence of non-negative numbers and set

$$\|a\|_p := \left(\sum_{j=1}^n a_j^p\right)^{1/p}.$$

It will suffice to show $\|a\|_p$ is a decreasing function of $p$. To see this is true, $q = p + r$. Then

$$\|a\|_q^q = \sum_{j=1}^n a_j^{p+r} \le \left(\max_j a_j\right)^r \cdot \sum_{j=1}^n a_j^p \le \|a\|_p^r \cdot \|a\|_p^p = \|a\|_p^q,$$

wherein we have used,

$$\max_j a_j = \left(\max_j a_j^p\right)^{1/p} \le \left(\sum_{j=1}^n a_j^p\right)^{1/p} = \|a\|_p.$$

∎

**Notation 28.10 (Partitions)** *Given* $\mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\}$, *a partition of* $[0,T]$, *let*

$$\Delta_i B := B_{t_i} - B_{t_{i-1}}, \text{ and } \Delta_i t := t_i - t_{i-1}$$

*for all* $i = 1, 2, \dots, n$. *Further let* $\operatorname{mesh}(\mathcal{P}) := \max_i |\Delta_i t|$ *denote the mesh of the partition,* $\mathcal{P}$.

**Corollary 28.11.** *For all* $p > 2$ *and* $T < \infty$, $v_p\left(B|_{[0,T]}\right) < \infty$ *a.s. (We will see later that* $v_p\left(B|_{[0,T]}\right) = \infty$ *a.s. for all* $p < 2$.)

**Proof.** By Theorem 28.3, there exists $K_p < \infty$ a.s. such that

$$|B_t - B_s| \le K_p |t-s|^{1/p} \text{ for all } 0 \le s,t \le T. \tag{28.9}$$

Thus we have

$$\sum_i |\Delta_i B|^p \le \sum_i \left(K_p |t_i - t_{i-1}|^{1/p}\right)^p \le \sum_i K_p^p |t_i - t_{i-1}| = K_p^p T$$

and therefore, $v_p\left(B|_{[0,T]}\right) \le K_p^p T < \infty$ a.s. ■

**Exercise 28.2 (Quadratic Variation).** Let

$$\mathcal{P}_m := \left\{0 = t_0^m < t_1^m < \cdots < t_{n_m}^m = T\right\}$$

be a sequence of partitions such that $\operatorname{mesh}(\mathcal{P}_m) \to 0$ as $m \to \infty$. Further let

$$Q_m := \sum_{i=1}^{n_m} (\Delta_i^m B)^2 := \sum_{i=1}^{n_m} \left(B_{t_i^m} - B_{t_{i-1}^m}\right)^2. \tag{28.10}$$

Show

$$\lim_{m\to\infty} \mathbb{E}\left[(Q_m - T)^2\right] = 0$$

and $\lim_{m\to\infty} Q_m = T$ a.s. if $\sum_{m=1}^\infty \operatorname{mesh}(\mathcal{P}_m) < \infty$. This result is often abbreviated by the writing, $dB_t^2 = dt$. **Hint:** it is useful to observe; 1)

$$Q_m - T = \sum_{i=1}^{n_m} \left[(\Delta_i^m B)^2 - \Delta_i t\right]$$

and 2) using Eq. (28.2) there is a constant, $c < \infty$ such that

$$\mathbb{E}\left[(\Delta_i^m B)^2 - \Delta_i t\right]^2 = c\,(\Delta_i t)^2.$$

**Proposition 28.12.** *Suppose that* $\{\mathcal{P}_m\}_{m=1}^\infty$ *is a sequence of partitions of* $[0,T]$ *such that* $\mathcal{P}_m \subset \mathcal{P}_{m+1}$ *for all* $m$ *and* $\operatorname{mesh}(\mathcal{P}_m) \to 0$ *as* $m \to \infty$. *Then* $Q_m \to T$ *a.s. where* $Q_m$ *is defined as in Eq. (28.10).*

**Proof.** It is always possible to find another sequence of partitions, $\{\mathcal{P}'_n\}_{n=1}^\infty$, of $[0,T]$ such that $\mathcal{P}'_n \subset \mathcal{P}'_{n+1}$, $\operatorname{mesh}(\mathcal{P}'_n) \to 0$ as $n \to \infty$, $\#\left(\mathcal{P}'_{n+1}\right) = \#\left(\mathcal{P}'_n\right) + 1$, and $\mathcal{P}_m = \mathcal{P}'_{n_m}$ where $\{n_m\}_{m=1}^\infty$ is a subsequence of $\mathbb{N}$. If we let $Q'_n$ denote the quadratic variations associated to $\mathcal{P}'_n$ and we can shown $Q'_n \to T$ a.s. then we will also have $Q_m = Q'_{n_m} \to T$ a.s. as well. So with these comments we may now assume that $\#(\mathcal{P}_{n+1}) = \#(\mathcal{P}_n) + 1$.

We already know form Exercise 28.2 that $Q_m \to T$ in $L^2(P)$. So it suffices to show $Q_m$ is almost surely convergent. We will do this by showing $\{Q_m\}_{m=1}^\infty$ is a backwards martingale relative to the filtration,

$$\mathcal{F}_m := \sigma(Q_m, Q_{m+1}, \dots).$$

To do this, suppose that $\mathcal{P}_{m+1} = \mathcal{P}_m \cup \{v\}$ and $u = t_{i-1}, v = t_{i+1} \in \mathcal{P}_m$ such that $u < v < w$. Let $X := B_v - B_w$ and $Y := B_w - B_u$. Then

$$\begin{aligned} Q_m &= Q_{m+1} - (B_v - B_w)^2 - (B_w - B_u)^2 + (B_v - B_u)^2 \\ &= Q_{m+1} - X^2 - Y^2 + (X+Y)^2 \\ &= Q_{m+1} + 2XY \end{aligned}$$

therefore,

$$\mathbb{E}[Q_m|\mathcal{F}_{m+1}] = Q_{m+1} + 2\mathbb{E}[XY|\mathcal{F}_{m+1}].$$

So to finish the proof it suffices to show $\mathbb{E}[XY|\mathcal{F}_{m+1}] = 0$ a.s.

To do this let

$$b_t := \begin{cases} B_t & \text{if } t \le v \\ B_v - (B_t - B_v) & \text{if } t \ge v, \end{cases}$$

that is after $t = v$, the increments of $b$ are the **reflections** of the increments of $B$. Clearly $b_t$ is still a continuous process and it is easily verified that $\mathbb{E}[b_t b_s] = s \wedge t$. Thus $\{b_t\}_{t\ge 0}$ is still a Brownian motion. Moreover, if $Q_{m+n}(b)$ is the quadratic variation of $b$ relative to $\mathcal{P}_{m+n}$, then

$$Q_{m+n}(b) = Q_{m+n} = Q_{m+n}(B) \text{ for all } n \in \mathbb{N}.$$

On the other hand, under this transformation, $X \to X$ and $Y \to -Y$. Since $(X, Y, Q_{m+1}, Q_{m+2}, \dots)$ and $(-X, Y, Q_{m+1}, Q_{m+2}, \dots)$ have the same distribution, if we write

$$\mathbb{E}[XY|\mathcal{F}_{m+1}] = f(Q_{m+1}, Q_{m+2}, \dots) \text{ a.s.}, \tag{28.11}$$

then it follows from Exercise 16.7, that

$$\mathbb{E}[-XY|\mathcal{F}_{m+1}] = f(Q_{m+1}, Q_{m+2}, \dots) \text{ a.s.} \tag{28.12}$$

Hence we may conclude,

$$\mathbb{E}[XY|\mathcal{F}_{m+1}] = \mathbb{E}[-XY|\mathcal{F}_{m+1}] = -\mathbb{E}[XY|\mathcal{F}_{m+1}],$$

and thus $\mathbb{E}[XY|\mathcal{F}_{m+1}] = 0$ a.s. ■

**Corollary 28.13.** *If $p<2$, then $v_p\left(B|_{[0,T]}\right)=\infty$ a.s.*

**Proof.** Choose partitions, $\{\mathcal{P}_m\}$, of $[0,T]$ such that $\lim_{m\to\infty}Q_m = T$ a.s. where $Q_m$ is as in Eq. (28.10) and let $\Omega_0 := \{\lim_{m\to\infty}Q_m = T\}$ so that $P(\Omega_0)=1$. If $v_p\left(B|_{[0,T]}(\omega)\right)<\infty$ for some $\omega\in\Omega_0$, then, with $b=B(\omega)$, we would have

$$\begin{aligned}Q_m(\omega) &= \sum_{i=1}^{n_m}(\Delta_i^m b)^2 = \sum_{i=1}^{n_m}|\Delta_i^m b|^p\cdot|\Delta_i^m b|^{2-p}\\ &\le \max_i|\Delta_i^m b|^{2-p}\cdot\sum_{i=1}^{n_m}|\Delta_i^m b|^p \le [v_p(b)]^p\cdot\max_i|\Delta_i^m b|^{2-p}\end{aligned}$$

which tends to zero as $m\to\infty$ by the uniform continuity of $b$. But this contradicts the fact that $\lim_{m\to\infty}Q_m(\omega)=T$. Thus we must $v_p\left(B|_{[0,T]}\right)=\infty$ on $\Omega_0$. ∎

*Remark 28.14.* The reader may find a proof that Corollary 28.13 also holds for $p=2$ in [16, Theorem 13.69 on p. 382]. You should consider why this result in not in contradiction with Exercise 28.2 and Theorem 28.12. **Hint:** unlike the case of $p=1$, when $p>1$ the quantity;

$$v_p^\Pi(Z) := \left(\sum_{j=1}^n\left\|Z_{t_j}-Z_{t_{j-1}}\right\|^p\right)^{1/p}$$

does not increase under refinement of partitions so that

$$v_p(Z)=\sup_\Pi v_p^\Pi(Z)\ne\lim_{|\Pi|\to0}v_p^\Pi(Z)$$

when $p>1$.

**Corollary 28.15 (Roughness of Brownian Paths).** *A Brownian motion, $\{B_t\}_{t\ge0}$, is **not** almost surely $\alpha$ – Hölder continuous for any $\alpha>1/2$.*

**Proof.** According to Exercise 28.2, we may choose partition, $\mathcal{P}_m$, such that $\mathrm{mesh}(\mathcal{P}_m)\to0$ and $Q_m\to T$ a.s. If $B$ were $\alpha$ – Hölder continuous for some $\alpha>1/2$, then

$$\begin{aligned}Q_m &= \sum_{i=1}^{n_m}(\Delta_i^m B)^2 \le C\sum_{i=1}^{n_m}(\Delta_i^m t)^{2\alpha}\le C\max\left([\Delta_i t]^{2\alpha-1}\right)\sum_{i=1}^{n_m}\Delta_i^m t\\ &\le C\left[\mathrm{mesh}(\mathcal{P}_m)\right]^{2\alpha-1}T\to0 \text{ as } m\to\infty\end{aligned}$$

which contradicts the fact that $Q_m\to T$ as $m\to\infty$. ∎

**Lemma 28.16.** *For any $\alpha>1/2$, $\limsup_{t\downarrow0}|B_t|/t^\alpha=\infty$ a.s. (See Exercise* **??** *below to see that $\alpha=1/2$ would work as well.)*

**Proof.** If $\limsup_{t\downarrow0}|\omega_t|/t^\alpha<\infty$ then there would exists $C<\infty$ such that $|\omega_t|\le Ct^\alpha$ for all $t\le1$ and in particular, $\left|\omega_{1/n}\right|\le Cn^{-\alpha}$ for all $n\in\mathbb{N}$. Hence we have shown

$$\left\{\limsup_{t\downarrow0}|B_t|/t^\alpha<\infty\right\}\subset\cup_{C\in\mathbb{N}}\cap_{n\in\mathbb{N}}\left\{\left|B_{1/n}\right|\le Cn^{-\alpha}\right\}.$$

This completes the proof because,

$$\begin{aligned}P\left(\cap_{n\in\mathbb{N}}\left\{\left|B_{1/n}\right|\le Cn^{-\alpha}\right\}\right)&\le\liminf_{n\to\infty}P\left(\left|B_{1/n}\right|\le Cn^{-\alpha}\right)\\&=\liminf_{n\to\infty}P\left(\frac{1}{\sqrt{n}}|B_1|\le Cn^{-\alpha}\right)\\&=\liminf_{n\to\infty}P\left(|B_1|\le Cn^{1/2-\alpha}\right)=P(|B_1|=0)=0\end{aligned}$$

if $\alpha>1/2$. ∎

**Theorem 28.17 (Nowhere $1/2+\varepsilon$ – Hölder Continuous).** *Let*

$$W:=\{\omega\in C([0,\infty)\to\mathbb{R}):\omega(0)=0\},$$

*$\mathcal{B}$ denote the $\sigma$ – field on $W$ generated by the projection maps, $b_t(\omega)=\omega(t)$ for all $t\in[0,\infty)$, and $\mu$ be **Wiener measure** on $(W,\mathcal{B})$, i.e. $\mu$ is the Law of a Brownian motion. For $\alpha>1/2$ and $E_\alpha$ denote the set of $\omega\in W$ such that $\omega$ is $\alpha$–Hölder continuous at some point $t=t_\omega\in[0,1]$. when $\mu^*(E_\alpha)=0$, i.e. there exists a set $\tilde{E}_\alpha\in\mathcal{B}$ such that*

$$E_\alpha=\left\{\inf_{0\le t\le1}\limsup_{h\to0}\frac{|\omega(t+h)-\omega(t)|}{|h|^\alpha}<\infty\right\}\subset\tilde{E}_\alpha$$

*and $\mu\left(\tilde{E}_\alpha\right)=0$. In particular, $\mu$ is concentrated on $\tilde{E}_\alpha^c$ which is a subset of the collection paths which are nowhere differentiable on $[0,1]$.*

**Proof.** Let $\alpha\in(0,1)$ and $\nu\in\mathbb{N}$ – to be chosen more specifically later. If $\omega\in E_\alpha$, then there exists, $t\in[0,1]$, $C<\infty$, such that

$$|\omega(t)-\omega(s)|\le C|t-s|^\alpha \text{ for all } |s|\le\nu+1.$$

For all $n\in\mathbb{N}$ we may choose $i\ge0$ so that $\left|t-\frac{i}{n}\right|<\frac{1}{n}$. By the triangle inequality, for all $j=1,2,\dots,\nu$, we have

$$\begin{aligned}\left|\omega\left(\frac{i+j}{n}\right)-\omega\left(\frac{i+j-1}{n}\right)\right| &\le \left|\omega\left(\frac{i+j}{n}\right)-\omega(t)\right|+\left|\omega(t)-\omega\left(\frac{i+j-1}{n}\right)\right| \\ &\le C\left[\left|\frac{i+j}{n}-t\right|^{\alpha}+\left|\frac{i+j-1}{n}-t\right|^{\alpha}\right] \\ &\le Cn^{-\alpha}\left[|\nu+1|^{\alpha}+|\nu|^{\alpha}\right] =: Dn^{-\alpha}.\end{aligned}$$

Therefore, $\omega \in E_\alpha$ implies there exists $D \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ there exists $i \le n$ such that

$$\left|\omega\left(\frac{i+j}{n}\right)-\omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\ \forall\ j=1,2,\dots,\nu.$$

Letting

$$A_D := \cap_{n=1}^{\infty}\cup_{i\le n}\cap_{j=1}^{\nu}\left\{\omega : \left|\omega\left(\frac{i+j}{n}\right)-\omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right\},$$

we have shown that $E_\alpha \subset \cup_{D\in\mathbb{N}} A_D$. We now complete the proof by showing $P(A_D)=0$. To do this, we compute,

$$\begin{aligned}P(A_D) &\le \liminf_{n\to\infty} P\left(\cup_{i\le n}\cap_{j=1}^{\nu}\left\{\omega : \left|\omega\left(\frac{i+j}{n}\right)-\omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right\}\right) \\ &\le \liminf_{n\to\infty}\sum_{i\le n}\prod_{j=1}^{\nu} P\left(\omega : \left|\omega\left(\frac{i+j}{n}\right)-\omega\left(\frac{i+j-1}{n}\right)\right| \le Dn^{-\alpha}\right) \\ &= \liminf_{n\to\infty} n\left[P\left(\frac{1}{\sqrt{n}}|N| \le Dn^{-\alpha}\right)\right]^{\nu} \\ &= \liminf_{n\to\infty} n\left[P\left(|N| \le Dn^{\frac{1}{2}-\alpha}\right)\right]^{\nu} \\ &\le \liminf_{n\to\infty} n\left[Cn^{\frac{1}{2}-\alpha}\right]^{\nu} = C^{\nu}\liminf_{n\to\infty} n^{1+\left(\frac{1}{2}-\alpha\right)\nu}. \qquad (28.13)\end{aligned}$$

wherein we have used

$$\mu(|N|\le\delta) = \frac{1}{\sqrt{2\pi}}\int_{|x|\le\delta} e^{-\frac{1}{2}x^2}dx \le \frac{1}{\sqrt{2\pi}}2\delta.$$

The last limit in Eq. (28.13) is zero provided we choose $\alpha > \frac{1}{2}$ and $\nu\left(\alpha-\frac{1}{2}\right) > 1$. ■

We end this section with a often useful, albeit heuristic, interpretation of Wiener measure, $\mu := \mathrm{Law}\left(B_{(\cdot)}\right)$. This interpretation empathizes the Gaussian nature of $\mu$.

**Theorem 28.18 (Fake Theorem!).** *Let $\mu := \mathrm{Law}\left(B_{(\cdot)}\right)$ thought of as a measure on $C([0,1],\mathbb{R})$ and let*

$$H := \left\{x \in C([0,1],\mathbb{R}) : x(0)=0 \text{ and } \int_0^1 |\dot{x}(t)|^2 dt < \infty\right\}.$$

*Then heuristically we expect $\mu(H)=1$,*

$$d\mu(x) = \frac{1}{Z}\exp\left(-\frac{1}{2}\int_0^1|\dot{x}(t)|^2 dt\right)dm(x)$$

*where $m$ is "Lebesgue" measure on $H$, and $Z$ is a normalization constant.*

**Fake Proof.** Let $\mathcal{P} := \{0=t_0<t_1<\dots<t_n=1\}$ be a partition of $[0,1]$, let

$$H_{\mathcal{P}} := \{x \in H : x(0)=0 \text{ and } x''(t)=0\ \forall\ t\notin\mathcal{P}\},$$

and for $x \in C([0,1],\mathbb{R})$ with $x(0)=0$, let $x^{\mathcal{P}}$ be the unique element of $H_{\mathcal{P}}$ such that $x^{\mathcal{P}}(t_i)=x(t_i)$ for $1\le i\le n$. Now suppose that $F : C([0,1],\mathbb{R})\to\mathbb{R}$ is a bounded continuous function. The key points are;

1. $\max_{0\le t\le 1}\left|B_t^{\mathcal{P}}(\omega)\to B_t(\omega)\right| \to 0$ as $\mathrm{mesh}(\mathcal{P})\to 0$, and therefore $F\left(B_{(\cdot)}^{\mathcal{P}}\right)\to F\left(B_{(\cdot)}\right)$ as $\mathrm{mesh}(\mathcal{P})\to 0$.
2. By the dominated convergence theorem,
$$\int_{C([0,1],\mathbb{R})} F(x)\,d\mu(x) := \mathbb{E}\left[F\left(B_{(\cdot)}\right)\right] = \lim_{\mathrm{mesh}(\mathcal{P})\to 0}\mathbb{E}\left[F\left(B_{(\cdot)}^{\mathcal{P}}\right)\right].$$
3. We have
$$\mathbb{E}\left[F\left(B_{(\cdot)}^{\mathcal{P}}\right)\right] = \int_{H_{\mathcal{P}}} f(x)\frac{1}{Z_{\mathcal{P}}}\exp\left(-\frac{1}{2}\int_0^1|\dot{x}(t)|^2 dt\right)dm_{\mathcal{P}}(x)$$
where $m_{\mathcal{P}}$ is a Lebesgue measure on $H_{\mathcal{P}}$ and $Z_{\mathcal{P}}$ is a normalization constant.
4. Given items 2. and 3. we formally have upon passing to the limit as $\mathrm{mesh}(\mathcal{P})\to 0$ that
$$\int_{C([0,1],\mathbb{R})} F(x)\,d\mu(x) = \int_H F(x)\frac{1}{Z}\exp\left(-\frac{1}{2}\int_0^1|\dot{x}(t)|^2 dt\right)dm(x).$$

The main point it to prove item 3. above which is done by direct computation. To this end, let $f : \mathbb{R}^n\to\mathbb{R}$ be defined so that $f(x(t_1),\dots,x(t_n)) = F\left(x^{\mathcal{P}}\right)$, then

$$\begin{aligned}\mathbb{E}\left[F\left(B_{(\cdot)}^{\mathcal{P}}\right)\right] &= \mathbb{E}\left[f\left(B_{t_1},\dots,B_{t_n}\right)\right] \\ &= \int_{\mathbb{R}^n} f(x_1,x_2,\dots,x_n)\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\Delta_j}}\exp\left(-\frac{1}{2\Delta_j}|x_j-x_{j-1}|^2\right)dx_j\end{aligned}$$

where $x_0 = 0$ by convention. The key point is that the above expression may be written as

$$\begin{aligned}\mathbb{E}\left[f\left(B_{t_1},\ldots,B_{t_n}\right)\right] \\ = \int_{H_{\mathcal{P}}} f\left(x\left(t_1\right),x\left(t_2\right),\ldots,x\left(t_n\right)\right)\frac{1}{Z_{\mathcal{P}}}\exp\left(-\frac{1}{2}\int_0^1\left|\dot{x}\left(t\right)\right|^2 dt\right)dm_{\mathcal{P}}\left(x\right)\end{aligned} \tag{28.14}$$

where

$$Z_{\mathcal{P}} := \prod_{j=1}^{n}\sqrt{2\pi\Delta_j} \text{ and } dm_{\mathcal{P}}\left(x\right) := d\left[x\left(t_1\right)\right]\ldots d\left[x\left(t_n\right)\right].$$

The reason this is true is that for $x \in H_{\mathcal{P}}$,

$$\dot{x}\left(t\right) = \frac{x\left(t_j\right)-x\left(t_{j-1}\right)}{\Delta_j} \text{ for } t_{j-1} < t < t_j$$

and therefore

$$\begin{aligned}\int_0^1\left|\dot{x}\left(t\right)\right|^2 dt &= \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\left|\dot{x}\left(t\right)\right|^2 dt = \sum_{j=1}^{n}\left|\frac{x\left(t_j\right)-x\left(t_{j-1}\right)}{\Delta_j}\right|^2\Delta_j \\ &= \sum_{j=1}^{n}\frac{\left|x\left(t_j\right)-x\left(t_{j-1}\right)\right|^2}{\Delta_j}.\end{aligned}$$

■

## 28.3 Scaling Properties of B. M.

**Theorem 28.19 (Transformations preserving B. M.).** *Let $\{B_t\}_{t\geq 0}$ be a Brownian motion and $\mathcal{B}_t := \sigma\left(B_s : s \leq t\right).$ Then;*

1. *$b_t = -B_t$ is again a Brownian motion.*
2. *if $c > 0$ and $b_t := c^{-1/2}B_{ct}$ is again a Brownian motion.*
3. *$b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion. In particular, $\lim_{t\downarrow 0} tB_{1/t} = 0$ a.s.*
4. *for all $T \in (0,\infty)$, $b_t := B_{t+T} - B_T$ for $t \geq 0$ is again a Brownian motion which is independent of $\mathcal{B}_T$.*
5. *for all $T \in (0,\infty)$, $b_t := B_{T-t} - B_T$ for $0 \leq t \leq T$ is again a Brownian motion on $[0,T]$.*

**Proof.** It is clear that in each of the four cases above $\{b_t\}_{t\geq 0}$ is still a Gaussian process. Hence to finish the proof it suffices to verify, $\mathbb{E}\left[b_t b_s\right] = s \wedge t$ which is routine in all cases. Let us work out item 3. in detail to illustrate the method. For $0 < s < t$,

$$\mathbb{E}\left[b_s b_t\right] = st\mathbb{E}\left[B_{s^{-1}}B_{t^{-1}}\right] = st\left(s^{-1}\wedge t^{-1}\right) = st\cdot t^{-1} = s.$$

Notice that $t \to b_t$ is continuous for $t > 0$, so to finish the proof we must show that $\lim_{t\downarrow 0} b_t = 0$ a.s. However, this follows from Kolmogorov's continuity criteria. Since $\{b_t\}_{t\geq 0}$ is a pre-Brownian motion, we know there is a version, $\tilde{b}$ which is a.s. continuous for $t \in [0,\infty)$. By exercise 27.1, we know that

$$E := \left\{\omega \in \Omega : b_t\left(\omega\right) \neq \tilde{b}_t\left(\omega\right) \text{ for some } t > 0\right\}$$

is a null set. Hence $\omega \notin E$ it follows that

$$\lim_{t\downarrow 0} b_t\left(\omega\right) = \lim_{t\downarrow 0}\tilde{b}_t\left(\omega\right) = 0.$$

■

**Corollary 28.20 (B. M. Law of Large Numbers).** *Suppose $\{B_t\}_{t\geq 0}$ is a Brownian motion, then almost surely, for each $\beta > 1/2$,*

$$\limsup_{t\to\infty}\frac{|B_t|}{t^\beta} = \begin{cases} 0 & if \quad \beta > 1/2 \\ \infty & if\ \beta \in (0,1/2).\end{cases} \tag{28.15}$$

**Proof.** Since $b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion, we know that for all $\alpha < 1/2$ there exists, $C_\alpha\left(\omega\right) < \infty$ such that, almost surely,

$$t\left|B_{1/t}\right| = \left|tB_{1/t}\right| = |b_t| \leq C_\alpha\left|t\right|^\alpha \text{ for all } t \leq 1.$$

Replacing $t$ by $1/t$ in this inequality implies, almost surely, that

$$\frac{1}{t}\left|B_t\right| \leq \frac{C_\alpha}{|t|^\alpha} \text{ for all } t \geq 1.$$

or equivalently that

$$|B_t| \leq C_\alpha t^{1-\alpha} \text{ for all } t \geq 1. \tag{28.16}$$

Hence if $\beta > 1/2$, let $\alpha < 1/2$ such that $\beta < 1-\alpha$. Then Eq. (28.15) follows from Eq. (28.16).

On the other hand, taking $\alpha > 1/2$, we know by Lemma 28.16 (or Theorem 28.17) that

$$\limsup_{t\downarrow 0}\frac{t\left|B_{1/t}\right|}{t^\alpha} = \limsup_{t\downarrow 0}\frac{|b_t|}{t^\alpha} = \infty \text{ a.s.}$$

This may be expressed as saying

$$\infty = \limsup_{t\to\infty}\frac{t^{-1}\left|B_t\right|}{t^{-\alpha}} = \limsup_{t\to\infty}\frac{|B_t|}{t^{1-\alpha}} \text{ a.s.}$$

Since $\beta := 1-\alpha$ is any number less that $1/2$, the proof is complete. ■

# 29

# Filtrations and Stopping Times

For our later development we need to go over some measure theoretic preliminaries about processes indexed by $\mathbb{R}_+ := [0, \infty)$. We will continue this discussion in more depth later. For this chapter we will always suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+}\right)$ is a **filtered measurable space**, i.e. $\Omega$ is as set, $\mathcal{B} \subset 2^\Omega$ is a $\sigma$ algebra, and $\{\mathcal{B}_t\}_{t \in \mathbb{R}_+}$ is a **filtration** which to say each $\mathcal{B}_t$ is a sub – $\sigma$ – algebra of $\mathcal{B}$ and $\mathcal{B}_s \subset \mathcal{B}_t$ for all $s \leq t$.

## 29.1 Measurability Structures

**Notation 29.1** ($\mathcal{B}_t^{\pm}$) *Let*

$$\mathcal{B}_\infty = \mathcal{B}_{\infty+} = \vee_{t \in \mathbb{R}_+} \mathcal{B}_t = \sigma\left(\cup_{t \in \mathbb{R}_+} \mathcal{B}_t\right) \subset \mathcal{B},$$

*and for* $t \in \mathbb{R}_+$, *let*

$$\mathcal{B}_t^+ = \mathcal{B}_{t+} := \cap_{s>t} \mathcal{B}_s.$$

*Also let* $\mathcal{B}_{0-} := \mathcal{B}_0$ *and for* $t \in (0, \infty]$ *let*

$$\mathcal{B}_{t-} := \vee_{s<t} \mathcal{B}_s = \sigma\left(\cup_{s<t} \mathcal{B}_s\right).$$

*(Observe that* $\mathcal{B}_{\infty-} = \mathcal{B}_\infty$.*)*

The filtration, $\left\{\mathcal{B}_t^+\right\}_{t \in \mathbb{R}_+}$, "peaks" infinitesimally into the future while $\mathcal{B}_t^-$ limits itself to knowing about the state of the system up to the times infinitesimally before time $t$.

**Definition 29.2 (Right continuous filtrations).** *The filtration* $\{\mathcal{B}_t\}_{t \geq 0}$ *is* ***right continuous*** *if* $\mathcal{B}_t^+ := \mathcal{B}_{t+} = \mathcal{B}_t$ *for all* $t \geq 0$.

The next result is trivial but we record it as a lemma nevertheless.

**Lemma 29.3 (Right continuous extension).** *Suppose* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}\right)$ *is a filtered space and* $\mathcal{B}_t^+ := \mathcal{B}_{t+} := \cap_{s>t} \mathcal{B}_s$. *Then* $\left\{\mathcal{B}_t^+\right\}_{t \geq 0}$ *is right continuous. (We refer to* $\left\{\mathcal{B}_t^+\right\}_{t \in \mathbb{R}_+}$ *as the* ***right continuous filtration*** *associated to* $\{\mathcal{B}_t\}_{t \in \mathbb{R}_+}$.*)*

**Exercise 29.1.** Suppose $(\Omega, \mathcal{F})$ is a measurable space, $(S, \rho)$ is a separable metric space[1], and $\mathcal{S}$ is the Borel $\sigma$ – algebra on $S$ – i.e. the $\sigma$ – algebra generated by all open subset of $S$.

1. Let $D \subset S$ be a countable dense set and $\mathbb{Q}_+ := \mathbb{Q} \cap \mathbb{R}_+$. Show $\mathcal{S}$ may be described as the $\sigma$ – algebra generated by all open (or closed) balls of the form
$$B(a, \varepsilon) := \{s \in S : \rho(s, a) < \varepsilon\} \tag{29.1}$$
$$(\text{or } C(a, \varepsilon) := \{s \in S : \rho(s, a) \leq \varepsilon\}) \tag{29.2}$$
with $a \in D$ and $\varepsilon \in \mathbb{Q}_+$.
2. Show a function, $Y : \Omega \to S$, is $\mathcal{F}/\mathcal{S}$ – measurable iff the functions, $\Omega \ni \omega \to \rho(x, Y(\omega)) \in \mathbb{R}_+$ are measurable for all $x \in D$. **Hint:** show, for each $x \in S$, that $\rho(x, \cdot) : S \to \mathbb{R}_+$ is a measurable map.
3. If $X_n : \Omega \to S$ is a sequence of $\mathcal{F}/\mathcal{S}$ – measurable maps such that $X(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists in $S$ for all $\omega \in \Omega$, then the limiting function, $X$, is $\mathcal{F}/\mathcal{S}$ – measurable as well. (**Hint:** use item 2.)

**Definition 29.4.** *Suppose* $S$ *is a metric space,* $\mathcal{S}$ *is the Borel* $\sigma$ *– algebra on* $S$, *and* $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+}\right)$ *is a filtered measurable space. A process,* $X_t : \Omega \to S$ *for* $t \in \mathbb{R}_+$ *is;*

1. ***adapted*** *if* $X_t$ *is* $\mathcal{B}_t/\mathcal{S}$*– measurable for all* $t \in \mathbb{R}_+$,
2. ***right continuous*** *if* $t \to X_t(\omega)$ *is right continuous for all* $\omega \in \Omega$,
3. ***left continuous*** *if* $t \to X_t(\omega)$ *is left continuous for all* $\omega \in \Omega$, *and*
4. ***progressively measurable***, *if for all* $T \in \mathbb{R}_+$, *the map* $\varphi^T : [0, T] \times \Omega \to S$ *defined by* $\varphi^T(t, \omega) := X_t(\omega)$ *is* $\mathcal{B}_{[0,T]} \otimes \mathcal{B}_T/\mathcal{S}$ *– measurable.*

**Lemma 29.5.** *Let* $\varphi(t, \omega) := X_t(\omega)$ *where we are continuing the notation in Definition 29.4. If* $X_t : \Omega \to S$ *is a progressively measurable process then* $X_\cdot$ *is adapted and* $\varphi : \mathbb{R}_+ \times \Omega \to S$ *is* $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}/\mathcal{S}$ *– measurable and the* $X$ *is adapted.*

**Proof.** For $T \in \mathbb{R}_+$, let $\eta_T : \Omega \to [0, T] \times \Omega$, be defined by $\eta_T(\omega) := (T, \omega)$. If $a \in [0, T]$ and $A \in \mathcal{B}_T$, then $\eta_T^{-1}([0, a] \times A) = \emptyset \in \mathcal{B}_T$ if $a \neq T$

[1] If you are unconfortable with this much generality, you may assume $S$ is a subset of $\mathbb{R}^d$ and $\rho(x, y) := \|x - y\|$ for all $x, y \in S$.

and $\eta_T^{-1}([0,a]\times A) = A \in \mathcal{B}_T$ if $a = T$. This shows $\eta_T$ is $\mathcal{B}_T/\mathcal{B}_{[0,T]}\otimes\mathcal{B}$ – measurable. Therefore, the composition, $\varphi^T\circ\eta_T = X_T$ is $\mathcal{B}_T/\mathcal{S}$ – measurable for all $T\in\mathbb{R}_+$ which is the statement that $X$ is adapted.

For $V\in\mathcal{S}$ and $T<\infty$, we have

$$\varphi^{-1}(V)\cap([0,T]\times\Omega) = \left(\varphi^T\right)^{-1}(V)\in\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\subset\mathcal{B}_{[0,T]}\otimes\mathcal{B}. \quad (29.3)$$

Since $\mathcal{B}_{[0,T]}\otimes\mathcal{B}$ and $\left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}$ are $\sigma$ – algebras which are generated by sets of the form $[0,a]\times A$ with $a\in[0,T]$ and $A\in\mathcal{B}$, they are equal – $\mathcal{B}_{[0,T]}\otimes\mathcal{B} = \left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}$. This observation along with Eq. (29.3) then implies,

$$\varphi^{-1}(V)\cap([0,T]\times\Omega)\in\left(\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right)_{[0,T]\times\Omega}\subset\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}$$

and therefore,

$$\varphi^{-1}(V) = \cup_{T\in\mathbb{N}}\left[\varphi^{-1}(V)\cap([0,T]\times\Omega)\right]\in\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}.$$

This shows $\varphi$ is $\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}/\mathcal{S}$ – measurable as claimed. ∎

**Lemma 29.6.** *Suppose $S$ is a separable metric space, $\mathcal{S}$ is the Borel $\sigma$ – algebra on $S$, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}\right)$ is a filtered measurable space, and $X_t:\Omega\to S$ for $t\in\mathbb{R}_+$ is an adapted right continuous process. Then $X_\cdot$ is progressively measurable and the map, $\varphi:\mathbb{R}_+\times\Omega\to S$ defined by $\varphi(t,\omega) = X_t(\omega)$ is $\left[\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right]/\mathcal{S}$ – measurable.*

**Proof.** Let $T\in\mathbb{R}_+$. To each $n\in\mathbb{N}$ let $\varphi_n(0,\omega) = X_0(\omega)$ and

$$\varphi_n(t,\omega) := X_{\frac{kT}{2^n}}(\omega) \text{ if } \frac{(k-1)T}{2^n} < t \le \frac{kT}{2^n} \text{ for } k\in\{1,2,\dots,2^n\}.$$

Then

$$\varphi_n^{-1}(A) = \left[\{0\}\times X_0^{-1}(A)\right]\cup_{k=1}^\infty\left[\left(\frac{(k-1)T}{2^n},\frac{kT}{2^n}\right]\times X^{-1}_{Tk2^{-n}}(A)\right]\in\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T,$$

showing that $\varphi_n$ is $\left[\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\right]/\mathcal{S}$ – measurable. Therefore, by Exercise 29.1, $\varphi^T = \lim_{n\to\infty}\varphi_n$ is also $\left[\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T\right]/\mathcal{S}$ – measurable. The fact that $\varphi$ is $\left[\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{B}\right]/\mathcal{S}$ – measurable now follows from Lemma 29.5. ∎

**Lemma 29.7.** *Suppose that $T\in(0,\infty)$, $\Omega_T := C([0,T],\mathbb{R})$, and $\mathcal{F}_T = \sigma\left(B_t^T : t\le T\right)$, where $B_t^T(\omega) = \omega(t)$ for all $t\in[0,T]$ and $\omega\in\Omega_T$. Then;*

1. *The map, $\pi:\Omega\to\Omega_T$ defined by $\pi(\omega) := \omega|_{[0,T]}$ is $\mathcal{B}_T/\mathcal{F}_T$ – measurable.*
2. *A function, $F:\Omega\to\mathbb{R}$ is $\mathcal{B}_T$ – measurable iff there exists a function, $f:\Omega_T\to\mathbb{R}$ which is $\mathcal{F}_T$ – measurable such that $F = f\circ\pi$.*
3. *Let $\|\omega\|_T := \max_{t\in[0,T]}|\omega(t)|$ so that $\left(\Omega_T,\|\cdot\|_T\right)$ is a Banach space. The Borel $\sigma$ – algebra, $\mathcal{B}_{\Omega_T}$ on $\Omega_T$ is the same as $\mathcal{F}_T$.*
4. *If $F = f\circ\pi$ where $f:\Omega_T\to\mathbb{R}$ is a $\|\cdot\|_T$ – continuous function, then $F$ is $\mathcal{B}_T$ – measurable.*

**Proof.** 1. Since $B_t^T\circ\pi = B_t$ is $\mathcal{B}_T$ – measurable for all $t\in[0,T]$, it follows that $\pi$ is measurable.

2. Clearly if $f:\Omega_T\to\mathbb{R}$ is $\mathcal{F}_T$ – measurable, then $F = f\circ\pi:\Omega\to\mathbb{R}$ is $\mathcal{B}_T$ – measurable. For the converse assertion, let $\mathbb{H}$ denote the bounded $\mathcal{B}_T$ – measurable functions of the form $F = f\circ\pi$ with $f:\Omega_T\to\mathbb{R}$ being $\mathcal{F}_T$ – measurable. It is a simple matter to check that $\mathbb{H}$ is a vector space which is closed under bounded convergence and contains all cylinder functions of the form, $G(B_{t_1},\dots B_{t_n}) = G\left(B_{t_1}^T,\dots B_{t_n}^T\right)\circ\pi$ with $\{t_i\}_{i=1}^n\subset[0,T]$. The latter set of functions generates the $\sigma$ – algebra, $\mathcal{B}_T$, and so by the multiplicative systems theorem, $\mathbb{H}$ contains all bounded $\mathcal{B}_T$ – measurable functions. For a general $\mathcal{B}_T$ – measurable function, $F:\Omega\to\mathbb{R}$, the truncation by $N\in\mathbb{N}$, $F_N = -N\vee(F\wedge N)$, is of the form $F_N = f_N\circ\pi$ for some $\mathcal{F}_T$ – measurable function, $f_N:\Omega_T\to\mathbb{R}$. Since every $\omega\in\Omega_T$ extends to an element of $\tilde\omega\in\Omega$, it follows that $\lim_{N\to\infty}f_N(\omega) = \lim_{N\to\infty}F_N(\tilde\omega) = F(\tilde\omega)$ exists. Hence if we let $f := \lim_{N\to\infty}f_N$, we will have $F = f\circ\pi$ with $f$ being a $\mathcal{F}_T$ – measurable function.

3. Recall that $\mathcal{B}_{\Omega_T} = \sigma(\text{open sets})$. Since $B_s:\Omega\to\mathbb{R}$ is continuous for all $s$, it follows that $\sigma\left(B_s^T\right)\subset\mathcal{B}_{\Omega_T}$ for all $s$ and hence $\mathcal{F}_T\subset\mathcal{B}_{\Omega_T}$. Conversely, since

$$\|\omega\| := \sup_{t\in\mathbb{Q}\cap[0,T]}|\omega(t)| = \sup_{t\in\mathbb{Q}\cap[0,T]}\left|B_t^T(\omega)\right|,$$

it follows that $\|\cdot-\omega_0\| = \sup_{t\in\mathbb{Q}\cap[0,T]}\left|B_t^T(\cdot)-\omega_0(t)\right|$ is $\mathcal{F}_T$ – measurable for every $\omega_0\in\Omega$. From this we conclude that each open ball, $B(\omega_0,r) := \{\omega\in\Omega : \|\omega-\omega_0\| < r\}$, is in $\mathcal{F}_T$. By the classical Weierstrass approximation theorem we know that $\Omega$ is separable and hence we may now conclude that $\mathcal{F}_T$ contains all open subsets of $\Omega$. This shows that $\mathcal{B}_{\Omega_T} = \sigma(\text{open sets})\subset\mathcal{F}_T$.

4. Any continuous function, $f:\Omega\to\mathbb{R}$ is $\mathcal{B}_{\Omega_T} = \mathcal{F}_T$ – measurable and therefore, $F = f\circ\pi$ is $\mathcal{B}_T$ – measurable since it is the composition of two measurable functions. ∎

## 29.2 Stopping and optional times

**Definition 29.8.** *A random time $T:\Omega\to[0,\infty]$ is a **stopping time** iff $\{T\le t\}\in\mathcal{B}_t$ for all $t\ge 0$ and is an **optional time** iff $\{T<t\}\in\mathcal{B}_t$ for all $t\ge 0$.*

If $T$ is an optional time, the condition $\{T<0\}=\emptyset\in\mathcal{B}_{0+}$ is vacuous. Moreover, since $\{T<t\}\downarrow\{T=0\}$ as $t\downarrow 0$, it follows that $\{T=0\}\in\mathcal{B}_{0+}$ when $T$ is an optional time.

**Proposition 29.9.** *Suppose $T:\Omega\to[0,\infty]$ is a random time. Then;*

1. *If $T(\omega)=s$ with $s\geq 0$ for all $\omega$, then $T$ is a stopping time.*
2. *Every stopping time is optional.*
3. *$T$ is a $\{\mathcal{B}_t\}$ – optional time iff $T$ is a $\{\mathcal{B}_t^+\}$ – stopping time. In particular, if $\mathcal{B}_t$ is right continuous, i.e. $\mathcal{B}_t^+=\mathcal{B}_t$ for all $t$, then the notion of optional time and stopping time are the same.*

**Proof.** 1.

$$\{T\leq t\}=\begin{cases}\emptyset \text{ if } t<s\\ \Omega \text{ if } t\geq s\end{cases}$$

which show $\{T\leq t\}$ is in any $\sigma$ – algebra on $\Omega$.

2. If $T$ is a stopping time, $t>0$ and $t_n\in(0,t)$ with $t_n\uparrow t$, then

$$\{T<t\}=\cup_n\{T\leq t_n\}\in\mathcal{B}_{t-}\subset\mathcal{B}_t.$$

This shows $T$ is an optional time.

3. If $T$ is $\{\mathcal{B}_t\}$ – optional and $t\geq 0$, choose $t_n>t$ such that $t_n\downarrow t$. Then $\{T<t_n\}\downarrow\{T\leq t\}$ which implies $\{T\leq t\}\in\mathcal{B}_{t+}=\mathcal{B}_t^+$. Conversely if $T$ is an $\{\mathcal{B}_{t+}\}$ – stopping time, $t>0$, and $t_n\in(0,t)$ with $t_n\uparrow t$, then $\{T\leq t_n\}\in\mathcal{B}_{t_n+}\subset\mathcal{B}_t$ for all $n$ and therefore,

$$\{T<t\}=\cup_{n=1}^{\infty}\{T\leq t_n\}\in\mathcal{B}_t.$$

■

**Exercise 29.2.** Suppose, for all $t\in\mathbb{R}_+$, that $X_t:\Omega\to\mathbb{R}$ is a function. Let $\mathcal{B}_t:=\mathcal{B}_t^X:=\sigma(X_s:s\leq t)$ and $\mathcal{B}=\mathcal{B}_\infty:=\vee_{0\leq t<\infty}\mathcal{B}_t$. (Recall that the general element, $A\in\mathcal{B}_t$ is of the form, $A=X_\Lambda^{-1}\left(\tilde{A}\right)$ where $\Lambda$ is a countable subset of $[0,\infty)$, $\tilde{A}\subset\mathbb{R}^\Lambda$ is a measurable set relative to the product $\sigma$ – algebra on $\mathbb{R}^\Lambda$, and $X_\Lambda:\Omega\to\mathbb{R}^\Lambda$ is defined by, $X_\Lambda(\omega)(t)=X_t(\omega)$ for all $t\in\Lambda$.) If $T$ is a stopping time and $\omega,\omega'\in\Omega$ satisfy $X_t(\omega)=X_t(\omega')$ for all $t\in[0,T(\omega)]\cap\mathbb{R}$, then show $T(\omega)=T(\omega')$.

**Definition 29.10.** *Given a process, $X_t:\Omega\to S$ and $A\subset S$, let*

$$T_A(\omega):=\inf\{t>0:X_t(\omega)\in A\}\ \text{and}$$
$$D_A(\omega):=\inf\{t\geq 0:X_t(\omega)\in A\}$$

*be the **first hitting time** and **Debut (first entrance time)** of $A$. As usual the infimum of the empty set is taken to be infinity.*

Clearly, $D_A\leq T_A$ and if $D_A(\omega)>0$ or more generally if $X_0(\omega)\notin A$, then $T_A(\omega)=D_A(\omega)$. Hence we will have $D_A=T_A$ iff $T_A(\omega)=0$ whenever $X_0(\omega)\in A$.

In the sequel will typically assume that $(S,\rho)$ is a metric space and $\mathcal{S}$ is the Borel $\sigma$ – algebra on $S$. We will also typically assume (or arrange) for our processes to have right continuous sample paths. If $A$ is an open subset of $S$ and $t\to X_t(\omega)$ is right continuous, then $T_A=D_A$. Indeed, if $X_0(\omega)\in A$, then by the right continuity of $X_\cdot(\omega)$, we know that $\lim_{t\downarrow 0}X_t(\omega)=X_0(\omega)\in A$ and hence $X_t(\omega)\in A$ for all $t>0$ sufficiently close to 0 and therefore, $T_A(\omega)=0$. On the other hand, if $A$ is a closed set and $X_0(\omega)\in\mathrm{bd}(A)$, there is no need for $T_A(\omega)=0$ and hence in this case, typically $D_A\lneqq T_A$.

**Proposition 29.11.** *Suppose $\left(\Omega,\{\mathcal{B}_t\}_{t\geq 0},\mathcal{B}\right)$ is a filtered measurable space, $(S,\rho)$ is a metric space, and $X_t:\Omega\to S$ is a right continuous $\{\mathcal{B}_t\}_{t\geq 0}$ – adapted process. Then;*

1. *If $A\subset S$ is a open set, $T_A=D_A$ is an optional time.*
2. *If $A\subset S$ is closed, on $\{T_A<\infty\}$ ($\{D_A<\infty\}$), $X_{T_A}\in A$ ($X_{D_A}\in A$).*
3. *If $A\subset S$ is closed and $X$ is a continuous process, then $D_A$ is a stopping time.*
4. *If $A\subset S$ is closed and $X$ is a continuous process, then $T_A$ is an optional time. In fact, $\{T_A\leq t\}\in\mathcal{B}_t$ for all $t>0$ while $\{T_A=0\}\in\mathcal{B}_{0+}$, see Figure 29.1.*

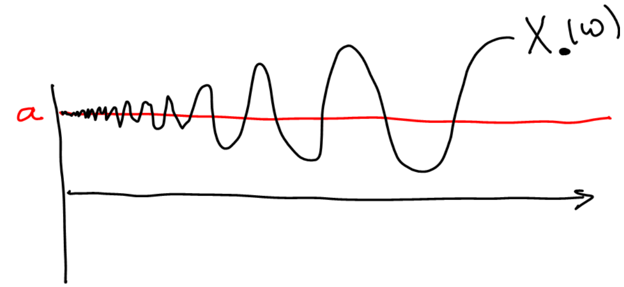



**Fig. 29.1.** A sample point, $\omega\in\Omega$, where $T_A(\omega)=0$ with $A=\{a\}\subset\mathbb{R}$.

**Proof.** 1. By definition, $D_A(\omega)<t$ iff $X_s(\omega)\in A$ for some $s<t$, which by right continuity of $X$ happens iff $X_s(\omega)\in A$ for some $s<t$ with $s\in\mathbb{Q}$. Therefore,

$$\{D_A<t\}=\bigcup\nolimits_{\mathbb{Q}\ni s<t}X_s^{-1}(A)\in\mathcal{B}_{t-}\subset\mathcal{B}_t.$$

2. If $A \subset S$ is closed and $T_A(\omega) < \infty$ (or $D_A(\omega) < \infty$), there exists $t_n > 0$ ($t_n \geq 0$) such that $X_{t_n} \in A$ and $t_n \downarrow T_A(\omega)$. Since $X$ is right continuous and $A$ is closed, $X_{t_n} \to X_{T_A(\omega)} \in A$ ($X_{t_n} \to X_{D_A}(\omega)$).

For the rest of the argument we will now assume that $X$ is a continuous process and $A$ is a closed subset of $S$.

3. Observe that $D_A(\omega) > t$ iff $X_{[0,t]}(\omega) \cap A = \emptyset$. Since $X$ is continuous, $X_{[0,t]}(\omega)$ is a compact subset of $S$ and therefore

$$\varepsilon := \rho\left(X_{[0,t]}(\omega), A\right) > 0$$

where

$$\rho(A,B) := \inf\{\rho(a,b) : a \in A \text{ and } b \in B\}.$$

Hence we have shown,

$$\begin{aligned}\{D_A > t\} &= \cup_{n=1}^{\infty}\left\{\omega : \rho\left(X_{[0,t]}(\omega), A\right) \geq 1/n\right\} \\ &= \cup_{n=1}^{\infty} \cap_{s\in\mathbb{Q}\cap[0,t]} \{\rho(X_s, A) \geq 1/n\} \in \mathcal{B}_t\end{aligned}$$

wherein we have used $\rho(\cdot, A) : S \to \mathbb{R}_+$ is continuous and hence measurable. As $\{D_A \leq t\} = \{D_A > t\}^c \in \mathcal{B}_t$ for all $t$, we have shown $D_A$ is an stopping time.

4. Suppose $t > 0$. Then $T_A(\omega) > t$ iff $X_{(0,t]}(\omega) \cap A = \emptyset$ which happens iff for all $\delta \in (0,t)$ we have $X_{[\delta,t]}(\omega) \cap A = \emptyset$ or equivalently iff for all $\delta \in (0,t)$, $\varepsilon := \rho\left(X_{[\delta,t]}(\omega), A\right) > 0$. Using these observations we find,

$$\begin{aligned}\{T_A > t\} &= \cap_{n>1/t} \cup_{m=1}^{\infty}\left\{\omega : \rho\left(X_{[1/n,t]}(\omega), A\right) \geq 1/m\right\} \\ &= \cap_{n>1/t} \cup_{m=1}^{\infty} \cap_{s\in\mathbb{Q}\cap[1/n,t]} \{\rho(X_s, A) \geq 1/m\} \in \mathcal{B}_t.\end{aligned}$$

This shows $\{T_A \leq t\} = \{T_A > t\}^c \in \mathcal{B}_t$ for all $t > 0$. Since, for $t > 0$, $\{T_A < t\} = \cup_{s\in\mathbb{Q}(0,t)} \{T_A \leq s\} \in \mathcal{B}_t$ we see that $T_A$ is an optional time. ■

The only thing keeping $T_A$ from being a stopping time in item 4 above is the fact that $\{T_A = 0\} \in \mathcal{B}_{0+}$ rather than $\{T_A = 0\} \in \mathcal{B}_0$. It should be clear that, in general, $\{T_A = 0\} \notin \mathcal{B}_0$ for $\{T_A = 0\} \in \mathcal{B}_0$ iff $1_{T_A=0} = f(X_0)$ for some measurable function, $f : S \to \{0,1\} \subset \mathbb{R}$. But it is clearly impossible to determine whether $T_A = 0$ by only observing $X_0$.

**Notation 29.12** *If $\tau : \Omega \to [0,\infty]$ is a random $\mathcal{B}_\infty$ – measurable time, let*

$$\mathcal{B}_\tau := \{A \in \mathcal{B}_\infty : \{\tau \leq t\} \cap A \in \mathcal{B}_t \text{ for all } t \in [0,\infty]\}. \tag{29.4}$$

*and*

$$\mathcal{B}_{\tau+} := \{A \in \mathcal{B}_\infty : A \cap \{\tau < t\} \in \mathcal{B}_t \text{ for all } t \leq \infty\}.$$

**Exercise 29.3.** If $\tau$ is a stopping time then $\mathcal{B}_\tau$ is a sub-$\sigma$ – algebra of $\mathcal{B}_\infty$ and if $\tau$ is an optional time then $\mathcal{B}_{\tau+}$ is a sub-$\sigma$ – algebra of $\mathcal{B}_\infty$.

**Exercise 29.4.** Suppose $\tau : \Omega \to [0,\infty]$ is the constant function, $\tau = s$, show $\mathcal{B}_\tau = \mathcal{B}_s$ and $\mathcal{B}_{\tau+} = \cap_{t>s}\mathcal{B}_t =: \mathcal{B}_{s+}$ so that the notation introduced in Notation 29.12 is consistent with the previous meanings of $\mathcal{B}_s$ and $\mathcal{B}_{s+}$.

**Exercise 29.5.** Suppose that $\tau$ is an optional time and let

$$\mathcal{B}_\tau^+ := \left\{A \in \mathcal{B}_\infty : A \cap \{\tau \leq t\} \in \mathcal{B}_{t+} = \mathcal{B}_t^+ \text{ for all } t \leq \infty\right\}.$$

Show $\mathcal{B}_{\tau+} = \mathcal{B}_\tau^+$. Hence $\mathcal{B}_{\tau+}$ is precisely the stopped $\sigma$ – algebra of the stopping time, $\tau$, relative to the filtration $\left\{\mathcal{B}_t^+\right\}$.

**Lemma 29.13.** *Suppose $T : \Omega \to [0,\infty]$ is a random time.*

1. *If $T$ is a $\{\mathcal{B}_t\}$ – stopping time, then $T$ is $\mathcal{B}_T$ – measurable.*
2. *If $T$ is a $\{\mathcal{B}_t\}$ – optional time, then $T$ is $\mathcal{B}_{T+} = \mathcal{B}_T^+$ – measurable.*

**Proof.** Because of Exercise 29.5, it suffices to prove the first assertion. For all $s,t \in \mathbb{R}_+$, we have

$$\{T \leq t\} \cap \{T \leq s\} = \{T \leq s \wedge t\} \in \mathcal{B}_{s\wedge t} \subset \mathcal{B}_s.$$

This shows $\{T \leq t\} \in \mathcal{B}_T$ for all $t \in \mathbb{R}_+$ and therefore that $T$ is $\mathcal{B}_T$ – measurable. ■

**Lemma 29.14.** *If $\tau$ is a $\{\mathcal{B}_t\}$ – stopping time and $X_t : \Omega \to S$ is a $\{\mathcal{B}_t\}$ – progressively measurable process, then $X_\tau$ defined on $\{\tau < \infty\}$ is $(\mathcal{B}_\tau)_{\{\tau<\infty\}}/\mathcal{S}$ – measurable. Similarly, if $\tau$ is a $\{\mathcal{B}_t\}$ – optional time and $X_t : \Omega \to S$ is a $\left\{\mathcal{B}_t^+\right\}$ – progressively measurable process, then $X_\tau$ defined on $\{\tau < \infty\}$ is $(\mathcal{B}_{\tau+})_{\{\tau<\infty\}}/\mathcal{S}$ – measurable.*

**Proof.** In view of Proposition 29.9 and Exercise 29.5, it suffices to prove the first assertion. For $T \in \mathbb{R}_+$, let $\psi_T : \{\tau \leq T\} \to [0,T] \times \Omega$ be defined by $\psi_T(\omega) = (\tau(\omega), \omega)$ and $\varphi^T : [0,T] \times \Omega \to S$ be defined by $\varphi^T(t,\omega) = X_t(\omega)$. By definition $\varphi^T$ is $\mathcal{B}_{[0,T]} \otimes \mathcal{B}_T/\mathcal{S}$ – measurable. Since, for all $A \in \mathcal{B}_T$ and $a \in [0,T]$,

$$\psi_T^{-1}([0,a] \times A) = \{\tau \leq a\} \cap A \in (\mathcal{B}_T)_{\{\tau\leq T\}},$$

it follows that $\psi_T$ is $(\mathcal{B}_T)_{\{\tau\leq T\}}/\mathcal{B}_{[0,T]}\otimes\mathcal{B}_T$ – measurable and therefore, $\varphi^T \circ \psi_T : \{\tau \leq T\} \to S$ is $(\mathcal{B}_T)_{\{\tau\leq T\}}/\mathcal{S}$ – measurable.

For $V \in \mathcal{S}$ and $T \in \mathbb{R}_+$,

$$\begin{aligned}X_\tau^{-1}(A) \cap \{\tau \leq T\} &= \left\{\omega \in \Omega : \tau(\omega) \leq T \text{ and } X_{\tau(\omega)}(\omega) \in A\right\} \\ &= \left\{\omega \in \Omega : \tau(\omega) \leq T \text{ and } \varphi^T \circ \psi_T(\omega) \in A\right\} \\ &= \{\tau \leq T\} \cap \left\{\varphi^T \circ \psi_T \in A\right\} \in (\mathcal{B}_T)_{\{\tau\leq T\}} \subset \mathcal{B}_T.\end{aligned}$$

This is true for arbitrary $T \in \mathbb{R}_+$ we conclude that $X_\tau^{-1}(A) \in \mathcal{B}_\tau$ and since, by definition, $X_\tau^{-1}(A) \subset \{\tau < \infty\}$, it follows that $X_\tau^{-1}(A) \in (\mathcal{B}_\tau)_{\{\tau<\infty\}}$. This completes the proof since $A \in \mathcal{S}$ was arbitrary. ■

**Lemma 29.15 (Properties of Optional/Stopping times).** *Let $T$ and $S$ be optional times and $\theta > 0$. Then;*

1. *$T + \theta$ is a stopping time.*
2. *$T + S$ is an optional time.*
3. *If $T > 0$ and $T$ is a stopping time then $T + S$ is again a stopping time.*
4. *If $T > 0$ and $S > 0$, then $T + S$ is a stopping time.*
5. *If we further assume that $S$ and $T$ are stopping times, then $T \wedge S$, $T \vee S$, and $T + S$ are stopping times.*
6. *If $\{T_n\}_{n=1}^\infty$ are optional times, then*
$$\sup_{n\geq 1} T_n, \quad \inf_{n\geq 1} T_n, \quad \liminf_{n\to\infty} T_n, \quad \text{and} \quad \limsup_{n\to\infty} T_n$$
*are all optional times. If $\{T_n\}_{n=1}^\infty$ are stopping times, then $\sup_{n\geq 1} T_n$ is a stopping time.*

**Proof.** 1. This follows from the observation that

$$\{T + \theta \leq t\} = \{T \leq t - \theta\} \in \mathcal{B}_{(t-\theta)+} \subset \mathcal{B}_t.$$

Notice that if $t < \theta$, then $\{T + \theta \leq t\} = \emptyset \in \mathcal{B}_0$.

2. – 4. For item 2., if $\tau > 0$, then

$$\{T + S < \tau\} = \cup \{T < t, S < s : s, t \in \mathbb{Q} \cap (0, \tau] \text{ with } s + t < \tau\} \in \mathcal{B}_{\tau-} \subset \mathcal{B}_\tau$$

and if $\tau = 0$, then $\{T + S < 0\} = \emptyset \in \mathcal{B}_0$. If $T > 0$ and $T$ is a stopping time and $\tau > 0$, then

$$\{T + S \leq \tau\} = \{S = 0, T \leq \tau\} \cup \{0 < S, S + T \leq \tau\}$$

and $\{S = 0, T \leq \tau\} \in \mathcal{B}_\tau$. Hence it suffices to show $\{0 < S, S + T \leq \tau\} \in \mathcal{B}_\tau$. To this end, observe that $0 < S$ and $S + T \leq \tau$ happens iff there exists $m \in \mathbb{N}$ such that for all $n \geq m$, there exists a $r = r_n \in \mathbb{Q}$ such that

$$0 < r_n < S < r_n + 1/n < \tau \text{ and } T \leq \tau - r_n.$$

Indeed, if the latter condition holds, then $S + T \leq \tau - r_n + (r_n + 1/n) = \tau + 1/n$ for all $n$ and therefore $S + T \leq \tau$. Thus we have shown

$$\{S + T \leq \tau\} = \cup_{m\in\mathbb{N}} \cap_{n\geq m} \cup \{r < S < r + 1/n, T \leq \tau - r : 0 < r < r + 1/n < \tau\}$$

which is in $\mathcal{B}_{\tau-} \subset \mathcal{B}_\tau$. In showing $\{0 < S, S + T \leq \tau\} \in \mathcal{B}_\tau$ we only need for $S$ and $T$ to be optional times and so if $S > 0$ and $T > 0$, then

$$\{T + S \leq \tau\} = \{0 < S, S + T \leq \tau\} \in \mathcal{B}_{\tau-} \subset \mathcal{B}_\tau.$$

5. If $T$ and $S$ are stopping times and $\tau \geq 0$, then

$$\{T \wedge S \leq \tau\} = \{T \leq \tau\} \cup \{S \leq \tau\} \in \mathcal{B}_\tau,$$
$$\{T \vee S \leq \tau\} = \{T \leq \tau\} \cap \{S \leq \tau\} \in \mathcal{B}_\tau,$$

and

$$\begin{aligned}\{S + T > \tau\} &= \{T = 0, S > \tau\} \cup \{0 < T, T + S > \tau\} \\ &= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \{T \geq \tau, T + S > \tau\} \\ &= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \\ &\quad \cup \{S = 0, T > \tau\} \cup \{S > 0, T \geq \tau\}.\end{aligned}$$

The first, third, and fourth events are easily seen to be in $\mathcal{B}_\tau$. As for the second event,

$$\{0 < T < \tau, T + S > \tau\} = \cup \{r < T < \tau, S > \tau - r : r \in \mathbb{Q} \text{ with } 0 < r < \tau\}.$$

6. We have

$$\left\{\sup_n T_n \leq t\right\} = \cap_{n=1}^\infty \{T_n \leq t\},$$
$$\left\{\inf_n T_n < t\right\} = \cup_{n=1}^\infty \{T_n < t\}$$

which shows that $\sup_n T_n$ is a stopping time if each $T_n$ is a stopping time and that $\inf_n T_n$ is optional if each $T_n$ is optional. Moreover, if each $T_n$ is optional, then $T_n$ is a $\mathcal{B}_{t+}$ stopping time and hence $\sup_n T_n$ is an $\mathcal{B}_{t+}$ stopping time and hence $\sup_n T_n$ is an $\mathcal{B}_t$ optional time, wherein we have used Proposition 29.9 twice. ■

**Lemma 29.16 (Stopped $\sigma$ – algebras).** *Suppose $\sigma$ and $\tau$ are stopping times.*

1. *$\mathcal{B}_\tau = \mathcal{B}_t$ on $\{\tau = t\}$.*
2. *If $t \in [0, \infty]$, then $\tau \wedge t$ is $\mathcal{B}_t$ – measurable.*
3. *If $\sigma \leq \tau$, then $\mathcal{B}_\sigma \subset \mathcal{B}_\tau$.*
4. *$(\mathcal{B}_\sigma)_{\{\sigma\leq\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$ and in particular $\{\sigma \leq \tau\}$, $\{\sigma < \tau\}$, $\{\tau \leq \sigma\}$, and $\{\tau < \sigma\}$ are all in $\mathcal{B}_{\sigma\wedge\tau}$.*
5. *$(\mathcal{B}_\sigma)_{\{\sigma<\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$.*
6. *$\mathcal{B}_\sigma \cap \mathcal{B}_\tau = \mathcal{B}_{\sigma\wedge\tau}$.*
7. *If $\mathbb{D}$ is a countable set and $\tau : \Omega \to \mathbb{D} \subset [0, \infty]$ is a function, then $\tau$ is a stopping time iff $\{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$.*
8. *If the range of $\tau$ is a countable subset, $\mathbb{D} \subset [0, \infty]$, then $A \subset \Omega$ is in $\mathcal{B}_\tau$ iff $A \cap \{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$.*
9. *If the range of $\tau$ is a countable subset, $\mathbb{D} \subset [0, \infty]$, then a function $f : \Omega \to \mathbb{R}$ is $\mathcal{B}_\tau$ – measurable iff $1_{\{\tau=t\}} f$ is $\mathcal{B}_t$ – measurable for all $t \in \mathbb{D}$.*

*Moreover, all of the above results hold if $\sigma$ and $\tau$ are optional times provided every occurrence of the letter $\mathcal{B}$ is replaced by $\mathcal{B}^+$.*

**Proof.** Recall from Definition 16.27 that if $\mathcal{G}$ is a $\sigma$ – algebra on $\Omega$ and $A \subset \Omega$, then $\mathcal{G}_A := \{B \cap A : B \in \mathcal{G}\}$ – a sub – $\sigma$ – algebra of $2^A$. Moreover if $\mathcal{G}$ and $\mathcal{F}$ are two $\sigma$ – algebras on $\Omega$ and $A \in \mathcal{G} \cap \mathcal{F}$, then (by definition) $\mathcal{G} = \mathcal{F}$ on $A$ iff $\mathcal{G}_A = \mathcal{F}_A$.

1. If $A \in \mathcal{B}_\tau$, then
$$A \cap \{\tau = t\} = A \cap \{\tau \le t\} \cap \{\tau < t\}^c \in \mathcal{B}_t.$$
Conversely if $A \in \mathcal{B}_t$ and $s \in \mathbb{R}_+$,
$$A \cap \{\tau = t\} \cap \{\tau \le s\} = \begin{cases} \emptyset & \text{if } s < t \\ A \cap \{\tau = t\} & \text{if } s \ge t \end{cases}$$
from which it follows that $A \cap \{\tau = t\} \in \mathcal{B}_\tau$.
2. To see $\tau \wedge t$ is $\mathcal{B}_t$ – measurable simply observe that
$$\{\tau \wedge t \le s\} = \begin{cases} \Omega \in \mathcal{B}_t & \text{if } t \le s \\ \{\tau \le s\} \in \mathcal{B}_s \subset \mathcal{B}_t & \text{if } t > s \end{cases}$$
and hence $\{\tau \wedge t \le s\} \in \mathcal{B}_t$ for all $s \in [0, \infty]$.
3. If $A \in \mathcal{B}_\sigma$ and $\sigma \le \tau$, then
$$A \cap \{\tau \le t\} = [A \cap \{\sigma \le t\}] \cap \{\tau \le t\} \in \mathcal{B}_t$$
for all $t \le \infty$ and therefore $A \in \mathcal{B}_\tau$.
4. If $A \in \mathcal{B}_\sigma$ then $A \cap \{\sigma \le \tau\}$ is the generic element of $(\mathcal{B}_\sigma)_{\{\sigma \le \tau\}}$. We now have
$$\begin{aligned}(A \cap \{\sigma \le \tau\}) \cap \{\tau \wedge \sigma \le t\} &= (A \cap \{\sigma \le \tau\}) \cap \{\sigma \le t\} \\ &= (A \cap \{\sigma \wedge t \le \tau \wedge t\}) \cap \{\sigma \le t\} \\ &= (A \cap \{\sigma \le t\}) \cap \{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{B}_t\end{aligned}$$
since $(A \cap \{\sigma \le t\}) \in \mathcal{B}_t$ and $\sigma \wedge t$ and $\tau \wedge t$ are $\mathcal{B}_t$ – measurable and hence $\{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{B}_t$. Since $\Omega \in \mathcal{B}_\sigma$, it follows from what we have just proved that $\{\sigma \le \tau\} = \Omega \cap \{\sigma \le \tau\} \in \mathcal{B}_{\sigma\wedge\tau}$ and hence also $\{\tau < \sigma\} = \{\sigma \le \tau\}^c \in \mathcal{B}_{\sigma\wedge\tau}$. By symmetry we may also conclude that $\{\tau \le \sigma\}$ and $\{\sigma < \tau\}$ are in $\mathcal{B}_{\sigma\wedge\tau}$.
5. By item 4., if $A \in \mathcal{B}_\sigma$, then
$$A \cap \{\sigma < \tau\} = A \cap \{\sigma \le \tau\} \cap \{\sigma < \tau\} \in \mathcal{B}_{\sigma\wedge\tau}.$$
6. Since $\sigma \wedge \tau$ is a stopping time which is no larger than either $\sigma$ or $\tau$, it follows that from item 2. that $\mathcal{B}_{\sigma\wedge\tau} \subset \mathcal{B}_\sigma \cap \mathcal{B}_\tau$. Conversely, if $A \in \mathcal{B}_\sigma \cap \mathcal{B}_\tau$ then
$$\begin{aligned}A \cap \{\sigma \wedge \tau \le t\} &= A \cap [\{\sigma \le t\} \cup \{\tau \le t\}] \\ &= [A \cap \{\sigma \le t\}] \cup [A \cap \{\tau \le t\}] \in \mathcal{B}_t\end{aligned}$$
for all $t \le \infty$. From this it follows that $A \in \mathcal{B}_{\sigma\wedge\tau}$.
7. If $\tau$ is a stopping time and $t \in \mathbb{D}$, then $\{\tau = t\} = \{\tau \le t\} \setminus [\cup_{\mathbb{D}\ni s<t} \{\tau \le s\}] \in \mathcal{B}_t$. Conversely if $\{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$ and $s \in \mathbb{R}_+$, then
$$\{\tau \le s\} = \cup_{\mathbb{D}\ni t\le s} \{\tau = t\} \in \mathcal{B}_s$$
showing $\tau$ is a stopping time.
8. If $A \cap \{\tau = t\} \in \mathcal{B}_t$ for all $t \in \mathbb{D}$, then for any $s \le \infty$,
$$A \cap \{\tau \le s\} = \cup_{\mathbb{D}\ni t\le s} [A \cap \{\tau = t\}] \in \mathcal{B}_s$$
which shows $A \in \mathcal{B}_\tau$. Conversely if $A \in \mathcal{B}_\tau$ and $t \in \mathbb{D}$, then
$$A \cap \{\tau = t\} = [A \cap \{\tau \le t\}] \setminus [\cup_{\mathbb{D}\ni s<t} (A \cap \{\tau \le s\})] \in \mathcal{B}_t.$$
9. If $f : \Omega \to \mathbb{R}$ is $\mathcal{B}_\tau$ measurable, then $f$ is a limit of $\mathcal{B}_\tau$ – simple functions, say $f_n \to f$. By item 7. it easily follows that $1_{\{\tau=t\}} f_n$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$ and therefore $1_{\{\tau=t\}} f = \lim_{n\to\infty} 1_{\{\tau=t\}} f_n$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$.
Conversely if $f : \Omega \to \mathbb{R}$ is a function such that $1_{\{\tau=t\}} f$ is $\mathcal{B}_t$ – measurable for each $t \in \mathbb{D}$, then for every $A \in \mathcal{B}_\mathbb{R}$ with $0 \notin A$ we have
$$\{\tau = t\} \cap \{f \in A\} = \left\{1_{\{\tau=t\}} f \in A\right\} \in \mathcal{B}_t \text{ for all } t \in \mathbb{D}.$$
Hence it follows by item 7. that $\{f \in A\} \in \mathcal{B}_\tau$. Similalry,
$$\{\tau = t\} \cap \{f = 0\} = \left\{1_{\{\tau=t\}} f = 0\right\} \cap \{\tau = t\} \in \mathcal{B}_t \text{ for all } t \in \mathbb{D}$$
and so again $\{f = 0\} \in \mathcal{B}_\tau$ by item 7. This suffices to show that $f$ is $\mathcal{B}_\tau$ – measurable.

■

**Corollary 29.17.** *If $\sigma$ and $\tau$ are stopping times and $F$ is a $\mathcal{B}_\sigma$ – measurable function then $1_{\{\sigma\le\tau\}} F$ and $1_{\{\sigma<\tau\}} F$ are $\mathcal{B}_{\sigma\wedge\tau}$ – measurable.*

**Proof.** If $F = 1_A$ with $A \in \mathcal{B}_\sigma$, then the assertion follows from items 4. and 5. from Lemma 29.16. By linearity, the assertion holds if $F$ is a $\mathcal{B}_\sigma$ – measurable simple function and then, by taking limits, for all $\mathcal{B}_\sigma$ – measurable functions. ■

**Lemma 29.18 (Optional time approximation lemma).** *Let $\tau$ be a $\{\mathcal{B}_t\}_{t\geq 0}$ – optional time and for $n \in \mathbb{N}$, let $\tau_n : \Omega \to [0,\infty]$ be defined by*

$$\tau_n := \frac{1}{2^n}\left[2^n \tau\right] = \infty 1_{\tau=\infty} + \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\frac{k-1}{2^n} \leq \tau < \frac{k}{2^n}}. \tag{29.5}$$

*Then $\{\tau_n\}_{n=1}^{\infty}$ are stopping times such that;*

1. *$\tau_n \downarrow \tau$ as $n \to \infty$,*
2. *$\mathcal{B}_\tau^+ \subset \mathcal{B}_{\tau_n}$ for all $n$, and*
3. *$\{\tau_n = \infty\} = \{\tau = \infty\}$ for all $n$.*

**Proof.** If $A \in \mathcal{B}_\tau^+$ then

$$\begin{aligned} A \cap \left\{\tau_n = k2^{-n}\right\} &= A \cap \left\{(k-1)\, 2^{-n} \leq \tau < k2^{-n}\right\} \\ &= \left[A \cap \left\{\tau < k2^{-n}\right\}\right] \setminus \left\{\tau < (k-1)\, 2^{-n}\right\} \in \mathcal{B}_{k2^{-n}}. \end{aligned}$$

Taking $A = \Omega$ in this equation shows $\{\tau_n = k2^{-n}\} \in \mathcal{B}_{k2^{-n}}$ for all $k \in \mathbb{N}$ and so is a stopping time by Lemma 29.16. Moreover this same lemma shows that $A \in \mathcal{B}_{\tau_n}$. The fact that $\tau_n \downarrow \tau$ as $n \to \infty$ and $\tau_n = \infty$ iff $\tau = \infty$ should be clear. ■

## 29.3 Filtration considerations

For this sction suppose that $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ is a given filtered probability space.

**Notation 29.19 (Null sets)** *Let $\mathcal{N}^P := \{N \in \mathcal{B} : P(N) = 0\}$ – be the collection of **null sets** of $P$.*

**Definition 29.20.** *If $\mathcal{A} \subset \mathcal{B}$ is a sub-sigma-algebra of $\mathcal{B}$, then the **augmentation of** $\mathcal{A}$ is the $\sigma$ – algebra, $\bar{\mathcal{A}} := \mathcal{A} \vee \mathcal{N} := \sigma(\mathcal{A} \cup \mathcal{N})$.*

**Definition 29.21 (Usual hypothesis).** *A filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, is said to satisfy the **weak usual hypothesis** if:*

1. *For each $t \in \mathbb{R}_+$, $\mathcal{N}^P \subset \mathcal{B}_t$, i.e. $\mathcal{B}_t$ contains all of the $P$ – null sets.*
2. *The filtration, $\{\mathcal{B}_t\}_{t\in\mathbb{R}_+}$ is **right continuous**, i.e. $\mathcal{B}_{t+} = \mathcal{B}_t$.*

*If in addition, $(\Omega, \mathcal{B}, P)$ is complete (i.e. if $N \in \mathcal{N}^P$, $A \subset N$, then $A \in \mathcal{N}^P$), then we say $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ satisfies the **usual hypothesis.***

It is always possible to make an arbitrary filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, into one satisfying the (weak) usual hypothesis by "augmenting" the filtration by the null sets and taking the "right continuous extension." We are going to develop these two concepts now. (For even more information on the usual hypothesis, see [37, pages 34-36.].)

**Lemma 29.22 (Augmentation lemma).** *Continuing the notation in Definition 29.20, we have*

$$\bar{\mathcal{A}} := \{B \in \mathcal{B} : \exists\, A \in \mathcal{A} \ni A \triangle B \in \mathcal{N}\}. \tag{29.6}$$

**Proof.** Let $\mathcal{G}$ denote the right side of Eq. (29.6). If $B \in \mathcal{G}$ and $A \in \mathcal{A}$ such that $N := A \triangle B \in \mathcal{N}$, then

$$B = [A \cap B] \cup [A \setminus B] = [A \setminus (A \setminus B)] \cup [B \setminus A]. \tag{29.7}$$

Since $A \setminus B \subset N$ and $B \setminus A \subset N$ implies $A \setminus B$ and $B \setminus A$ are in $\mathcal{N}$, it follows that $B \in \mathcal{A} \vee \mathcal{N} = \bar{\mathcal{A}}$. Thus we have shown, $\mathcal{G} \subset \bar{\mathcal{A}}$. Since it is clear that $\mathcal{A} \subset \mathcal{G}$ and $\mathcal{N} \subset \mathcal{G}$, to finish the proof it suffices to show $\mathcal{G}$ is a $\sigma$ – algebra. For if we do this, then $\bar{\mathcal{A}} = \mathcal{A} \vee \mathcal{N} \subset \mathcal{G}$.

Since $A^c \triangle B^c = A \triangle B$, we see that $\mathcal{G}$ is closed under complementation. Moreover, if $B_j \in \mathcal{G}$, there exists $A_j \in \mathcal{A}$ such that $A_j \triangle B_j \in \mathcal{N}$ for all $j$. So letting $A = \cup_j A_j \in \mathcal{A}$ and $B = \cup_j B_j \in \mathcal{B}$, we have

$$\mathcal{B} \ni A \triangle B \subset \cup_j \left[A_j \triangle B_j\right] \in \mathcal{N}$$

from which we conclude that $A \triangle B \in \mathcal{N}$ and hence $B \in \mathcal{G}$. This shows that $\mathcal{G}$ is closed under countable unions, complementation, and contains $\mathcal{A}$ and hence the empty set and $\Omega$, thus $\mathcal{G}$ is a $\sigma$ – algebra. ■

**Lemma 29.23 (Commutation lemma).** *If $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$ is a filtered probability space, then $\bar{\mathcal{B}}_{t+} = \overline{\mathcal{B}_{t+}}$. In words the augmentation procedure and the right continuity extension procedure commute.*

**Proof.** Since for any $s > t$, $\mathcal{B}_{t+} \subset \mathcal{B}_s$ it follows that $\overline{\mathcal{B}_{t+}} \subset \bar{\mathcal{B}}_s$ and therefore that

$$\overline{\mathcal{B}_{t+}} \subset \cap_{s>t} \bar{\mathcal{B}}_s = \bar{\mathcal{B}}_{t+}.$$

Conversely if $B \in \bar{\mathcal{B}}_{t+} = \cap_{s>t} \bar{\mathcal{B}}_s$ and $t_n > t$ such that $t_n \downarrow 0$, then for each $n \in \mathbb{N}$ there exists $A_n \in \mathcal{B}_{t_n}$ such that $A_n \triangle B \in \mathcal{N}$. We will now show that $B \in \overline{\mathcal{B}_{t+}}$, by showing $B \triangle A \in \mathcal{N}$ where

$$A := \{A_n \text{ i.o.}\} = \cap_{m\in\mathbb{N}} \cup_{n\geq m} A_n \in \mathcal{B}_{t+}.$$

To prove this let $A'_m := \cup_{n\geq m} A_n$ so that $A'_n \downarrow A$ as $n \uparrow \infty$. Then

$$B \triangle A = B \triangle [\cap_m A'_m] = (B \setminus [\cap_m A'_m]) \cup ([\cap_m A'_m] \setminus B)$$
$$\subset [\cup_m (B \setminus A'_m)] \cup (A'_1 \setminus B) \in \mathcal{N}$$

because measurable subsets of elements in $\mathcal{N}$ are still in $\mathcal{N}$, $\mathcal{N}$ is closed under countable unions,

$$B \setminus A'_m \subset B \setminus A_m \subset B \triangle A_m \in \mathcal{N}, \text{ and}$$
$$A'_1 \setminus B = \cup_{n=1}^\infty [A_n \setminus B] \subset \cup_{n=1}^\infty [A_n \triangle B] \in \mathcal{N}.$$

■

### 29.3.1 ***More Augmentation Results (This subsection neeed serious editing.)

In this subsection we generalize the augmentation results above to the setting where we adjoin our favorite collection of "null like" sets.

**Definition 29.24.** *Suppose $(\Omega, \mathcal{B})$ is a measurable space. A collection of subsets, $\mathcal{N} \subset \mathcal{B}$ is a **null like collection in** $\mathcal{B}$ if; 1) $\mathcal{N}$ is closed under countable union, and 2) if $A \in \mathcal{B}$ and there exists $N \in \mathcal{N}$ such that $A \subset N$, then $A \in \mathcal{N}$.*

*Example 29.25.* Let $\{P_i\}_{i\in I}$ be any collection of probability measures on a measurable space, $(\Omega, \mathcal{B})$. Then

$$\mathcal{N} := \{N \in \mathcal{B} : P_i(N) = 0 \text{ for all } i \in I\}$$

is a null like collections of subsets in $\mathcal{B}$.

*Example 29.26.* If $\mathcal{N}$ is a null like collection in $\mathcal{B}$, then

$$\bar{\mathcal{N}} := \left\{A \subset 2^\Omega : A \subset N \text{ for some } N \in \mathcal{N}\right\}$$

is a null like collection in $2^\Omega$.

*Example 29.27.* Let $\{P_i\}_{i\in I}$ be any collection of probability measures on a measurable space, $(\Omega, \mathcal{B})$. Then

$$\mathcal{N} := \left\{N \subset 2^\Omega : \exists\, B \in \mathcal{B} \ni P_i(B) = 0 \text{ and } N \subset B \text{ for all } i \in I\right\}$$

is a null like collections of subsets of $2^\Omega$. Similarly,

$$\mathcal{N} := \left\{N \subset 2^\Omega : \exists\, B_i \in \mathcal{B} \ni P_i(B_i) = 0 \text{ and } N \subset B_i \text{ for all } i \in I\right\}$$

is a null like collections of subsets of $2^\Omega$. These two collections are easily seen to be the same if $I$ is countable, otherwise they may be different.

*Example 29.28.* If $\mathcal{N}_i \subset \mathcal{B}$ are null like collections in $\mathcal{B}$ for all $i \in I$, then $\mathcal{N} := \cap_{i\in I}\mathcal{N}_i$ is another null like collection in $\mathcal{B}$. Indeed, if $B \in \mathcal{B}$ and $B \subset N \in \mathcal{N}$, then $B \subset N \in \mathcal{N}_i$ for all $i$ and therefore, $B \in \mathcal{N}_i$ for all $i$ and hence $B \in \mathcal{N}$. Moreover, it is clear that $\mathcal{N}$ is still closed under countable unions.

**Definition 29.29.** *If $(\Omega, \mathcal{B})$ is a measurable space, $\mathcal{A} \subset \mathcal{B}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N} \subset \mathcal{B}$ is a null like collection in $\mathcal{B}$, we say $\mathcal{A}^\mathcal{N} := \mathcal{A} \vee \mathcal{N}$ is the **augmentation of** $\mathcal{A}$ by $\mathcal{N}$.*

**Lemma 29.30 (Augmentation lemma).** *If $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N}$ is a null like collection in $\mathcal{B}$, then the augmentation of $\mathcal{A}$ by $\mathcal{N}$ is*

$$\mathcal{A}^\mathcal{N} := \{B \in \mathcal{B} : \exists\, A \in \mathcal{A} \ni A \triangle B \in \mathcal{N}\}. \tag{29.8}$$

**Proof.** Let $\mathcal{G}$ denote the right side of Eq. (29.8). If $B \in \mathcal{G}$ and $A \in \mathcal{A}$ such that $N := A \triangle B \in \mathcal{N}$, then

$$B = [A \cap B] \cup [A \setminus B] = [A \setminus (A \setminus B)] \cup [B \setminus A]. \tag{29.9}$$

Since $A \setminus B \subset N$ and $B \setminus A \subset N$ implies $A \setminus B$ and $B \setminus A$ are in $\mathcal{N}$, it follows that $B \in \mathcal{A} \vee \mathcal{N} = \mathcal{A}^\mathcal{N}$. Thus we have shown, $\mathcal{G} \subset \mathcal{A}^\mathcal{N}$. Since it is clear that $\mathcal{A} \subset \mathcal{G}$ and $\mathcal{N} \subset \mathcal{G}$, to finish the proof it suffices to show $\mathcal{G}$ is a $\sigma$ – algebra. For if we do this, then $\mathcal{A}^\mathcal{N} = \mathcal{A} \vee \mathcal{N} \subset \mathcal{G}$.

Since $A^c \triangle B^c = A \triangle B$, we see that $\mathcal{G}$ is closed under complementation. Moreover, if $B_j \in \mathcal{G}$, there exists $A_j \in \mathcal{A}$ such that $A_j \triangle B_j \in \mathcal{N}$ for all $j$. So letting $A = \cup_j A_j \in \mathcal{A}$ and $B = \cup_j B_j \in \mathcal{B}$, we have

$$\mathcal{B} \ni A \triangle B \subset \cup_j [A_j \triangle B_j] \in \mathcal{N}$$

from which we conclude that $A \triangle B \in \mathcal{N}$ and hence $B \in \mathcal{G}$. This shows that $\mathcal{G}$ is closed under countable unions, complementation, and contains $\mathcal{A}$ and hence the empty set and $\Omega$, thus $\mathcal{G}$ is a $\sigma$ – algebra. ■

**Lemma 29.31 (Commutation lemma).** *Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}\right)$ be a filtered space and $\mathcal{N} \subset \mathcal{B}$ be null like collection and for $\mathcal{G} \subset \mathcal{B}$, let $\bar{\mathcal{G}} := \mathcal{G} \vee \mathcal{N}$. Then $\bar{\mathcal{B}}_{t+} = \overline{\mathcal{B}_{t+}}$.*

**Proof.** Since for any $s > t$, $\mathcal{B}_{t+} \subset \mathcal{B}_s$ it follows that $\overline{\mathcal{B}_{t+}} \subset \bar{\mathcal{B}}_s$ and therefore that

$$\overline{\mathcal{B}_{t+}} \subset \cap_{s>t}\bar{\mathcal{B}}_s = \bar{\mathcal{B}}_{t+}.$$

Conversely if $B \in \bar{\mathcal{B}}_{t+} = \cap_{s>t}\bar{\mathcal{B}}_s$ and $t_n > t$ such that $t_n \downarrow 0$, then for each $n \in \mathbb{N}$ there exists $A_n \in \mathcal{B}_{t_n}$ such that $A_n \triangle B \in \mathcal{N}$. We will now show that $B \in \overline{\mathcal{B}_{t+}}$, by shown $B \triangle A \in \mathcal{N}$ where

$$A := \{A_n \text{ i.o.}\} = \cap_{m\in\mathbb{N}} \cup_{n\geq m} A_n \in \mathcal{B}_{t+}.$$

To prove this let $A'_m := \cup_{n\geq m} A_n$ so that $A'_n \downarrow A$ as $n\uparrow\infty$. Then

$$\begin{aligned}B \triangle A &= B \triangle [\cap_m A'_m] = (B\setminus[\cap_m A'_m]) \cup ([\cap_m A'_m]\setminus B) \\ &\subset [\cup_m (B\setminus A'_m)] \cup (A'_1 \setminus B) \in \mathcal{N}\end{aligned}$$

because measurable subsets of elements in $\mathcal{N}$ are still in $\mathcal{N}$, $\mathcal{N}$ is closed under countable unions,

$$\begin{aligned}&B\setminus A'_m \subset B\setminus A_m \subset B\triangle A_m \in \mathcal{N}, \text{ and} \\ &A'_1 \setminus B = \cup_{n=1}^\infty [A_n\setminus B] \subset \cup_{n=1}^\infty [A_n \triangle B] \in \mathcal{N}.\end{aligned}$$

■

**Corollary 29.32.** *Suppose $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$ and $\mathcal{N}$ is a null like collection in $\mathcal{B}$ with the additional property, that for all $N\in\mathcal{N}$ there exists $N'\in\mathcal{A}\cap\mathcal{N}$ such that $N\subset N'$. Then*

$$\mathcal{A}^{\mathcal{N}} := \{A\cup N : A\in\mathcal{A} \text{ and } N\in\mathcal{N}\}. \tag{29.10}$$

**Proof.** Let $\mathcal{G}$ denote the right side of Eq. (29.10). It is clear that $\mathcal{G}\subset\mathcal{A}^{\mathcal{N}}$. Conversely if $B\in\mathcal{A}^{\mathcal{N}}$, we know by Lemma 29.22 that there exists, $A\in\mathcal{A}$ such that $N := A\triangle B\in\mathcal{N}$. Since $C := A\setminus B\subset N$, $C\in\mathcal{N}$ and so by assumption there exists $N'\in\mathcal{A}\cap\mathcal{N}$ such that $C\subset N'$. Therefore, according to Eq. (29.7), we have

$$B = [A\setminus C]\cup[B\setminus A] = [A\setminus N']\cup[(A\cap N')\setminus C]\cup[B\setminus A].$$

Since $A\setminus N'\in\mathcal{A}$ and $[(A\cap N')\setminus C]\cup[B\setminus A]\in\mathcal{N}$, it follows that $B\in\mathcal{G}$. ■

*Example 29.33.* Let $\nu$ be a probability measure on $\mathbb{R}$. As in Notation **??**, let $P_\nu := \int_{\mathbb{R}} d\nu(x)\ P_x$ be the Wiener measure on $\Omega := C([0,\infty),\mathbb{R})$, $B_t:\Omega\to\mathbb{R}$ be the projection map, $B_t(\omega)=\omega(t)$, $\mathcal{B}_t = \sigma(B_s : s\leq t)$, and $\mathcal{N}_{t+}(\nu) := \{N\in\mathcal{B}_{t+} : P_\nu(N)=0\}$. Then by Corollary **??**, $\mathcal{B}_{t+} = \mathcal{B}_t\vee\mathcal{N}_{t+}(\nu)$. Hence if we let

$$\mathcal{N}(\nu) := \{N\in\mathcal{B} : P_\nu(N)=0\}$$

and

$$\bar{\mathcal{N}}(\nu) := \left\{B\subset 2^\Omega : B\subset N \text{ for some } N\in\mathcal{N}(\nu)\right\}$$

then $\mathcal{B}_{t+}\vee\mathcal{N}(\nu) = \mathcal{B}_t\vee\mathcal{N}(\nu) = \bar{\mathcal{B}}_t$ and $\mathcal{B}_{t+}\vee\bar{\mathcal{N}}(\nu) = \mathcal{B}_t\vee\bar{\mathcal{N}}(\nu)$ for all $t\in\mathbb{R}_+$. This shows that the augmented Brownian filtration, $\left\{\bar{\mathcal{B}}_t\right\}_{t\geq 0}$, is already right continuous.

**Definition 29.34.** *Recall from Proposition 6.77, if $(\Omega,\mathcal{B},P)$ is a probability space and $\bar{\mathcal{N}}^P := \{A\subset\Omega : P^*(A)=0\}$, then the **completion,** $\bar{P} := P^*|_{\mathcal{B}\vee\bar{\mathcal{N}}^P}$, is a probability measure on $\mathcal{B}\vee\bar{\mathcal{N}}^P$ which extends $P$ so that $\bar{P}(A)=0$ for all $A\in\bar{\mathcal{N}}^P$.*

Suppose that $(\Omega,\mathcal{B})$ is a measurable space and $\mathcal{N}\subset\mathcal{B}$ is a collection of sets closed under countable union and also satisfying, $A\in\mathcal{N}$ if $A\in\mathcal{B}$ with $A\subset N$ for some $N\in\mathcal{N}$. The main example that we will use below is to let $\{P_i\}_{i\in I}$ to be a collection of probability measures on $(\Omega,\mathcal{B})$ and then let

$$\mathcal{N} := \{N\in\mathcal{B} : P_i(N)=0 \text{ for all } i\in I\}.$$

Let us also observe that if $\mathcal{N}_i$ is a collection of null sets above for each $i\in I$, then $\mathcal{N} = \cap_i\mathcal{N}_i$ is also a collection of null sets. Indeed, if $B\in\mathcal{B}$ and $B\subset N\in\mathcal{N}$, then $B\subset N\in\mathcal{N}_i$ for all $i$ and therefore, $B\in N_i$ for all $i$ and hence $B\in\mathcal{N}$. Moreover, it is clear that $\mathcal{N}$ is still closed under countable unions.

**Lemma 29.35 (Augmentation).** *Let us not suppose that $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$. Then the **augmentation** of $\mathcal{A}$ by $\mathcal{N}$,*

$$\mathcal{A}^{\mathcal{N}} := \{B\in\mathcal{B} :\ \exists\ A\in\mathcal{A}\ \ni\ A\triangle B\in\mathcal{N}\},$$

*is a sub-sigma-algebra of $\mathcal{B}$. Moreover if $\mathcal{N} = \cap\mathcal{N}_i$ and $\mathcal{A}_i = \mathcal{A}^{\mathcal{N}_i}$ is the augmentation of $\mathcal{A}$ by $\mathcal{N}_i$, then*

$$\mathcal{A}^{\mathcal{N}} = \cap_i\mathcal{A}_i.$$

**Proof.** To prove this, first observe that

$$\begin{aligned}A\triangle B &= (A\setminus B)\cup(B\setminus A) = (A\cap B^c)\cup(B\cap A^c) \\ &= (B^c\setminus A^c)\cup(A^c\setminus B^c) = A^c\triangle B^c\end{aligned}$$

from which it follows that $\mathcal{A}^{\mathcal{N}}$ is closed under complementation. Moreover, if $B_j\in\mathcal{A}^{\mathcal{N}}$, then there exists $A_j\in\mathcal{A}$ such that $A_j\triangle B_j\in\mathcal{N}$ for all $j$. So letting $A=\cup_j A_j\in\mathcal{A}$ and $B=\cup_j B_j\in\mathcal{B}$, we have

$$\mathcal{B}\ni A\triangle B\subset\cup_j [A_j\Delta B_j]\in\mathcal{N}$$

from which we conclude that $A\triangle B\in\mathcal{N}$ and hence $B\in\mathcal{A}^{\mathcal{N}}$. This shows that $\mathcal{A}^{\mathcal{N}}$ is closed under unions and hence we have show $\mathcal{A}\subset\mathcal{A}^{\mathcal{N}}\subset\mathcal{B}$ and $\mathcal{A}^{\mathcal{N}}$ is sigma algebra.

???Now to prove the second assertion of this lemma. It is clear that if $\mathcal{N}\subset\mathcal{N}'$, then $\mathcal{A}^{\mathcal{N}}\subset\mathcal{A}^{\mathcal{N}'}$ and hence it follows that

$$\mathcal{A}^{\mathcal{N}}\subset\cap_i\mathcal{A}_i = \cap_i\mathcal{A}^{\mathcal{N}_i}.$$

For the converse inclusion, suppose that $B\in\cap_i\mathcal{A}^{\mathcal{N}_i}$ in which case there exists $A_i\in\mathcal{A}$ such that $B\triangle A_i\in\mathcal{N}_i$ for all $i\in I$. ■

Suppose that $(\Omega,\mathcal{B},P)$ is a probability space and $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$. The **augmentation,** $\mathcal{A}^P$, of $\mathcal{A}$ by the $P$ – null sets of $\mathcal{B}$ is the collection of sets:

$$\mathcal{A}^P := \{B\in\mathcal{B} : \exists\ A\in\mathcal{A}\ \ni\ P(B\Delta A)=0\}.$$

**Notation 29.36** *Let* $\bar{\mathcal{B}}^P$ *denote the completion of* $\mathcal{B}$*. Let* $\mathcal{B}_t^P$ *denote the augmentation of* $\mathcal{B}_t$ *by the* $P$ *– null subsets of* $\mathcal{B}$*. We also let* $\bar{\mathcal{B}}_t^P$ *denote the augmentation of* $\mathcal{B}_t$ *by the* $P$ *– null subsets of* $\bar{\mathcal{B}}_t^P$*.*

# Continuous time (sub)martingales

For this chapter, let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space as described in Chapter 29.

**Definition 30.1.** *Given a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, an adapted process, $X_t : \Omega \to \mathbb{R}$, is said to be a $(\{\mathcal{B}_t\})$* ***martingale*** *provided, $\mathbb{E}|X_t| < \infty$ for all $t$ and $\mathbb{E}[X_t - X_s|\mathcal{B}_s] = 0$ for all $0 \leq s \leq t < \infty$. If $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \geq 0$ or $\mathbb{E}[X_t - X_s|\mathcal{B}_s] \leq 0$ for all $0 \leq s \leq t < \infty$, then $X$ is said to be* ***submartingale*** *or* ***supermartingale*** *respectively.*

*Remark 30.2.* If $\sigma$ and $\tau$ are two $\{\mathcal{B}_t\}$ – optional times, then $\sigma \wedge \tau$ is as well. Indeed, it $t \in \mathbb{R}_+ \cup \{\infty\}$, then

$$\{\sigma \wedge \tau < t\} = \{\sigma < t\} \cup \{\tau < t\} \in \mathcal{B}_t.$$

The following results are of fundamental importance for a number or results in this chapter. The first result is a simple consequence of the optional sampling Theorem 20.40.

**Proposition 30.3 (Discrete optional sampling).** *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, and $\tau$ and $\sigma$ are two $\{\mathcal{B}_t\}_{t\geq 0}$ – stopping times with values in $\mathbb{D}_n := \left\{\frac{k}{2^n} : k \in \bar{\mathbb{N}}\right\}$ for some $n \in \mathbb{N}$. If $M := \sup_\omega \tau(\omega) < \infty$, then $X_\tau \in L^1(\Omega, \mathcal{B}_\tau, P)$, $X_{\sigma\wedge\tau} \in L^1(\Omega, \mathcal{B}_{\sigma\wedge\tau}, P)$, and*

$$X_{\sigma\wedge\tau} \leq \mathbb{E}[X_\tau|\mathcal{B}_\sigma].$$

**Proof.** For $k \in \bar{\mathbb{N}}$, let $\mathcal{F}_k := \mathcal{B}_{k2^{-n}}$ and $Y_k := X_{k2^{-n}}$. Then $\{Y_k\}_{k=0}^\infty$ is a $\{\mathcal{F}_k\}$ – submartingale and $2^n\sigma$, $2^n\tau$ are two $\bar{\mathbb{N}}$ – valued stopping times with $2^n\tau \leq 2^n M < \infty$. Therefore we may apply the optional sampling Theorem 20.40 to find

$$X_{\sigma\wedge\tau} = Y_{(2^n\sigma)\wedge(2^n\tau)} \leq \mathbb{E}[Y_{2^n\tau}|\mathcal{F}_{2^n\sigma}] = \mathbb{E}[X_\tau|\mathcal{B}_\sigma].$$

We have used $\mathcal{F}_{2^n\sigma} = \mathcal{B}_\sigma$ (you prove) in the last equality. ∎

**Lemma 30.4 ($L^1$ – convergence I).** *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$. If $t \in \mathbb{R}_+$ and $\{t_n\}_{n=1}^\infty \subset (t, \infty)$ such that $t_n \downarrow t$, then $\lim_{n\to\infty} X_{t_n}$ exists almost surely and in $L^1(P)$.*

**Proof.** Let $Y_n := X_{t_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{t_{-n}}$ for $n \in -\mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that $\inf \mathbb{E}Y_n \geq \mathbb{E}X_t$ and hence the result follows by by Theorem 20.79. ∎

**Lemma 30.5 ($L^1$ – convergence II).** *Suppose $\{X_t\}_{t\in\mathbb{R}_+}$ is a submartingale on a filtered probability space, $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\geq 0}, P\right)$, $\tau$ is a bounded $\{\mathcal{B}_t\}$ – optional time, and $\{\tau_n\}_{n=1}^\infty$ is the sequence of approximate stopping times defined in Lemma 29.18. Then $X_{\tau+} := \lim_{n\to\infty} X_{\tau_n}$ exists a.s. and in $L^1(P)$.*

**Proof.** Let $M := \sup_{\omega\in\Omega} \tau(\omega)$. If $m < n$, then $\tau_m$ and $\tau_n$ take values in $\mathbb{D}_n$, $0 \leq \tau_m \leq \tau_n$, and $\tau_n \leq M + 1$. Therefore by Proposition 30.3, $X_{\tau_m} \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_{\tau_m}]$ and $X_0 \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_0]$. Hence if we let $Y_n := X_{\tau_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{\tau_{-n}}$ for $n \in -\mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that

$$\inf_{n\in-\mathbb{N}} \mathbb{E}Y_n = \inf_{n\in\mathbb{N}} \mathbb{E}X_{\tau_n} \geq \mathbb{E}X_0 > -\infty.$$

The result now follows by an application of Theorem 20.79. ∎

**Lemma 30.6 ($L^1$ – convergence III).** *Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{\mathcal{B}_n\}_{n=1}^\infty$ is a decreasing sequence of sub– $\sigma$– algebras of $\mathcal{B}$. Then for all $Z \in L^1(P)$,*

$$\lim_{n\to\infty} \mathbb{E}[Z|\mathcal{B}_n] = \mathbb{E}[Z|\cap_{n=1}^\infty \mathcal{B}_n] \tag{30.1}$$

*where the above convergence is almost surely and in $L^1(P)$.*

**Proof.** This is a special case of Corollary 20.81 applied to the reverse martingale, $M_m = \mathbb{E}[Z|\mathcal{F}_m]$ where, for $m \in -\mathbb{N}$, $\mathcal{F}_m := \mathcal{B}_{-m}$. This may also be proved by Hilbert space projection methods when $Z \in L^2(P)$ and then by a limiting argument for all $Z \in L^1(P)$. ∎

**Proposition 30.7.** *Suppose that $Z \in L^1(\Omega, \mathcal{B}, P)$ and $\sigma$ and $\tau$ are two stopping times. Then*

1. *$\mathbb{E}[Z|\mathcal{B}_\sigma] = \mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$ on $\{\sigma \leq \tau\}$ and hence on $\{\sigma < \tau\}$.*
2. *$\mathbb{E}[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau] = \mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$.*

*Moreover, both results hold if $\sigma$ and $\tau$ are optional times provided every occurrence of the letter $\mathcal{B}$ is replaced by $\mathcal{B}^+$.*

**Proof.** 1. From Corollary 29.17, $1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma]$ is $\mathcal{B}_{\sigma\wedge\tau}$ – measurable and therefore,

$$\begin{aligned}1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma] &= \mathbb{E}\left[1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\sigma\wedge\tau}\right] \\ &= 1_{\sigma\le\tau}\mathbb{E}\left[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\sigma\wedge\tau}\right] = 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]\end{aligned}$$

as desired.

2. Writing

$$Z = 1_{\sigma\le\tau}Z + 1_{\tau<\sigma}Z$$

we find, using item 1. that,

$$\begin{aligned}\mathbb{E}[Z|\mathcal{B}_\sigma] &= 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_\sigma] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma] \\ &= 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma].\end{aligned} \tag{30.2}$$

Another application of item 1. shows,

$$\begin{aligned}\mathbb{E}\left[1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau\right] &= 1_{\tau<\sigma}\mathbb{E}\left[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau\right] \\ &= 1_{\tau<\sigma}\mathbb{E}\left[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_{\tau\wedge\sigma}\right] = 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}].\end{aligned}$$

Using this equation and the fact that $1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}]$ is $\mathcal{B}_{\sigma\wedge\tau}$ – measurable, we may condition Eq. (30.2) on $\mathcal{B}_{\tau\wedge\sigma}$ to find

$$\mathbb{E}\left[\mathbb{E}[Z|\mathcal{B}_\sigma]|\mathcal{B}_\tau\right] = 1_{\sigma\le\tau}\mathbb{E}[Z|\mathcal{B}_{\sigma\wedge\tau}] + 1_{\tau<\sigma}\mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}] = \mathbb{E}[Z|\mathcal{B}_{\tau\wedge\sigma}].$$

■

**Lemma 30.8.** *Suppose $\sigma$ is an optional time and $\{\sigma_m\}_{m=1}^\infty$ are stopping times such that $\sigma_m\downarrow\sigma$ as $m\uparrow\infty$ and $\sigma<\sigma_m$ on $\{\sigma<\infty\}$ for all $m\in\mathbb{N}$. Then $\mathcal{B}_{\sigma_m}\downarrow\mathcal{B}_\sigma^+$ as $m\to\infty$, i.e. $\mathcal{B}_{\sigma_m}$ is decreasing in $m$ and*

$$\mathcal{B}_\sigma^+ = \cap_{m=1}^\infty\mathcal{B}_{\sigma_m}. \tag{30.3}$$

**Proof.** If $A\in\mathcal{B}_\sigma^+$, then $A\in\mathcal{B}_\infty$ and for all $t\in\mathbb{R}_+$ and $m\in\mathbb{N}$ we have

$$A\cap\{\sigma_m\le t\} = A\cap\{\sigma<t\}\cap\{\sigma_m\le t\}\in\mathcal{B}_t.$$

This shows $A\in\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}$. For the converse, observe that

$$\{\sigma<t\} = \cup_{m=1}^\infty\{\sigma_m\le t\}\ \forall\, t\in\mathbb{R}_+.$$

Therefore if $A\in\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}$ then $A\in\mathcal{B}_\infty$ and

$$A\cap\{\sigma<t\} = \cup_{m=1}^\infty\left[A\cap\{\sigma_m\le t\}\right]\in\mathcal{B}_t\ \forall\, t\in\mathbb{R}_+.$$

■

**Theorem 30.9 (Continuous time optional sampling theorem).** *Let $\{X_t\}_{t\geqslant0}$ be a right continuous $\{\mathcal{B}_t\}$ (or $\{\mathcal{B}_t^+\}$) – submartingale and $\sigma$ and $\tau$ be two $\{\mathcal{B}_t\}$ – optional (or stopping) times such that $M:=\sup_{\omega\in\Omega}\tau(\omega)<\infty$.*[1] *Then $X_\tau\in L^1(\Omega,\mathcal{B}_\tau^+,P)$, $X_{\sigma\wedge\tau}\in L^1\left(\Omega,\mathcal{B}_{\sigma\wedge\tau}^+,P\right)$ and*

$$X_{\sigma\wedge\tau}\le\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right]. \tag{30.4}$$

**Proof.** Let $\{\sigma_m\}_{m=1}^\infty$ and $\{\tau_n\}_{n=1}^\infty$ be the sequences of approximate times for $\sigma$ and $\tau$ respectively defined Lemma 29.18, i.e.

$$\tau_n := \infty 1_{\tau=\infty} + \sum_{k=1}^\infty\frac{k}{2^n}1_{\frac{k-1}{2^n}\le\tau<\frac{k}{2^n}}.$$

By the discrete optional sampling Proposition 30.3, we know that

$$X_{\sigma_m\wedge\tau_n}\le\mathbb{E}\left[X_{\tau_n}|\mathcal{B}_{\sigma_m}\right]\text{ a.s.} \tag{30.5}$$

Since $X_t$ is right continuous, $X_{\tau_n}(\omega)\to X_\tau(\omega)$ for all $\omega\in\Omega$ which combined with Lemma 30.5 implies $X_{\tau_n}\to X_\tau$ in $L^1(P)$ and in particular $X_\tau\in L^1(P)$. Similarly, $X_{\sigma_n\wedge\tau_n}\to X_{\sigma\wedge\tau}$ in $L^1(P)$ and therefore $X_{\sigma\wedge\tau}\in L^1(P)$. Using the $L^1(P)$ – contractivity of conditional expectation along with the fact that $X_{\sigma_m\wedge\tau_n}\to X_{\sigma_m\wedge\tau}$ on $\Omega$, we may pass to the limit $(n\to\infty)$ in Eq. (30.5) to find

$$X_{\sigma_m\wedge\tau}\le\mathbb{E}\left[X_\tau|\mathcal{B}_{\sigma_m}\right]\text{ a.s.} \tag{30.6}$$

From the right continuity of $\{X_t\}$ and making use of Lemma 30.6 (or Corollary 20.81) and Lemma 30.8, we may let $m\to\infty$ in Eq. (30.6) to find

$$X_{\sigma\wedge\tau}\le\lim_{m\to\infty}\mathbb{E}\left[X_\tau|\mathcal{B}_{\sigma_m}\right] = \mathbb{E}\left[X_\tau|\cap_{m=1}^\infty\mathcal{B}_{\sigma_m}\right] = \mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right]$$

which is Eq. (30.4). ■

**Corollary 30.10 (Optional stopping).** *Let $\{X_t\}_{t\geqslant0}$ be a right continuous $\{\mathcal{B}_t\}$ (or $\{\mathcal{B}_t^+\}$) – submartingale and $\tau$ be any $\{\mathcal{B}_t\}$ – optional (or stopping) time. Then the stopped process, $X_t^\sigma:=X_{\sigma\wedge t}$ is a right continuous $\{\mathcal{B}_t^+\}$ – submartingale.*

**Proof.** Let $0\le s\le t<\infty$ and apply Theorem 30.9 with to the two stopping times, $\sigma\wedge s$ and $\sigma\wedge t$ to find

$$X_s^\sigma = X_{\sigma\wedge s}\le\mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge s}^+\right] = \mathbb{E}\left[X_t^\sigma|\mathcal{B}_{\sigma\wedge s}^+\right].$$

From Proposition 30.7,

$$\mathbb{E}\left[X_t^\sigma|\mathcal{B}_{\sigma\wedge s}^+\right] = \mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge t\wedge s}^+\right] = \mathbb{E}\left[\mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_{\sigma\wedge t}^+\right]|\mathcal{B}_s^+\right] = \mathbb{E}\left[X_{\sigma\wedge t}|\mathcal{B}_s^+\right]$$

and therefore, we have shown $X_s^\sigma\le\mathbb{E}\left[X_t^\sigma|\mathcal{B}_s^+\right]$. Since $X_s^\sigma=X_{\sigma\wedge s}$ is $\mathcal{B}_{\sigma\wedge s}^+$ measurable and $\mathcal{B}_{\sigma\wedge s}^+\subset\mathcal{B}_s^+$, it follows that $X_s^\sigma$ is $\mathcal{B}_s^+$ – measurable. ■

[1] We will see below in Theorem 30.28, that the boundedness restriction on $\tau$ may be replaced by the assumption that $\{X_t^+\}_{t\ge0}$ is uniformly integrable.

## 30.1 Submartingale Inequalities

Let $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\in\mathbb{R}_+}, P\right)$ be a filtered probability space, $\mathbb{D}$ be any dense subset of $\mathbb{R}_+$ containing 0, and let $\mathbb{T}$ denote either $\mathbb{D}$ or $\mathbb{R}_+$. Throughout this section, $\{X_t\}_{t\in\mathbb{T}}$ will be a submartingale which is *assumed* to be *right continuous* if $\mathbb{T} = \mathbb{R}_+$. To keep the notation unified, for $T \in \mathbb{R}_+$, we will simply denote $\sup_{\mathbb{D}\ni t\le T} X_t = \inf_{\mathbb{D}\ni t\le T} X_t$, and $\sup_{s\in\mathbb{D}\cap[0,T]} |X_s|$ by $\sup_{t\le T} X_t$, $\inf_{t\le T} X_t$, and $X_T^*$ respectively. It is worth observing that if $\mathbb{T} = \mathbb{R}_+$ and $T \in \mathbb{D}$, we have (by the assumed right continuity of $X_t$) that

$$\sup_{t\le T} X_t = \sup_{\mathbb{D}\ni t\le T} X_t, \ \ \inf_{t\le T} X_t = \inf_{\mathbb{D}\ni t\le T} X_t \text{ and , } \sup_{s\in[0,T]} |X_s| = \sup_{s\in\mathbb{D}\cap[0,T]} |X_s| . \tag{30.7}$$

Our immediate goal is to generalize the submartingale inequalities of Section 20.5 to this context.

**Proposition 30.11 (Maximal Inequalities of Bernstein and Lévy).** *With $\mathbb{T} = \mathbb{D}$ for $\mathbb{T} = \mathbb{R}_+$, for any $a \ge 0$ and $T \in \mathbb{T}$, we have,*

$$aP\left(\sup_{t\le T} X_t \ge a\right) \le \mathbb{E}\left[X_T : \sup_{t\le T} X_t \ge a\right] \le \mathbb{E}\left[X_T^+\right], \tag{30.8}$$

$$aP\left(\inf_{t\le T} X_t \le -a\right) \le \mathbb{E}\left[X_T : \inf_{t\le T} X_t > -a\right] - \mathbb{E}\left[X_0\right] \tag{30.9}$$

$$\le \mathbb{E}\left[X_T^+\right] - \mathbb{E}\left[X_0\right], \tag{30.10}$$

*and*

$$aP\left(X_T^* \ge a\right) \le 2\mathbb{E}\left[X_T^+\right] - \mathbb{E}\left[X_0\right]. \tag{30.11}$$

*In particular if $\{M_t\}_{t\in\mathbb{T}}$ is a martingale and $a > 0$, then*

$$P\left(M_T^* \ge a\right) \le \frac{1}{a}\mathbb{E}\left[|M|_T : M_T^* \ge a\right] \le \frac{1}{a}\mathbb{E}\left[|M_T|\right] \tag{30.12}$$

**Proof.** First assume $\mathbb{T} = \mathbb{D}$. For each $k \in \mathbb{N}$ let

$$\Lambda_k = \{0 = t_0 < t_1 < \cdots < t_m = T\} \subset \mathbb{D}\cap[0,T]$$

be a finite subset of $\mathbb{D}\cap[0,T]$ containing $\{0,T\}$ such that $\Lambda_k \uparrow \mathbb{D}\cap[0,T]$. Noting that $\{X_{t_n}\}_{n=0}^m$ is a discrete $\left(\Omega, \mathcal{B}, \{\mathcal{B}_{t_n}\}_{n=0}^m, P\right)$ submartingale, Proposition 20.43 implies all of the inequalities in Eqs. (30.8) – (30.11) hold provided we replace $\sup_{t\le T} X_t$ by $\max_{t\in\Lambda_k} X_t$, $\inf_{t\le T} X_t$ by $\min_{t\in\Lambda_k} X_t$, and $X_T^*$ by $\max_{t\in\Lambda_k} |X_t|$. Since $\max_{t\in\Lambda_k} X_t \uparrow \sup_{t\le T} X_t$, $\max_{t\in\Lambda_k} |X_t| \uparrow X_T^*$, and $\min_{t\in\Lambda_k} X_t \downarrow \inf_{t\le T} X_t$, we may use the MCT and the DCT to pass to the limit $(k\to\infty)$ in order to conclude Eqs. (30.8) – (30.11) are valid as stated. Equation (30.12) follows from Eq. (30.8) applied to $X_t := |M_t|$.

Now suppose that $\{X_t\}_{t\in\mathbb{R}_+}$ and $\{M_t\}_{t\in\mathbb{R}_+}$ are right continuous. Making use of the observations in Eq. (30.7), we see that Eqs. (30.8) – (30.12) remain valid for $\mathbb{T} = \mathbb{R}_+$ by what we have just proved in the case $\mathbb{T} = \mathbb{D}\cup\{T\}$. ■

**Proposition 30.12 (Doob's Inequality).** *Suppose that $X_t$ is a non-negative submartingale (for example $X_t = |M_t|$ where $M_t$ is a martingale) and $1 < p < \infty$, then for any $T \in \mathbb{T}$,*

$$\mathbb{E}X_T^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_T^p. \tag{30.13}$$

**Proof.** Using the notation in the proof of Proposition 30.11, it follows from Corollary 20.47 that

$$\mathbb{E}\left[\max_{t\in\Lambda_k} |X_t|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_T^p.$$

Using the MCT, we may let $k \uparrow \infty$ in this equation to arrive at Eq. (30.13) when $\mathbb{T} = \mathbb{D}$. The case when $\mathbb{T} = \mathbb{R}_+$ follows immediately using the comments at the end of the proof of Proposition 30.11. ■

**Lemma 30.13.** *Suppose that $F_n$ is a sequence of bounded functions on $[a,b)$ which are uniformly convergent to a function $F$. If $\xi_n := \lim_{t\downarrow a} F_n(t)$ exists for all $n$, then $\xi := \lim_{t\downarrow a} F(t)$ exists and $\xi_n \to \xi$ as $n \to \infty$. An analogous statement holds for left limits. In particular right (left) continuous functions are preserved under uniform limits.*

**Proof.** Let $\varepsilon_n := \sup_{t\in[a,b)} |F(t) - F_n(t)|$ which by assumption tends to zero as $n\to\infty$. Thus for $s, t > a$, we have

$$\begin{aligned} |F(t) - F(s)| &\le |F(t) - F_n(t)| + |F_n(t) - F_n(s)| + |F_n(s) - F(s)| \\ &\le 2\varepsilon_n + |F_n(t) - F_n(s)| . \end{aligned}$$

Therefore we have

$$\limsup_{s,t\downarrow a} |F(t) - F(s)| \le 2\varepsilon_n \to 0 \text{ as } n\to\infty$$

which shows that $\xi := \lim_{t\downarrow a} F(t)$ exists. Similarly, for any $t > a$,

$$\begin{aligned} |\xi - \xi_n| &\le |\xi - F(t)| + |F(t) - F_n(t)| + |F_n(t) - \xi_n| \\ &\le |\xi - F(t)| + |F_n(t) - \xi_n| + \varepsilon_n \end{aligned}$$

and hence by passing to the limit as $t \downarrow a$ in the previous inequality we have $|\xi - \xi_n| \le \varepsilon_n \to 0$ as $n\to\infty$. ■

**Corollary 30.14.** *Suppose that* $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ *is a filtered probability space such that* $\mathcal{B}_t$ *contains all* $P$ *– null subsets*[2] *of* $\mathcal{B}$ *for all* $t\in\mathbb{R}_+$. *For any* $T\in\mathbb{R}_+$, *let* $\mathbb{M}_T$ *denote the collection of (right) continuous* $L^2$ *– martingales,* $M:=\{M_t\}_{t\leq T}$ *equipped with the inner product,*

$$(M,N)_T:=\mathbb{E}\left[M_TN_T\right].$$

*(More precisely, two (right) continuous* $L^2$ *– martingales,* $M$ *and* $N$, *are taken to be equal if* $P(M_t=N_t\ \forall\ t\leq T)=1$.*) Then the space,* $(\mathbb{M}_T,(\cdot,\cdot)_T)$, *is a Hilbert space and the map,* $U:\mathbb{M}_T\to L^2(\Omega,\mathcal{B}_T,P)$ *defined by* $UM:=M_T$, *is an isometry.*

**Proof.** Since $M_t=\mathbb{E}[M_T|\mathcal{B}_t]$ a.s., if $(M_T,M_T)=\mathbb{E}|M_T|^2=0$ then $M=0$ in $\mathbb{M}_T$. This shows that $U$ is injective and by definition $U$ is an isometry and $(\cdot,\cdot)_T$ is an inner product on $\mathbb{M}_T$. To finish the proof, we need only show $H:=\mathrm{Ran}(U)$ is a closed subspace of $L^2(\Omega,\mathcal{B}_T,P)$ or equivalently that $\mathbb{M}_T$ is complete.

Suppose that $\{M^n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{M}_T$, then by Doob's inequality (Proposition 30.12) and Hölder's inequality, we have

$$\begin{aligned}\mathbb{E}\left[(M^n-M^m)_T^*\right]&\leq\sqrt{\mathbb{E}\left[(M^n-M^m)_T^{*2}\right]}\\&\leq\sqrt{4\mathbb{E}|M_T^n-M_T^m|^2}=2\|M^n-M^m\|_T\to 0\text{ as }m,n\to\infty.\end{aligned}$$

By passing to a subsequence if necessary, we may assume

$$\sum_{n=1}^\infty\mathbb{E}\left[\left(M^{n+1}-M^n\right)_T^*\right]\leq 2\sum_{n=1}^\infty\left\|M^{n+1}-M^n\right\|_T<\infty$$

from which it follows that

$$\mathbb{E}\left[\sum_{n=1}^\infty\left(M^{n+1}-M^n\right)_T^*\right]=\sum_{n=1}^\infty\mathbb{E}\left[\left(M^{n+1}_\cdot-M^n_\cdot\right)_T^*\right]<\infty.$$

So if we let

$$\Omega_0:=\left\{\sum_{n=1}^\infty\left(M^{n+1}_\cdot-M^n_\cdot\right)_T^*<\infty\right\},$$

then $P(\Omega_0)=1$. Hence if $m<l$, the triangle inequality implies

$$\left(M^l_\cdot-M^m_\cdot\right)_T^*\leq\sum_{n=m}^{l-1}\left(M^{n+1}_\cdot-M^n_\cdot\right)_T^*\to 0\text{ on }\Omega_0\text{ as }m,l\to\infty,$$

[2] Lemma 30.15 below shows that this hypothesis can always be fulfilled if one is willing to "augment" the filtration by the $P$ – null sets.

which shows that $\{M^n_\cdot(\omega)\}_{n=1}^\infty$ is a uniformly Cauchy sequence and hence uniformly convergent for all $\omega\in\Omega_0$. Therefore by Lemma 30.13, $t\to M_t(\omega)$ is (right) continuous for all $\omega\in\Omega_0$. We complete the definition of $M$ by setting $M_\cdot(\omega)\equiv 0$ for $\omega\notin\Omega_0$. Since $\mathcal{B}_t$ contains all of the null subset in $\mathcal{B}$, it is easy to see that $M_\cdot$ is a $\mathcal{B}_t$ – adapted process. Moreover, by Fatou's lemma, we have

$$\begin{aligned}\mathbb{E}\left[(M_\cdot-M^m_\cdot)_T^{*2}\right]&=\mathbb{E}\left[\liminf_{n\to\infty}(M^n_\cdot-M^m_\cdot)_T^{*2}\right]\\&\leq\liminf_{n\to\infty}\mathbb{E}\left[(M^n_\cdot-M^m_\cdot)_T^{*2}\right]\to 0\text{ as }m\to\infty.\end{aligned}$$

In particular $M_t^m\to M_t$ in $L^2(P)$ for all $t\leq T$ from which follows that $M$ is still an $L^2$ – martingale. As $M$ is (right) continuous, $M\in\mathbb{M}_T$ and

$$\|M-M^n\|_T=\|M_T-M_T^n\|_{L^2(P)}\to 0\text{ as }n\to\infty.$$

■

## 30.2 Regularizing a submartingale

**Lemma 30.15.** *Suppose that* $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ *is a filtered probability space and* $\{X_t\}_{t\geq 0}$ *be a* $\{\mathcal{B}_t\}_{t\geq 0}$ *– submartingale. Then* $\{X_t\}_{t\geq 0}$ *is also a* $\left\{\bar{\mathcal{B}}_t\right\}_{t\geq 0}$ *– submartingale. Moreover, we may first replace* $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\geq 0},P\right)$ *by its completion,* $\left(\Omega,\bar{\mathcal{B}},\bar{P}\right)$ *(see Proposition 6.77), then* $\{X_t\}_{t\geq 0}$ *is still a submartingale relative to the filtration* $\bar{\mathcal{B}}_t:=\mathcal{B}_t\vee\bar{\mathcal{N}}$ *where*

$$\bar{\mathcal{N}}:=\left\{B\in\bar{\mathcal{B}}:\bar{P}(B)=0\right\}.$$

**Proof.** It suffices to prove the second assertion. By the augmentation Lemma 29.22 we know that $B\in\bar{\mathcal{B}}_s:=\mathcal{B}_s\vee\bar{\mathcal{N}}$ iff there exists $A\in\mathcal{B}_s$ such that $B\bigtriangleup A\in\bar{\mathcal{N}}$. Then for any $t>s$ we have

$$\mathbb{E}_{\bar{P}}[X_t-X_s:B]=\mathbb{E}_{\bar{P}}[X_t-X_s:A]=\mathbb{E}_P[X_t-X_s:A]\geq 0.$$

■

**Proposition 30.16.** *Suppose that* $\{X_t\}_{t\in\mathbb{R}_+}$ *is an* $\{\mathcal{B}_t\}$ *– submartingale such that* $t\to X_t$ *is right continuous in probability, i.e.* $X_t\overset{P}{\to}X_s$ *as* $t\downarrow s$ *for all* $s\in\mathbb{R}_+$. *(For example, this hypothesis will hold if there exists* $\varepsilon>0$ *such that* $\lim_{t\downarrow s}\mathbb{E}|X_t-X_s|^\varepsilon=0$ *for all* $s\in\mathbb{R}_+$.*) Then* $\{X_t\}_{t\in\mathbb{R}_+}$ *is also an* $\left\{\mathcal{B}_t^+\right\}$ *– submartingale.*

**Proof.** Let $0 \le s < t < \infty$, $A \in \mathcal{B}_s^+$, and $s_n \in (s,t)$ such that $s_n \downarrow s$. By Lemma 30.4 we know that $\hat{X}_s := \lim_{n\to\infty} X_{s_n}$ exists a.s. and in $L^1(P)$ and using the assumption that $X_{s_n} \overset{P}{\to} X_s$ we may conclude that $X_{s_n} \to X_s$ in $L^1(P)$. Since $A \in \mathcal{B}_s^+ \subset \mathcal{B}_{s_n}$ for all $n$, we have

$$\mathbb{E}[X_t - X_s : A] = \lim_{n\to\infty} \mathbb{E}[X_t - X_{s_n} : A] \ge 0.$$

■

**Corollary 30.17.** *Suppose $\left(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t\ge 0}, P\right)$ is a filtered probability space and $\{X_t\}_{t\in\mathbb{R}_+}$ is an $\{\mathcal{B}_t\}$ – submartingale such that $X_t \overset{P}{\to} X_s$ as $t \downarrow s$ for all $s \in \mathbb{R}_+$. Let $\left(\Omega, \bar{\mathcal{B}}, \bar{P}\right)$ denote the completion of $(\Omega, \mathcal{B}, P)$, $\mathcal{N}$ and $\bar{\mathcal{N}}$ be the $P$ and $\bar{P}$ null sets respectively, then $\left(\Omega, \mathcal{B}, \{\mathcal{B}_{t+} \vee \mathcal{N}\}_{t\ge 0}, P\right)$ satisfies the weak usual hypothesis (see Definition 29.21), $\left(\Omega, \bar{\mathcal{B}}, \left\{\mathcal{B}_{t+} \vee \bar{\mathcal{N}}\right\}_{t\ge 0}, \bar{P}\right)$ satisfies the usual hypothesis, and $\{X_t\}_{t\ge 0}$ is a submartingale relative to each of these filtrations.*

**Proof.** This follows directly from Proposition 30.16, Lemma 30.15, and Lemma 29.23. We use Lemma 29.23 to guarantee that $\{\mathcal{B}_{t+} \vee \mathcal{N}\}_{t\ge 0}$ and $\left\{\mathcal{B}_{t+} \vee \bar{\mathcal{N}}\right\}_{t\ge 0}$ are right continuous. ■

In all of the examples of submartingales appearing in this book, the hypothesis and hence the conclusions of Proposition 30.16 will apply. For this reason there is typically no harm in assuming that our filtration is right continuous. By Corollary 30.17 we may also assume that $\mathcal{B}_t$ contains all $P$ – null sets. The results in the following exercise are useful to keep in mind as you are reading the rest of this section.

**Exercise 30.1 (Continuous version of Example 20.7).** Suppose that $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$, and $P = m$ – Lebesgue measure. Further suppose that $\varepsilon : [0,\infty) \to \{0,1\}$ is **any** function of your choosing. Then define, for $t \ge 0$ and $x \in \Omega$,

$$M_t^\varepsilon(x) := e^t \left(\varepsilon(t) 1_{0<x\le e^{-t}} + (1-\varepsilon(t)) 1_{0<x<e^{-t}}\right) = e^t \left(1_{0<x<e^{-t}} + \varepsilon(t) 1_{x=e^{-t}}\right).$$

Further let $\mathcal{B}_t^\varepsilon := \sigma(M_s^\varepsilon : s \le t)$ for all $t \ge 0$ and for $a \in (0,1]$ let

$$\mathcal{F}_{(0,a]} := \left\{[0,a] \cup A : A \in \mathcal{B}_{(a,1]}\right\} \cup \mathcal{B}_{(a,1]}$$

and

$$\mathcal{F}_{(0,a)} := \left\{(0,a) \cup A : A \in \mathcal{B}_{[a,1]}\right\} \cup \mathcal{B}_{[a,1]}.$$

Show:

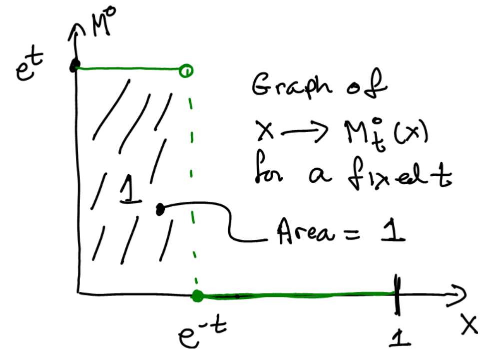



**Fig. 30.1.** The graph of $x \to M_t^0(x)$ for some fixed $t$.

1. $\mathcal{F}_{(0,a]}$ and $\mathcal{F}_{(0,a)}$ are sub – sigma – algebras of $\mathcal{B}$ such that $\mathcal{F}_{(0,a]} \subsetneq \mathcal{F}_{(0,a)}$ and
$$\mathcal{B} = \vee_{a\in(0,1]} \mathcal{F}_{(0,a]} = \vee_{a\in(0,1]} \mathcal{F}_{(0,a)}.$$
2. For all $b \in (0,1]$,
$$\mathcal{F}_{(0,b)} = \cap_{a<b} \mathcal{F}_{(0,a]} = \cap_{a<b} \mathcal{F}_{(0,a)}. \tag{30.14}$$
3. $M_{t+}^\varepsilon = M_t^0$ for all $t \ge 0$ and $M_{t-}^\varepsilon = M_t^1$ for all $t > 0$. In particular, the sample paths, $t \to M_{t+}^\varepsilon(x)$, are right continuous and possess left limits for all $x \in \Omega$.

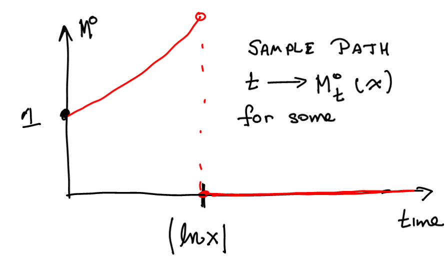



**Fig. 30.2.** A typical sample path of $M_\cdot^0(x)$.

4. $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t}]}$ if $\varepsilon(t) = 1$ and $\mathcal{B}_t^\varepsilon = \mathcal{F}_{(0,e^{-t})}$ if $\varepsilon(t) = 0$.
5. No matter how $\varepsilon$ is chosen, $\mathcal{B}_{t+}^\varepsilon = \mathcal{B}_t^0 := \mathcal{F}_{(0,e^{-t})}$ for all $t \ge 0$.
6. $M_t^\varepsilon$ is a $\{\mathcal{B}_t^\varepsilon\}_{t\ge 0}$ – martingale and in fact it is a $\left\{\mathcal{B}_{t+}^\varepsilon = \mathcal{B}_t^0\right\}_{t\ge 0}$ – martingale.
7. The map, $[0,\infty) \times (0,1] \ni (t,x) \to M_t^\varepsilon(x) \in \mathbb{R}_+$ is measurable iff $\{t \in [0,\infty) : \varepsilon(t) = 1\} \in \mathcal{B}_{\mathbb{R}_+}$.

8. Let
$$\mathcal{N} := \left\{x : M_t^\varepsilon(x) \neq M_{t+}^\varepsilon(x) \text{ for some } t \geq 0\right\}.$$
Show $\mathcal{N} = \{x : \varepsilon(|\ln x|) = 1\}$ and observe that $\mathcal{N}$ is measurable iff $\varepsilon$ is measurable. Also observe that if $\varepsilon \equiv 1$, then $P(\mathcal{N}) = 1$ and hence $M_{t+}^\varepsilon$ and $M_t^\varepsilon$ are certainly not indistinguishable, see Definition 27.2.
9. Show $\{M_t^\varepsilon\}_{t\geq 0}$ is **not** uniformly integrable.
10. Let $Z \in L^1(\Omega, \mathcal{B}, P)$ find a version, $N_t^\varepsilon$, of $\mathbb{E}[Z|\mathcal{B}_t^\varepsilon]$. Verify that for any sequence, $\{t_n\}_{n=1}^\infty \subset [1,\infty)$, that $N_{t_n}^\varepsilon \to Z$ almost surely and in $L^1(P)$ as $n \to \infty$.

**Definition 30.18 (Upcrossings).** *Let $\{x_t\}_{t\in\mathbb{T}}$ be a real valued function which is right continuous if $\mathbb{T} = \mathbb{R}_+$. Given $-\infty < a < b < \infty$, $T \in \mathbb{T}$, and a finite subset, $F$, of $[0,T] \cap \mathbb{T}$, let $U_F^x(a,b)$ denote the **number of upcrossings of** $\{x_t\}_{t\in F}$ **across** $[a,b]$, see Section 20.6. Also let*
$$U_T^x(a,b) := \sup\{U_F^x(a,b) : F \subset_f \mathbb{T} \cap [0,T]\} \tag{30.15}$$
*be the **number of upcrossings of** $\{x_t\}_{t\in\mathbb{T}\cap[0,T]}$ **across** $[a,b]$.*

**Lemma 30.19.** *If $\mathbb{T} = \mathbb{D}$ and $\{F_n\}_{n=1}^\infty$ is a sequence of finite subsets of $\mathbb{D} \cap [0,T]$ such that $F_n \uparrow \mathbb{D} \cap [0,T]$, then*
$$U_T^x(a,b) := \lim_{n\to\infty} U_{F_n}^x(a,b). \tag{30.16}$$
*In particular, $U_T^X(a,b)$ is a $\mathcal{B}_T$ – measurable random variable when $\mathbb{T} = \mathbb{D}$.*

**Proof.** It is clear that $U_{F_n}^x(a,b) \leq U_T^x(a,b)$ for all $n$ and $U_{F_n}^x(a,b)$ is increasing with $n$ and therefore the limit in Eq. (30.16) exists and satisfies, $\lim_{n\to\infty} U_{F_n}^x(a,b) \leq U_T^x(a,b)$. Moreover, for any $F \subset_f \mathbb{D} \cap [0,T]$ we may find an $n \in \mathbb{N}$ sufficiently large so that $F \subset F_n$. For this $n$ we will have
$$U_F^x(a,b) \leq U_{F_n}^x(a,b) \leq \lim_{n\to\infty} U_{F_n}^x(a,b).$$
Taking supremum over all $F \subset_f \mathbb{D} \cap [0,T]$ in this estimate then shows $U_T^x(a,b) \leq \lim_{n\to\infty} U_{F_n}^x(a,b)$. ■

*Remark 30.20.* It is easy to see that if $\mathbb{T} = \mathbb{R}_+$, $x_t$ is right continuous, and $a < \alpha < \beta < b$, then
$$U_T^x(a,b) \leq \sup\{U_F^x(\alpha,\beta) : F \subset_f \mathbb{D} \cap [0,T]\}.$$

**Lemma 30.21.** *Let $T \in \mathbb{R}_+$ and $\{x_t\}_{t\in\mathbb{D}}$ be a real valued function such that $U_T^x(a,b) < \infty$ for all $-\infty < a < b < \infty$ with $a,b \in \mathbb{Q}$. Then*
$$x_{t-} := \lim_{\mathbb{D}\ni s\uparrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in (0,T] \text{ and} \tag{30.17}$$
$$x_{t+} := \lim_{\mathbb{D}\ni s\downarrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in [0,T). \tag{30.18}$$
*Moreover, if we let $U_\infty^x(a,b) = \lim_{T\uparrow\infty} U_T^x(a,b)$ and further assume that $U_\infty^x(a,b) < \infty$ for all $-\infty < a < b < \infty$ with $a,b \in \mathbb{Q}$, then $x_\infty := \lim_{t\uparrow\infty} x_t$ exists in $\bar{\mathbb{R}}$ as well.*

**Proof.** I will only prove the statement in Eq. (30.17) since all of the others are similar. If $x_{t-}$ does not exists in $\bar{\mathbb{R}}$ then we can find $a,b \in \mathbb{Q}$ such that
$$\liminf_{\mathbb{D}\ni s\uparrow t} x_s < a < b < \limsup_{\mathbb{D}\ni s\uparrow t} x_s.$$
From the definition of the lim inf and the lim sup, it follows that for every $\varepsilon \in (0,t)$ there are infinitely many $s \in (t-\varepsilon, t)$ such that $x_s < a$ and infinitely may $s \in (t-\varepsilon,t)$ such that $x_s > b$. From this observation it is easy to see that $\infty = U_t^x(a,b) \leq U_T^x(a,b)$. ■

**Lemma 30.22.** *Suppose that $\mathbb{T} = \mathbb{D}$, $S$ is a metric space, and $\{x_t \in S\}_{t\in\mathbb{D}}$.*

1. *If for all $t \in \mathbb{R}_+$,*
$$x_t^+ := x_{t+} = \lim_{\mathbb{D}\ni s\downarrow t} x_s \text{ exists in } S,$$
*then $\mathbb{R}_+ \ni t \to x_t^+ \in S$ is right continuous.*
2. *If we further assume that*
$$x_{t-} := \lim_{\mathbb{D}\ni s\uparrow t} x_s \text{ exists in } S$$
*for all $t > 0$, then $\lim_{\tau\uparrow t} x_{\tau+} = x_{t-}$ for all $t > 0$.*
3. *Moreover, if $\lim_{\mathbb{D}\ni t\uparrow\infty} x_t$ exists in $S$ then again $\lim_{t\uparrow\infty} x_{t+} = \lim_{\mathbb{D}\ni t\uparrow\infty} x_t$.*

**Proof.** 1. Suppose $t \in \mathbb{R}_+$ and $\varepsilon > 0$ is given. By assumption, there exists $\delta > 0$ such that for $s \in (t, t+\delta) \cap \mathbb{D}$, we have $\rho(x_{t+}, x_s) \leq \varepsilon$. Therefore if $\tau \in (t, t+\delta)$, then
$$\rho(x_{t+}, x_{\tau+}) = \lim_{\mathbb{D}\ni s\downarrow\tau} \rho(x_{t+}, x_s) \leq \varepsilon$$
from which it follows that $x_{\tau+} \to x_{t+}$ as $\tau \downarrow t$.

2. Now suppose $t > 0$ such that $x_{t-}$ exists in $S$. Then for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $\rho(x_{t-}, x_s) \leq \varepsilon$ if $s \in (t-\delta, t) \cap \mathbb{D}$. Hence, if $\tau \in (t-\delta, t)$ we may conclude,
$$\rho(x_{t-}, x_{\tau+}) = \lim_{s\downarrow\tau} \rho(x_{t+}, x_s) \leq \varepsilon$$

from which it follows that $x_{\tau+} \uparrow x_{t-}$ as $\tau \uparrow t$.

3. Now suppose $x_\infty := \lim_{\mathbb{D}\ni s\uparrow\infty} x_s$ exists in $S$. Then for every $\varepsilon > 0$, there exists $M = M(\varepsilon) < \infty$ such that $\rho(x_\infty, x_s) \le \varepsilon$ if $s \in \mathbb{D}\cap(M,\infty)$. Hence if $t \in (M,\infty)$ we have

$$\rho(x_\infty, x_{t+}) = \lim_{\mathbb{D}\ni s\downarrow t} \rho(x_\infty, x_s) \le \varepsilon$$

from which we conclude that, $\lim_{t\uparrow\infty} x_{t+}$ exists in $S$ and is equal to $x_\infty$. ■

**Theorem 30.23 (Doob's upcrossing inequality).** *Let $\{X_t\}_{t\in\mathbb{D}}$ be a submartingale and $-\infty < a < b < \infty$. Then for all $T \in \mathbb{D}$,*

$$\mathbb{E}\left[U_T^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_T - a)_+ - \mathbb{E}(X_0 - a)_+\right]. \tag{30.19}$$

**Proof.** Let $\{F_n\}_{n=1}^\infty$ be a sequence as in Lemma 30.19 and assume without loss of generality that $0, T \in F_n$ for all $n$. It then follows From Theorem 20.53 that

$$\mathbb{E}\left[U_{F_n}^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_T - a)_+ - \mathbb{E}(X_0 - a)_+\right] \ \forall\, n \in \mathbb{N}.$$

By letting $n \uparrow \infty$, Eq. (30.19) follows from this inequality, Lemma 30.19, and the MCT. ■

**Theorem 30.24.** *Let $\{X_t\}_{t\in\mathbb{D}}$ be a submartingale,*

$$\Omega_0 := \bigcap_{T\in\mathbb{N}} \left(\left\{\sup_{\mathbb{D}\ni t\le T} |X_t| < \infty\right\} \cap \left[\bigcap \{U_T^X(a,b) < \infty : \ a < b \text{ with } a,b \in \mathbb{Q}\}\right]\right), \tag{30.20}$$

*for all $t \in \mathbb{R}_+$,*

$$Y_t := \limsup_{\mathbb{D}\ni s\downarrow t} X_s \text{ and } \bar{X}_t := Y_t \cdot 1_{|Y_t|<\infty}. \tag{30.21}$$

*Then;*

1. *$P(\Omega_0) = 1$.*
2. *on $\Omega_0$, $\sup_{t\le T} |X_t| < \infty$ and $X_{t+}$ and $X_{t-}$ exist for all $t \in \mathbb{R}_+$ where by convention $X_{0-} := X_0$.*
3. *$\{X_{t+}(\omega)\}_{t\in\mathbb{R}_+}$ is right continuous with left hand limits for all $\omega \in \Omega_0$.*
4. *For any $t \in \mathbb{R}_+$ and any sequence $\{s_n\}_{n=1}^\infty \subset \mathbb{D}\cap(t,\infty)$ such that $s_n \downarrow t$, then $X_{s_n} \to X_{t+}$ in $L^1(P)$ as $n \to \infty$.*
5. *The process $\{\bar{X}_t\}_{t\in\mathbb{R}_+}$ is a $\{\mathcal{B}_t^+\}_{t\ge0}$ submartingale such that $t \to \bar{X}_t$ is right continuous and has left limits on $\Omega_0$.*
6. *$X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. for all $t \in \mathbb{D}$ with equality at some $t \in \mathbb{D}$ iff $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$.*
7. *If $X_s \xrightarrow{P} X_t$ as $\mathbb{D} \ni s \downarrow t$ at some $t \in \mathbb{D}$,[3] then $\bar{X}_t = X_t$ a.s.*
8. *If $C := \sup_{t\in\mathbb{D}} \mathbb{E}|X_t| < \infty$ (or equivalently $\sup_{t\in\mathbb{D}} \mathbb{E}X_t^+ < \infty$), then $X_\infty := \lim_{\mathbb{D}\ni t\uparrow\infty} \bar{X}_t = \lim_{\mathbb{D}\ni t\uparrow\infty} X_t$ exists in $\mathbb{R}$ a.s. and $\mathbb{E}|X_\infty| < C < \infty$.*
   ***Note:** if $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable then $\sup_{t\in\mathbb{D}} \mathbb{E}\left|X_t^+\right| < \infty$.*
9. *If $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable iff there exists $X_\infty \in L^1(\Omega,\mathcal{B},P)$ such that $\{X_t\}_{t\in\mathbb{D}\cup\{\infty\}}$ is a submartingale. In other words, $\{X_t^+\}_{t\in\mathbb{D}}$ is uniformly integrable iff here exists $X_\infty \in L^1(\Omega,\mathcal{B},P)$ such that $X_t \le \mathbb{E}[X_\infty|\mathcal{B}_t]$ a.s. for all $t \in \mathbb{D}$.*

**Proof.** 1. – 3. The fact that $P(\Omega_0) = 1$ follows from Doob's upcrossing inequality and the maximal inequality in Eq. (30.11). The assertions in items 2. and 3. are now a consequence of the definition of $\Omega_0$ and Lemmas 30.21 and 30.22.

4. Let $Y_n := X_{s_{-n}}$ and $\mathcal{F}_n := \mathcal{B}_{s_{-n}}$ for $-n \in \mathbb{N}$. Then $\{(Y_n, \mathcal{F}_n)\}_{n\in-\mathbb{N}}$ is a backwards submartingale such that $\inf \mathbb{E}Y_n \ge \mathbb{E}X_t$ and hence by Theorem 20.79, $Y_n = X_{s_{-n}} \to X_{t+}$ in $L^1(P)$ as $n \to -\infty$.

5. Since $\bar{X}_t = X_{t+}$ on $\Omega_0$ and $X_{t+}$ is right continuous with left hand limits, $\bar{X}$ has these properties on $\Omega_0$ as well. Now let $0 \le s < t < \infty$, $\{s_n\}, \{t_n\} \subset \mathbb{D}$, such that $s_n \downarrow s$, $t_n \downarrow t$ with $s_n < t$ for all $n$. Then by item 4. and the submartingale property of $X$,

$$\mathbb{E}\left[\bar{X}_t - \bar{X}_s : A\right] = \mathbb{E}\left[X_{t+} - X_{s+} : A\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n} - X_{s_n} : A\right] \ge 0$$

for all and $A \in \mathcal{B}_{s+}$.

6. Let $A \in \mathcal{B}_t$ and $\{t_n\} \subset \mathbb{D}$ with $t_n \downarrow t \in \mathbb{D}$, then

$$\mathbb{E}\left[\bar{X}_t : A\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n} : A\right] \ge \lim_{n\to\infty} \mathbb{E}\left[X_t : A\right].$$

Since $A \in \mathcal{B}_t$ is arbitrary it follows that $X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. If equality holds, then, taking $A = \Omega$ above, we find

$$\mathbb{E}X_t = \mathbb{E}\bar{X}_t = \lim_{n\to\infty} \mathbb{E}\left[X_{t_n}\right].$$

Since $\{t_n\} \subset \mathbb{D}$ with $t_n \downarrow t$ was arbitrary, we may conclude that $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$. Conversely if $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}X_s = \mathbb{E}X_t$, then along any sequence, $\{s_n\} \subset \mathbb{D}$ with $s_n \downarrow s$, we have

$$\mathbb{E}X_t = \lim_{n\to\infty} \mathbb{E}X_{s_n} = \mathbb{E}\lim_{n\to\infty} X_{s_n} = \mathbb{E}\bar{X}_t = \mathbb{E}\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right].$$

As $X_t \le \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s. this identity implies $X_t = \mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]$ a.s.

[3] For example, this will hold if $\lim_{\mathbb{D}\ni s\downarrow t} \mathbb{E}|X_t - X_s| = 0$.

7. Let $t_n \in \mathbb{D}$ such that $t_n \downarrow t$, then as we have already seen $X_{t_n} \to \bar{X}_t$ in $L^1(P)$. However by assumption, $X_{t_n} \overset{P}{\to} X_t$, and therefore we must have $\bar{X}_t = X_t$ a.s. since limits in probability are unique up to null sets.

The proof or items 8. and 9. will closely mimic their discrete versions given in Corollary 20.56.

8. The proof here mimics closely the discrete version given in Corollary 20.56. For any $-\infty < a < b < \infty$, Doob's upcrossing inequality (Theorem 30.23) and the MCT implies,

$$\begin{aligned}\mathbb{E}\left[U_\infty^X(a,b)\right] &= \lim_{\mathbb{D}\ni T\to\infty} \mathbb{E}\left[U_T^X(a,b)\right] \\ &\le \frac{1}{b-a}\left[\sup_{T\in\mathbb{D}} \mathbb{E}(X_T - a)_+ - \mathbb{E}(X_0-a)_+\right] < \infty\end{aligned}$$

where

$$U_\infty^X(a,b) = \lim_{\mathbb{D}\ni T\to\infty} U_T^X(a,b)$$

is the total number of upcrossings of $X$ across $[a,b]$. In particular it follows that

$$\tilde{\Omega}_0 := \cap\left\{U_\infty^X(a,b) < \infty : a,b\in\mathbb{Q} \text{ with } a<b\right\}$$

has probability one. Hence by Lemma 30.21, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{\mathbb{D}\ni t\to\infty} X_t(\omega)$ exists in $\bar{\mathbb{R}}$. By Fatou's lemma with $\mathbb{D} \ni t_n \uparrow \infty$, it follows that

$$\mathbb{E}\left[|X_\infty|\right] = \mathbb{E}\left[\liminf_{n\to\infty} |X_n|\right] \le \liminf_{n\to\infty} \mathbb{E}\left[|X_n|\right] \le C < \infty$$

and therefore that $X_\infty \in \mathbb{R}$ a.s.

9. If $\left\{X_t^+\right\}_{t\ge 0}$ is uniformly integrable, then, by Vitalli's convergence Theorem 14.51 and the fact that $X_t^+ \to X_\infty^+$ a.s. (as we have already shown), $X_t^+ \to X_\infty^+$ in $L^1(P)$. Therefore for $A \in \mathcal{B}_t$ we have, by Fatou's lemma, that

$$\begin{aligned}\mathbb{E}\left[X_t 1_A\right] &\le \limsup_{\mathbb{D}\ni s\to\infty} \mathbb{E}\left[X_s 1_A\right] = \limsup_{\mathbb{D}\ni s\to\infty}\left(\mathbb{E}\left[X_s^+ 1_A\right] - \mathbb{E}\left[X_s^- 1_A\right]\right) \\ &= \mathbb{E}\left[X_\infty^+ 1_A\right] - \liminf_{\mathbb{D}\ni s\to\infty} \mathbb{E}\left[X_s^- 1_A\right] \le \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[\liminf_{\mathbb{D}\ni s\to\infty} X_s^- 1_A\right] \\ &= \mathbb{E}\left[X_\infty^+ 1_A\right] - \mathbb{E}\left[X_\infty^- 1_A\right] = \mathbb{E}\left[X_\infty 1_A\right].\end{aligned}$$

Since $A \in \mathcal{B}_t$ was arbitrary we may conclude that $X_t \le \mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ a.s. for all $t \in \mathbb{R}_+$.

Conversely if we suppose that $X_t \le \mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ a.s. for all $t\in\mathbb{R}_+$, then by Jensen's inequality, $X_t^+ \le \mathbb{E}\left[X_\infty^+|\mathcal{B}_t\right]$ and therefore $\left\{X_t^+\right\}_{t\ge0}$ is uniformly integrable by Proposition 20.8 and Exercise 14.5. ■

*Example 30.25.* In this example we show that there exists a right continuous submartingale, $\{X_t\}_{t\ge0}$, such that $\{X_{s_n}\}_{n=1}^\infty$ is **not** uniformly integrable for some bounded increasing sequence $\{s_n\}_{n=1}^\infty$. Indeed, let

$$X_t := -M^0_{\tan\left(\frac{\pi}{2}t\wedge 1\right)},$$

where $\{M_t\}_{t\ge0}$ is the martingale constructed in Exercise 30.1. Then it is easily checked that $\{X_t\}_{t\ge0}$ is a $\left\{\mathcal{B}_{\tan\left(\frac{\pi}{2}t\wedge1\right)}\right\}_{t\ge0}$ – submartingale. Moreover if $s_n \in [0,1)$ with $s_n \uparrow 1$, the collection, $\{X_{s_n}\}_{n=1}^\infty$ is not uniformly integrable for if it were we would have

$$-1 = \lim_{n\to\infty} \mathbb{E}X_{s_n} = \mathbb{E}\left[\lim_{n\to\infty} X_{s_n}\right] = \mathbb{E}[0] = 0.$$

In particular this shows that in item 4. of Theorem 30.24, we can **not** suppose $\{s_n\}_{n=1}^\infty \subset \mathbb{D}\cap[0,t)$ with $s_n \uparrow t$.

**Exercise 30.2.** If $\{X_t\}_{t\ge0}$ is a right continuous submartingale on a filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0},P\right)$, then $s\to\mathbb{E}X_s$ is right continuous at $t$ and $X$ is a $\left\{\mathcal{B}_t^+\right\}$ – submartingale.

**Exercise 30.3.** Let $\{X_t\}_{t\ge0}$ be a submartingale on a filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0},P\right)$, $t\in\mathbb{R}_+$, and $X_s \overset{P}{\to} X_t$ as $s\downarrow t$, then $s\to\mathbb{E}X_s$ is right continuous at $t$.

**Theorem 30.26 (Regularizing Submartingales).** *Let $\{X_t\}_{t\ge0}$ be a submartingale on a filtered probability space, $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0},P\right)$, and let $\Omega_0$ and $\bar{X}_t$ be as in Theorem 30.24 applied to $\{X_t\}_{t\in\mathbb{D}}$. Further let*

$$\hat{X}_t(\omega) := \begin{cases}\bar{X}_t(\omega) & \text{if } \omega\in\Omega_0\\ 0 & \text{if } \omega\notin\Omega_0,\end{cases}$$

*and $\left\{\bar{\mathcal{B}}_{t+} := \mathcal{B}_{t+}\vee\mathcal{N}\right\}_{t\ge0}$ where $\mathcal{N}$ is the collection of $P$ – null subsets of $\Omega$ in $\mathcal{B}$. Then:*

1. *$\left\{\hat{X}_t\right\}_{t\ge0}$ is a $\left\{\bar{\mathcal{B}}_{t+}\right\}_{t\ge0}$ – submartingale which is right continuous with left hand limits.*
2. *$\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right] \ge X_t$ a.s. for all $t\in\mathbb{R}_+$ with equality holding for all $t\in\mathbb{R}_+$ iff $t\to\mathbb{E}X_t$ is right continuous.*
3. *If $\{\mathcal{B}_t\}_{t\ge0}$ is right continuous, then $\hat{X}_t \ge X_t$ a.s. for all $t\in\mathbb{R}_+$ with equality holding for all $t\in\mathbb{R}_+$ iff $t\to\mathbb{E}X_t$ is right continuous.*

*In particular if* $\{X_t\}_{t\ge0}$ *is right continuous in probability, then* $\{X_t\}_{t\ge0}$ *has a right continuous modification possessing left hand limits,* $\left\{\hat{X}_t\right\}_{t\ge0}$*, such that* $\left\{\hat{X}_t\right\}_{t\ge0}$ *is a* $\{\bar{\mathcal{B}}_{t+}\}_{t\ge0}$ *– submartingale.*

**Proof.** 1. Since $\Omega\setminus\Omega_0\in\mathcal{N}$ and $\bar{X}_t$ is a $\mathcal{B}_{t+}$ – measurable, it follows that $\hat{X}_t$ is $\bar{\mathcal{B}}_{t+}$ – measurable. Hence $\left\{\hat{X}_t\right\}_{t\ge0}$ is an adapted process. Since $\hat{X}$ is a modification of $\bar{X}$ which is already a $\{\bar{\mathcal{B}}_{t+}\}_{t\ge0}$ – submartingale (see Lemma 30.15) it follows that $\hat{X}$ is also a $\{\bar{\mathcal{B}}_{t+}\}_{t\ge0}$ – submartingale.

2. Since $\hat{X}_t=\bar{X}_t$, we may replace $\hat{X}$ by $\bar{X}$ in the statement 2. We now need only follow the proof of item 6. in Theorem 30.26. Indeed, if $t\in\mathbb{R}_+$, $A\in\mathcal{B}_t$, and $\{t_n\}\subset\mathbb{D}$ with $t_n\downarrow t\in\mathbb{D}$, then

$$\mathbb{E}\left[\bar{X}_t:A\right]=\lim_{n\to\infty}\mathbb{E}\left[X_{t_n}:A\right]\ge\lim_{n\to\infty}\mathbb{E}\left[X_t:A\right].$$

Since $A\in\mathcal{B}_t$ was arbitrary, it follows that

$$X_t\le\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]\text{ a.s.}\tag{30.22}$$

If equality holds in Eq. (30.22), $\mathbb{E}X_t=\mathbb{E}\hat{X}_t$ which is right continuous by Exercise 30.2. Conversely if $t\to\mathbb{E}X_t$ is right continuous, it follows from item 6. of Theorem 30.26 that $\hat{X}_t=X_t$ a.s.

3. Since $\hat{X}_t=\bar{X}_t$ a.s. and $\bar{X}_t$ is $\mathcal{B}_{t+}=\mathcal{B}_t$ – measurable, we find

$$\hat{X}_t=\bar{X}_t=\mathbb{E}\left[\bar{X}_t|\mathcal{B}_t\right]=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]\text{ a.s.}$$

With this observation (i.e. $\hat{X}_t=\mathbb{E}\left[\hat{X}_t|\mathcal{B}_t\right]$ a.s.) the assertions in item 3. follow directly from those in item 2.

Now suppose that $\{X_t\}_{t\ge0}$ is right continuous in probability. By Proposition 30.16 and Lemma 30.15, $\{X_t\}_{t\ge0}$ is a $\{\bar{\mathcal{B}}_{t+}\}_{t\ge0}$ – submartingale such that, by Exercise 30.3, $t\to\mathbb{E}X_t$ is right continuous. Therefore, by items 1. and 3. of the theorem, $\hat{X}$ defined above is the desired modification of $X$. ■

*Example 30.27.* Let be a Poisson process, $\{N_t\}_{t\ge0}$, with parameter $\lambda$ as described in Example 27.9. Since $\{N\}_{t\ge0}$ has independent increments, it follows that $\{N_t\}_{t\ge0}$ is a $\{\mathcal{B}_t:=\sigma(N_s:s\le t)\}_{t\ge0}$ – martingale. By example 24.14) we know that $\mathbb{E}|N_t-N_s|=\lambda|t-s|$ and in particular, $t\to N_t$ is continuous in probability. Hence it follows from Theorem 30.26 that there is a modification, $\hat{N}$ and $N$ such that $\left\{\hat{N}_t\right\}_{t\ge0}$ is a $\{\bar{\mathcal{B}}_{t+}\}_{t\ge0}$ – martingale which has continuous sample paths possessing left hand limits.

The ideas of this example significantly generalize to produce good modifications of large classes of Markov processes, see for example [5, Theorem I.9.4 on p. 46], [17] and [30]. See [36, Chapter I] where this is carried out in the context of Lévy processes. We end this section with another version of the optional sampling theorem.

**Theorem 30.28 (Optional sampling II).** *Suppose* $\{X_t\}_{t\ge0}$ *is a right continuous submartingale on a filtered probability space,* $\left(\Omega,\mathcal{B},\{\mathcal{B}_t\}_{t\ge0},P\right)$ *such that* $\{X_t^+\}_{t\ge0}$ *is uniformly integrable. Then for any two optional times,* $\sigma$ *and* $\tau$*,* $X_\tau\in L^1(P)$ *and*

$$X_{\sigma\wedge\tau}\le\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right].\tag{30.23}$$

*In particular if* $\{M_t\}_{t\ge0}$ *is a right continuous uniformly integrable martingale, then*

$$M_{\sigma\wedge\tau}=\mathbb{E}\left[M_\tau|\mathcal{B}_\sigma^+\right].\tag{30.24}$$

**Proof.** Let $X_\infty:=\lim_{\mathbb{D}\ni t\uparrow\infty}X_t\in L^1(P)$ as in Theorem 30.24 so that $X_t\le\mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$ for all $t\in\mathbb{D}$. For $t\in\mathbb{R}_+$, let $\{t_n\}_{n=1}^\infty\subset\mathbb{D}\cap(t,\infty)$ be such that $t_n\downarrow t$, then by Corollary 20.81,

$$X_t=\lim_{n\to\infty}X_{t_n}\le\lim_{n\to\infty}\mathbb{E}\left[X_\infty|\mathcal{B}_{t_n}\right]=\mathbb{E}\left[X_\infty|\mathcal{B}_t^+\right]\text{ a.s.}$$

Conditioning this inequality on $\mathcal{B}_t$ also allows us to conclude that $X_t\le\mathbb{E}\left[X_\infty|\mathcal{B}_t\right]$.[4] We may now reduce the inequality in Eq. (30.23) to the case in Theorem 30.9 where $\tau$ is a bounded stopping time by simply identifying $[0,\pi/2]$ with $[0,\infty]$ via the map, $t\to\tan t$. More precisely, let $\{Y_t\}_{0\le t\le\pi/2}$ be the right continuous $\left\{\tilde{\mathcal{B}}_t:=\mathcal{B}_{\tan t}\right\}_{0\le t\le\pi/2}$ – submartingale defined by $Y_t:=X_{\tan t}$. As $\tan^{-1}(\sigma)$ and $\tan^{-1}(\tau)$ are two bounded $\left\{\tilde{\mathcal{B}}_t\right\}_{0\le t\le\pi/2}$ – optional times, we may apply Theorem 30.9 to find;

$$X_{\sigma\wedge\tau}=Y_{\tan^{-1}(\sigma)\wedge\tan^{-1}(\tau)}\le\mathbb{E}\left[Y_{\tan^{-1}(\tau)}|\tilde{\mathcal{B}}^+_{\tan^{-1}(\sigma)}\right]=\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma^+\right]\text{ a.s.}$$

For the martingale assertions, simply apply Eq. (30.23) with $X_t=M_t$ and $X_t=-M_t$. ■

[4] According to Exercise 30.2, $\{X_t\}_{t\ge0}$ is also a $\{\mathcal{B}_t^+\}$ – submartingale. Therefore for the purposes of this Theorem, there is no loss in generality in assuming that $\mathcal{B}_t^+=\mathcal{B}_t$.

# References


1. David Applebaum, *Lévy processes and stochastic calculus*, second ed., Cambridge Studies in Advanced Mathematics, vol. 116, Cambridge University Press, Cambridge, 2009. MR MR2512800
2. Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR MR1324786 (95k:60001)
3. ———, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR MR1700749 (2000e:60008)
4. Ottó J. Björnsson, *A note on the characterization of standard Borel spaces*, Math. Scand. **47** (1980), no. 1, 135–136. MR MR600083 (82a:54070)
5. R. M. Blumenthal and R. K. Getoor, *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York, 1968. MR MR0264757 (41 #9348)
6. Leo Breiman, *Probability*, Addison-Wesley Publishing Company, Reading, Mass., 1968. MR MR0229267 (37 #4841)
7. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
8. J. L. Doob, *Stochastic processes*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR MR1038526 (91d:60002)
9. L. E. Dubins and E. H. Spanier, *How to cut a cake fairly*, Amer. Math. Monthly **68** (1961), 1–17. MR MR0129031 (23 #B2068)
10. Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. MR MR1009162 (90g:47001a)
11. ———, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. MR MR1009162 (90g:47001a)
12. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
13. John Elton, Theodore P. Hill, and Robert P. Kertz, *Optimal-partitioning inequalities for nonatomic probability measures*, Trans. Amer. Math. Soc. **296** (1986), no. 2, 703–725. MR MR846603 (87m:60012)
14. Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, Characterization and convergence. MR MR838085 (88a:60130)
15. William Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. MR MR0270403 (42 #5292)
16. Peter K. Friz and Nicolas B. Victoir, *Multidimensional stochastic processes as rough paths*, Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, 2010, Theory and applications. MR MR2604669
17. Masatoshi Fukushima, Yōichi Ōshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR MR1303354 (96f:60126)

18. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
19. Paul R. Halmos, *Lectures on ergodic theory*, (1960), vii+101. MR MR0111817 (22 #2677)
20. Theodore P. Hill, *Partitioning general probability measures*, Ann. Probab. **15** (1987), no. 2, 804–813. MR MR885145 (88d:60014)
21. Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 1987. MR MR959133 (89k:60044)
22. ———, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR MR1943877 (2003j:60001)
23. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
24. Martin L. Jones, *A note on a cake cutting algorithm of Banach and Knaster*, Amer. Math. Monthly **104** (1997), no. 4, 353–355. MR MR1450669 (99e:60012)
25. Shizuo Kakutani, *On equivalence of infinite product measures*, Ann. of Math. (2) **49** (1948), 214–224. MR MR0023331 (9,340e)
26. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
27. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR MR1121940 (92h:60127)
28. Oliver Knill, *Probability and stochastic processes with applications*, Havard Web-Based, 1994.
29. Leonid B. Koralov and Yakov G. Sinai, *Theory of probability and random processes*, second ed., Universitext, Springer, Berlin, 2007. MR MR2343262 (2008k:60002)
30. Zhi Ming Ma and Michael Röckner, *Introduction to the theory of (nonsymmetric) Dirichlet forms*, Universitext, Springer-Verlag, Berlin, 1992. MR MR1214375 (94d:60119)
31. Ivan Netuka, *The change-of-variables theorem for the Lebesgue integral*, Acta Univ. M. Belii Ser. Math. **19** (2011), 37–42. MR 2853669
32. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
33. ———, *Probability and measure*, Tech. report, Mathematics Department, University of Cambridge, 2009.
34. K. R. Parthasarathy, *Probability measures on metric spaces*, AMS Chelsea Publishing, Providence, RI, 2005, Reprint of the 1967 original. MR MR2169627 (2006d:60004)
35. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
36. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
37. Philip E. Protter, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR MR2273672
38. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR 2001g:60188
39. H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR MR1013117 (90g:00004)
40. Gennady Samorodnitsky and Murad S. Taqqu, *Stable non-Gaussian random processes*, Stochastic Modeling, Chapman & Hall, New York, 1994, Stochastic models with infinite variance. MR MR1280932 (95f:60024)
41. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)
42. Barry Simon, *Functional integration and quantum physics*, second ed., AMS Chelsea Publishing, Providence, RI, 2005. MR MR2105995 (2005f:81003)
43. Daniel W. Stroock, *Probability theory, an analytic view*, Cambridge University Press, Cambridge, 1993. MR MR1267569 (95f:60003)
44. S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes in Mathematics, vol. 7, New York University Courant Institute of Mathematical Sciences, New York, 2001. MR MR1852999 (2003a:60001)

# 31

## Homework Problem Solutions

### 31.1 Resnik repeats