# Contents

## Part Homework Problems

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3 Math 280A Homework Problems Fall 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.1 Homework 1. Due Friday, October 5, 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.2 Homework 2. Due Friday, October 12, 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.3 Homework 3. Due Friday, October 19, 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.4 Homework 4. Due a, October 26, 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.5 Homework 5. Due Friday, November 2, 2018</td>
<td>3</td>
</tr>
<tr>
<td>-3.6 Homework 6. Due Friday, November 9, 2018</td>
<td>3</td>
</tr>
</tbody>
</table>

## Part I Background Material

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Limsups, Liminfs and Extended Limits</td>
<td>7</td>
</tr>
<tr>
<td>1.1 Infinite sums</td>
<td>9</td>
</tr>
<tr>
<td>2 Basic Metric and Topological Space Notions</td>
<td>13</td>
</tr>
<tr>
<td>2.1 Metric spaces as topological spaces</td>
<td>14</td>
</tr>
<tr>
<td>2.1.1 Continuity</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Completeness in Metric Spaces</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Sequential compactness</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Supplementary Remarks</td>
<td>18</td>
</tr>
<tr>
<td>2.4.1 Word of Caution</td>
<td>18</td>
</tr>
<tr>
<td>2.4.2 Riemannian Metrics</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Exercises</td>
<td>20</td>
</tr>
<tr>
<td>3 Basic Probabilistic Notions</td>
<td>21</td>
</tr>
</tbody>
</table>
## Part II Formal Development

### 4 Preliminaries
- 4.1 Set Operations ............................................................................................................. 27
- 4.2 Exercises ......................................................................................................................... 27
- 4.3 Algebraic sub-structures of sets ....................................................................................... 29

### 5 Finitely Additive Measures / Integration
- 5.1 Examples of Measures .................................................................................................. 34
- 5.2 Additive probabilities on infinite products of a discrete space
  - 5.2.1 Finitely additive measures on $\mathbb{R}$ .................................................................... 35
- 5.3 *Finitely additive measures on product spaces
- 5.4 Simple Random Variables
  - 5.4.1 The algebraic structure of simple functions*
- 5.5 Simple Integration......................................................................................................... 39
- 5.5.1 *Appendix: Bonferroni Inequalities ............................................................................. 45
- 5.6 *Finitely Additive Measures on $\mathbb{R}^d$ and $[0,1]^d$ .................................................. 46
- 5.7 *Appendix: Riemann Stieljes integral............................................................................ 48
- 5.8 *Appendix: Tonelli and Fubini’s Theorem I ............................................................... 49
- 5.9 Conditional probabilities and expectations ................................................................. 50
- 5.10 Simple Independence and the Weak Law of Large Numbers
  - 5.10.1 Complex Weierstrass Approximation Theorem................................................... 55
- 5.11 *Appendix: A Multi-dimensional Weierstrass Approximation Theorem ................. 58

### 6 Countably Additive Measures
- 6.1 Outer Measures ............................................................................................................. 62
- 6.2 Construction of $\sigma$–Additive Finite Measures .......................................................... 63
- 6.3 A Discrete Kolmogorov’s Extension Theorem ............................................................... 66
- 6.4 Construction of $\sigma$ – Finite Measures ........................................................................ 68
- 6.5 Radon Measures on $\mathbb{R}$ ........................................................................................... 69
- 6.6 Lebesgue Measure ......................................................................................................... 73
- 6.7 Metric-Measure Space Regularity Results
  - 6.7.1 Regularity Exercises for Borel measures on $(X, \mathcal{B}_X)$ ........................................ 76
- 6.8 * The $\pi$ – $\lambda$ theorem .............................................................................................. 76
- 6.9 * Supplement: Generalizations of Theorem 6.34 to $\mathbb{R}^d$ ........................................ 78
- 6.10 *Appendix: Completions of Measure Spaces ............................................................ 79

### 7 *Measure construction extras
- 7.1 Overview ....................................................................................................................... 81
- 7.1.1 An Extension of Measure Strategy .......................................................................... 81
- 7.2 Monotone Class Theorems ............................................................................................ 82
- 7.3 * $\sigma$ – Algebra Regularity and Uniqueness Results .................................................... 82

---

**Contents**

4.3 Algebraic sub-structures of sets

5.11 *Appendix: A Multi-dimensional Weirstrass Approximation Theorem...

6.4 Construction of

7.3 * $\sigma$ – Algebra Regularity and Uniqueness Results
12 Multiple and Iterated Integrals

12.1 Iterated Integrals ............................................................................................................ 149
12.2 Tonelli’s Theorem and Product Measure ........................................................................... 149
12.3 Fubini’s Theorem .......................................................................................................... 149
12.4 Fubini’s Theorem and Completions* ................................................................................. 151
12.5 Lebesgue Measure on $\mathbb{R}^n$ and the Change of Variables Theorem ......................... 154
12.6 Other change of variables proofs ..................................................................................... 161
  12.6.1 $\delta$-function localization proof of the change of variables Theorem 12.6.2 ............................................................................................................. 162
  12.6.2 Radon-Nykodym proof ................................................................................................ 162
12.7 The Polar Decomposition of Lebesgue Measure* ............................................................... 166
12.8 More Spherical Coordinates* .......................................................................................... 168
12.9 Gaussian Random Vectors ............................................................................................. 170
  12.9.1 *Gaussian measures with possibly degenerate covariances .................................... 172
12.10 Kolmogorov’s Extension Theorems .................................................................................. 173
  12.10.1 Regularity and compactness results ........................................................................... 173
  12.10.2 Kolmogorov’s Extension Theorem and Infinite Product Measures ....................... 174
12.11 Appendix: Standard Borel Spaces* .................................................................................. 176
12.12 More Exercises ............................................................................................................... 179
12.13 Exercises ....................................................................................................................... 180

13 Independence ...................................................................................................................... 183

13.1 Basic Properties of Independence ................................................................................... 183
13.2 Examples of Independence .............................................................................................. 189
  13.2.1 An Example of Ranks ................................................................................................ 189
13.3 Independence for Gaussian Random Vectors .................................................................... 190
13.4 Summing independent random variables ......................................................................... 191
13.5 A Strong Law of Large Numbers ..................................................................................... 193
13.6 A Central Limit Theorem ............................................................................................... 193
13.7 The Second Borel-Cantelli Lemma .................................................................................. 197
13.8 Kolmogorov and Hewitt-Savage Zero-One Laws ............................................................... 200
  13.8.1 Hewitt-Savage Zero-One Law .................................................................................... 201
13.9 Another Construction of Independent Random Variables* .......................................... 202

14 The Standard Poisson Process .......................................................................................... 207

14.1 Poisson Random Variables ............................................................................................. 207
14.2 Exponential Random Variables ....................................................................................... 208
  14.2.1 Appendix: More properties of Exponential random Variables* ............................ 208
14.3 The Standard Poisson Process ....................................................................................... 210
14.4 Poisson Process Extras* ............................................................................................... 213
## 15 $L^p$ – spaces

15.1 Modes of Convergence .......................................................... 215
15.2 Almost Everywhere and Measure Convergence ......................... 216
15.3 Jensen’s, Holder’s and Minkowski’s Inequalities ....................... 219
15.4 Completeness of $L^p$ – spaces ................................................. 221
15.5 Density Results ................................................................. 222
15.6 Relationships between different $L^p$ – spaces ......................... 223
15.6.1 Summary: ........................................................................ 225
15.7 Uniform Integrability ............................................................. 225
15.7.1 Atoms ............................................................................. 229
15.8 Exercises ............................................................................. 230
15.9 Appendix: Convex Functions ................................................ 230

## 16 Hilbert Space Basics

16.1 Compactness Results for $L^p$ – Spaces* .................................. 237
16.2 Exercises ............................................................................. 242

## 17 Conditional Expectation

17.1 Examples: ............................................................................. 245
17.1.1 Conditioning Gaussian Random Vectors ......................... 247
17.1.2 Probability Kernels and Regular Conditional Distributions .. 249
17.2 Additional Properties of Conditional Expectations ................. 252
17.3 Conditional Independence .................................................... 253
17.4 Construction of Regular Conditional Distributions* ............... 255

## 18 The Radon-Nikodym Theorem

18.1 Proof of the Change of Variables Theorem* 12.21 .................. 261

## 19 Some Ergodic Theory

20 The Markov Property

20.1 Markov Processes .................................................................. 275
20.2 Discrete Time Homogeneous Markov Processes ..................... 276
20.3 Continuous time homogeneous Markov processes .................. 280
20.4 Continuous Time Markov Chains on Denumerable State Spaces .. 281
20.5 First Step Analysis and Hitting Probabilities ......................... 287
20.6 Finite state space chains ....................................................... 290
  20.6.1 Invariant distributions and return times ......................... 293
  20.6.2 Some worked examples ................................................. 302
  20.6.3 Exercises ..................................................................... 303

Part III Stochastic Processes I

## 20.6 Finite state space chains

20.6.1 Invariant distributions and return times ......................... 302
20.6.2 Some worked examples ................................................. 303
20.6.3 Exercises ..................................................................... 305
### 20.9 Appendix: More Probability Kernel Constructions

- 20.9.1 Removing the standard Borel restriction
- 20.9.2 Appendix: Kolmogorov’s extension theorem II

### 21 (Sub and Super) Martingales

- 21.1 (Sub and Super) Martingale Examples
- 21.2 Decompositions
- 21.3 Stopping Times
  - 21.3.1 Summary of some of the more notable Chapter 21 convergence results
- 21.4 Stochastic Integrals and Optional Stopping
- 21.5 Submartingale Maximal Inequalities
- 21.6 Submartingale Upcrossing Inequality and Convergence Theorems
- 21.7 *Supermartingale inequalities
  - 21.7.1 Maximal Inequalities
  - 21.7.2 The upcrossing inequality and convergence result
- 21.8 Martingale Closure and Regularity Results
- 21.9 Backwards (Reverse) Submartingales
- 21.10 Some More Martingale Exercises
  - 21.10.1 More Random Walk Exercises
- 21.11 Appendix: Some Alternate Proofs

### 22 Some Martingale Examples and Applications

- 22.1 Aside on Large Deviations
- 22.2 A Polya Urn Model
- 22.3 Galton Watson Branching Process
- 22.4 Kakutani’s Theorem

### Part IV (Weak) Convergence of Random Sums

### 23 Random Sums

- 23.1 Weak Laws of Large Numbers
  - 23.1.1 A WLLN Example
- 23.2 Kolmogorov’s Convergence Criteria
- 23.3 The Strong Law of Large Numbers Revisited
  - 23.3.1 Strong Law of Large Number Examples
- 23.4 Kolmogorov’s Three Series Theorem
  - 23.4.1 Examples
- 23.5 Maximal Inequalities
- 23.6 Bone Yards
  - 23.6.1 Kronecker’s Lemma
  - 23.6.2 Older variants on the proof of Kolmogorov’s three series Theorem
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>28 Versions and Modifications</strong></td>
<td></td>
</tr>
<tr>
<td>28.1 Kolmogorov’s Continuity Criteria</td>
<td>463</td>
</tr>
<tr>
<td>28.2 Kolmogorov’s Tightness Criteria</td>
<td>466</td>
</tr>
<tr>
<td><strong>29 Brownian Motion I</strong></td>
<td></td>
</tr>
<tr>
<td>29.1 Donsker’s Invariance Principle</td>
<td>469</td>
</tr>
<tr>
<td>29.2 Path Regularity Properties of BM</td>
<td>470</td>
</tr>
<tr>
<td>29.3 Scaling Properties of B. M.</td>
<td>474</td>
</tr>
<tr>
<td><strong>30 Filtrations and Stopping Times</strong></td>
<td></td>
</tr>
<tr>
<td>30.1 Measurability Structures</td>
<td>475</td>
</tr>
<tr>
<td>30.2 Stopping and optional times</td>
<td>476</td>
</tr>
<tr>
<td>30.3 Filtration considerations</td>
<td></td>
</tr>
<tr>
<td>30.3.1 ***More Augmentation Results (This subsection need serious editing.)</td>
<td>481</td>
</tr>
<tr>
<td><strong>31 Continuous time (sub)martingales</strong></td>
<td></td>
</tr>
<tr>
<td>31.1 Submartingale Inequalities</td>
<td>485</td>
</tr>
<tr>
<td>31.2 Regularizing a submartingale</td>
<td>487</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td></td>
</tr>
<tr>
<td><strong>32 Homework Problem Solutions</strong></td>
<td></td>
</tr>
<tr>
<td>32.1 Resnik repeats</td>
<td>497</td>
</tr>
</tbody>
</table>
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. In the lecture note problems listed in the assignments, you should look up the corresponding problem in the lecture notes where more context is given including extra standing assumptions for the problem. This context and standing assumptions are not always extracted out when I construct the homework sheets.

-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

-3.2 Homework 2. Due Friday, October 12, 2018
- Lecture note Exercises: 4.2, 4.3, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12
- Look at Resnick, p. 20-27: 9, 12, 17, 23
- Hand in Resnick, p. 20-27: 5, 18, 40*

*Notes on Resnick’s #40: (i) $B((0,1])$ should be $B([0,1))$ in the statement of this problem, (ii) $k$ is an integer, (iii) $r \geq 2$.

-3.3 Homework 3. Due Friday, October 19, 2018
- Look at Resnick, p. 20-27: 9, 12, 17, 23
- Hand in Resnick, p. 20-27: 5, 18, 40*

-3.4 Homework 4. Due a, October 26, 2018
- Hand in Resnick exercises: § 2.6, #7* and § 2.6, #13.

*Hint: For Resnick #7 you might label the coupons as $\{1, 2, \ldots, N\}$ and let $A_i$ be the event that the collector does not have the $i^{th}$ - coupon after buying $n$ - boxes of cereal.

-3.5 Homework 5. Due Friday, November 2, 2018
- Look at Resnick, § 2.6, p. 63–70; 3, 14
- Look at lecture note exercises: 6.1
- Hand in Resnick, § 2.6, p. 63–70; 6, 11
- Hand in lecture note exercises: 6.2, 6.3, 9.4

-3.6 Homework 6. Due Friday, November 9, 2018
- Hand in Lecture note Exercises: 6.5, 6.6, 9.1, 9.2, 9.5
- Look at Lecture note Exercises: 9.3, 9.6
- Look at Resnick, p. 85–90: 3, 7, 8, 12, 17, 21
- Hand in from Resnick, p. 85–90: 4, 6*, 9, 15, 18**. Notes. * In #6, the random variable $X$ is understood to take values in the extended real numbers.
- ** In #18, I would write the left side in terms of an expectation.
Background Material
Limsups, Liminfns 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, \( \infty \cdot 0 = 0, \infty \cdot a = \infty \) if \( a \in \mathbb{R} \) with \( a > 0 \), \( \infty \cdot a = -\infty \) if \( a \in \mathbb{R} \) with \( a < 0 \), \( \infty + \infty = \infty \) and \( -\infty - \infty = -\infty \) while \( \infty - \infty \) is not defined. A sequence \( a_n \in \bar{\mathbb{R}} \) is said to converge to \( -\infty \) 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 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. \( \{a_nb_n\}_{n=1}^\infty \) is convergent and
\[
\lim_{n \to \infty} (a_nb_n) = \lim_{n \to \infty} a_n \cdot \lim_{n \to \infty} b_n \tag{1.2}
\]
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
\[
\lim a_n + \lim b_n = "\infty - \infty".
\]
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. (1.2). 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
\[
\lvert a - a_n \rvert \leq \varepsilon \text{ and } \lvert b - b_n \rvert \leq \varepsilon \text{ for all } n \geq N.
\]
Therefore,
\[
\lvert a + b - (a_n + b_n) \rvert = \lvert a - a_n + b - b_n \rvert \leq \lvert a - a_n \rvert + \lvert b - b_n \rvert \leq 2\varepsilon
\]
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 \( \alpha := \min \left( \frac{\alpha}{2}, \frac{1}{4} \right) \). 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 \( N \), \( a_nb_n \geq M\alpha \) for all \( n \geq N \). Since \( \alpha > 0 \) is fixed and \( M \) is arbitrary it follows that \( \lim_{n \to \infty} (a_nb_n) = \infty \) as desired. 

Notation 1.3 Suppose that \( \{x_n\}_{n=1}^\infty \subset \mathbb{R} \) is a sequence of numbers. Then
\[
\liminf_{n \to \infty} x_n = \lim_{n \to \infty} \inf \{x_k : k \geq n\} \text{ and } \limsup_{n \to \infty} x_n = \lim_{n \to \infty} \sup \{x_k : k \geq n\}. \tag{1.3}
\]
We will also write \( \lim \) for \( \liminf_{n \to \infty} \) and \( \overline{\lim} \) for \( \limsup_{n \to \infty} \).

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

\[
\begin{align*}
\liminf_{n \to \infty} x_n &= \sup_{n} \inf\{x_k : k \geq n\} \\
\limsup_{n \to \infty} x_n &= \inf_{n} \sup\{x_k : k \geq n\}.
\end{align*}
\]

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

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

1. \( \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \) and \( \lim_{n \to \infty} a_n \) exists in \( \mathbb{R} \) iff
   \[
   \liminf_{n \to \infty} a_n = \limsup_{n \to \infty} a_n \in \mathbb{R}.
   \]

2. There is a subsequence \( \{a_{n_k}\}_{k=1}^{\infty} \) of \( \{a_n\}_{n=1}^{\infty} \) such that \( \lim_{n \to \infty} a_{n_k} = \liminf_{n \to \infty} a_n \). Similarly, there is a subsequence \( \{a_{n_k}\}_{k=1}^{\infty} \) of \( \{a_n\}_{n=1}^{\infty} \) such that \( \lim_{n \to \infty} a_{n_k} = \limsup_{n \to \infty} a_n \).

3. \[
\limsup_{n \to \infty} (a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n \tag{1.5}
\]
   whenever the right side of this equation is not of the form \( \infty - \infty \).

4. If \( a_n \geq 0 \) and \( b_n \geq 0 \) for all \( n \in \mathbb{N} \), then
   \[
   \limsup_{n \to \infty} (a_n b_n) \leq \limsup_{n \to \infty} a_n \cdot \limsup_{n \to \infty} b_n, \tag{1.6}
   \]
   provided the right hand side of (1.6) is not of the form \( 0 \cdot \infty \) or \( \infty \cdot 0 \).

**Proof.**

1. Since
   \[
   \inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\} \forall n,
   \]
   \[
   \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n.
   \]

Now suppose that \( \liminf_{n \to \infty} a_n = \limsup_{n \to \infty} a_n = a \in \mathbb{R} \). Then for all \( \varepsilon > 0 \), there is an integer \( N \) such that
   \[
a - \varepsilon \leq \liminf_{n \to \infty} a_n \leq \sup\{a_k : k \geq N\} \leq a + \varepsilon.
   \]
   i.e.
   \[
a - \varepsilon \leq a_k \leq a + \varepsilon \text{ for all } k \geq N.
   \]

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

and hence \( \lim_{n \to \infty} a_n = \infty \). The case where \( \limsup_{n \to \infty} a_n = -\infty \) is handled similarly.

Conversely, suppose that \( \lim_{n \to \infty} a_n = A \in \mathbb{R} \) exists. If \( A \in \mathbb{R} \), then for every \( \varepsilon > 0 \) there exists \( N(\varepsilon) \in \mathbb{N} \) such that \( |A - a_n| \leq \varepsilon \) for all \( n \geq N(\varepsilon) \), i.e.
   \[
   A - \varepsilon \leq a_n \leq A + \varepsilon \text{ for all } n \geq N(\varepsilon).
   \]

From this we learn that
   \[
   A - \varepsilon \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \leq A + \varepsilon.
   \]

Since \( \varepsilon > 0 \) is arbitrary, it follows that
   \[
   A \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \leq A,
   \]

i.e. that \( A = \lim_{n \to \infty} a_n = \limsup_{n \to \infty} a_n \). If \( A = \infty \), then for all \( M > 0 \) there exists \( N = N(M) \) such that \( a_n \geq M \) for all \( n \geq N \). This show that \( \liminf_{n \to \infty} a_n \geq M \) and since \( M \) is arbitrary it follows that
   \[
   \infty \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n.
   \]

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
   \[
   \limsup_{n \to \infty} (a_n + b_n) < \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n \text{ and }
   \]
   \[
   \liminf_{n \to \infty} (a_n + b_n) > \liminf_{n \to \infty} a_n + \liminf_{n \to \infty} b_n.
   \]
Remark 1.6. If $a_n \leq b_n$ for a.a. $n$, (i.e. there exists $N \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n \geq N$), then it is easy to verify that
\[
\limsup_{n \to \infty} a_n \leq \limsup_{n \to \infty} b_n \quad \text{and} \quad \liminf_{n \to \infty} a_n \leq \liminf_{n \to \infty} b_n.
\]
In particular if $a_n = b_n$ for a.a. $n$, then
\[
\limsup_{n \to \infty} a_n = \limsup_{n \to \infty} b_n \quad \text{and} \quad \liminf_{n \to \infty} a_n = \liminf_{n \to \infty} b_n.
\]

It is also easy to verify that if $b \in \mathbb{R}$, then
\[
\limsup_{n \to \infty} (a_n + b) = \limsup_{n \to \infty} a_n + b \quad \text{and} \quad \liminf_{n \to \infty} (a_n + b) = \liminf_{n \to \infty} a_n + b.
\]

Lemma 1.7. If \(\{a_n\}_{n=1}^{\infty}\) and \(\{b_n\}_{n=1}^{\infty}\) are two sequences of real numbers and \(b = \lim_{n \to \infty} b_n\) exists in \(\mathbb{R}\), then
\[
\liminf_{n \to \infty} (a_n + b_n) = \liminf_{n \to \infty} a_n + b.
\]

**Proof.** Let \(\varepsilon > 0\) be given and choose \(N = N(\varepsilon) \in \mathbb{N}\) so that \(|b - b_n| \leq \varepsilon\) for all \(n \geq N\). Then
\[
a_n + b - \varepsilon \leq a_n + b_n \leq a_n + b + \varepsilon
\]
and hence taking the limit inf of this inequality using Remark 1.6 we find
\[
\liminf_{n \to \infty} a_n + b - \varepsilon \leq \liminf_{n \to \infty} (a_n + b_n) \leq \liminf_{n \to \infty} a_n + b + \varepsilon.
\]
As this is valid for all \(\varepsilon > 0\) we may let \(\varepsilon \downarrow 0\) in the previous inequality to find
\[
\liminf_{n \to \infty} a_n + b \leq \liminf_{n \to \infty} (a_n + b_n) \leq \liminf_{n \to \infty} a_n + b.
\]
\[\blacksquare\]

1.1 Infinite sums

**Definition 1.8.** 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 1.9.** 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,
\[
\sum_{n=1}^{\infty} (a_n + \lambda b_n) = \lim_{N \to \infty} \sum_{n=1}^{N} (a_n + \lambda b_n) = \lim_{N \to \infty} \left[ \sum_{n=1}^{N} a_n + \lambda \sum_{n=1}^{N} b_n \right]
\]
\[
= \lim_{N \to \infty} \sum_{n=1}^{N} a_n + \lambda \lim_{N \to \infty} \sum_{n=1}^{N} b_n = \sum_{n=1}^{\infty} a_n + \lambda \cdot \sum_{n=1}^{\infty} b_n.
\]

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

**Proposition 1.10 (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
\[
\lim_{n \to \infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} f(i), \quad i.e.
\]
\[
\lim_{n \to \infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} \lim_{n \to \infty} f_n(i).
\]

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,
\[
\sum_{i=1}^{N} f(i) = \sum_{i=1}^{N} \lim_{n \to \infty} f_n(i) = \lim_{n \to \infty} \sum_{i=1}^{N} f_n(i) = \lim_{n \to \infty} \sum_{i=1}^{\infty} f_n(i) = M.
\]
Letting \(N \uparrow \infty\) in this equation then shows \(\sum_{i=1}^{\infty} f(i) \leq M\) which completes the proof. \(\blacksquare\)

**Proposition 1.11 (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 1.10),
\[
\lim_{N \to \infty} \sum_{k=1}^{\infty} S_N(k) = \sum_{k=1}^{\infty} \lim_{N \to \infty} S_N(k) = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn}.
\]
On the other hand,
\[
\sum_{k=1}^{\infty} S_N(k) = \sum_{k=1}^{N} \sum_{n=1}^{\infty} a_{kn} = \sum_{n=1}^{\infty} \sum_{k=1}^{N} a_{kn}
\]
so that
\[
\lim_{N \to \infty} \sum_{k=1}^{\infty} S_N(k) = \lim_{N \to \infty} \sum_{n=1}^{\infty} \sum_{k=1}^{N} a_{kn} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn}.
\]

Second Proof. Let
\[
M := \sup \left\{ \sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn} \mid K, N \in \mathbb{N} \right\} = \sup \left\{ \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn} \mid K, N \in \mathbb{N} \right\}
\]
and
\[
L := \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn}.
\]
Since
\[
L = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = \lim_{K \to \infty} \sum_{k=1}^{K} \sum_{n=1}^{\infty} a_{kn} = \lim_{N \to \infty} \lim_{K \to \infty} \sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn}
\]
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,
\[
\sum_{k=1}^{K} \sum_{n=1}^{N} a_{kn} \leq \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} \leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn} = L
\]
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 1.12.** A sequence \(\{a_n\}_{n=1}^{\infty} \subseteq \mathbb{R}\) is **summable** (absolutely convergent) if \(\sum_{n=1}^{\infty} |a_n| < \infty\). When \(\{a_n\}_{n=1}^{\infty} \subseteq \mathbb{R}\) 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 1.13.** From Eq. (1.7) it follows that
\[
\left| \sum_{n=1}^{\infty} a_n \right| \leq \sum_{n=1}^{\infty} a_{n}^{+} + \sum_{n=1}^{\infty} a_{n}^{-} = \sum_{n=1}^{\infty} (a_{n}^{+} + a_{n}^{-}) = \sum_{n=1}^{\infty} |a_n|.
\]

**Proposition 1.14 (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 \(|c_n| \leq |a_n| + |\lambda||b_n|\) and hence
\[
\sum_{n=1}^{\infty} |c_n| = \sum_{n=1}^{\infty} |a_n + \lambda b_n| \leq \sum_{n=1}^{\infty} |a_n| + |\lambda| \sum_{n=1}^{\infty} |b_n| < \infty.
\]
This shows \(\{c_n\}_{n=1}^{\infty}\) is summable. Let us now suppose that \(\lambda \geq 0\) 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 1.9 then shows,
\[
\sum_{n=1}^{\infty} c_n^{+} + \lambda \sum_{n=1}^{\infty} b_n^{-} = \sum_{n=1}^{\infty} (c_n^{+} + a_n^{-} + \lambda b_n^{-}) = \sum_{n=1}^{\infty} (c_n^{+} + a_n^{+} + \lambda b_n^{+}) = \sum_{n=1}^{\infty} c_n^{+} + \lambda \sum_{n=1}^{\infty} b_n^{+}.
\]
Rearranging these terms gives,
\[
\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} c_n^{+} - \sum_{n=1}^{\infty} c_n^{-} = \sum_{n=1}^{\infty} a_n^{+} + \lambda \sum_{n=1}^{\infty} b_n^{+} - \sum_{n=1}^{\infty} a_n^{-} - \lambda \sum_{n=1}^{\infty} b_n^{-} = \sum_{n=1}^{\infty} a_n + \lambda \sum_{n=1}^{\infty} b_n.
\]
To finish the proof we need only observe that \((-1 \cdot a_n)^{\pm} = a_n^{\mp}\) and hence
\[ \sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^- - \sum_{n=1}^{\infty} a_n^+ = - \left( \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^- \right) = - \sum_{n=1}^{\infty} a_n. \]

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

**Proposition 1.15 (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

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

(In this simple setting, this is in fact a special case of the statement that absolutely convergent sums are independent of rearrangements.]

**Example 1.16 (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 \( m \).

For example, take \( S_{mn} = 1 \) or \( S_{mn} = \frac{m}{m+n} \). Then define \( \{a_{ij}\}_{i,j=1}^{\infty} \) so that

\[ S_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \tag{1.8} \]

in which case we will have

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

To find \( a_{ij} \) which give Eq. \( \tag{1.8} \) if we define \( S_{mn} = 0 \) if \( m = 0 \) or \( n = 0 \), then

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

and so

\[ a_{mn} = \sum_{j=1}^{n} a_{mj} - \sum_{j=1}^{n-1} a_{mj} = (S_{mn} - S_{m-1,n}) - (S_{m,n-1} - S_{m-1,n-1}) = S_{mn} - S_{m-1,n} - S_{m,n-1} + S_{m-1,n-1}. \]

Applying this to the example where \( S_{mn} = 1 \) gives,

\[ a_{m,n} = 1_{m \geq n} - 1_{m-1 \geq n} - 1_{m \geq n-1} + 1_{m-1 \geq n-1} \]

\[ = 2 \cdot 1_{m \geq n} - 1_{m \geq n+1} - 1_{m \geq n-1} \]

\[ = 1_{m=n} - 1_{m=n-1}. \]

Thus it follows that \( n \to a_{m,n} \) is summable and \( m \to a_{m,n} \) is summable and

\[ \sum_{m=1}^{\infty} a_{m,n} = 1_{n=1} \text{ and } \sum_{n=1}^{\infty} a_{m,n} = 0 \]

and therefore

\[ \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{m,n} = 1 \neq 0 = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{m,n}. \]

Note: the graph of \( a_{m,n} \) is 0 everywhere except on the two lines, \( n = m \) where \( a_{m,m} = 1 \) and \( n = m+1 \) where \( a_{m,m+1} = -1 \). For an integral variant of this example, fatten up \( a_{m,n} \) to produce that function,

\[ h(x,y) = 1 \text{ for all } x \leq y \leq x+1 - 1 \]

see Figure 1.1. For this \( h \) we have

Fig. 1.1. The function \( h \) is 1 between the black and the green line, it is \(-1\) between the green and red line, at it 0 on the remaining regions in \([0, \infty)^2\).
and hence
\[ \int_0^\infty dx \int_0^\infty dy \, k(x, y) = 0 \neq 1 = \int_0^\infty dy \int_0^\infty dx \, k(x, y). \]

**Proposition 1.17 (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
\[ \sum_{i=1}^\infty \liminf_{n \to \infty} h_n(i) \leq \liminf_{n \to \infty} \sum_{i=1}^\infty h_n(i). \]

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

**Proposition 1.18 (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. \( f_n(i) \leq g(i) \) 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
\[ \lim_{n \to \infty} \sum_{i=1}^\infty f_n(i) = \sum_{i=1}^\infty f(i). \]

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

**Exercise 1.1 (Prove the Fubini Proposition 1.15).** 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 Tonelli’s theorem (Proposition 1.11) with \( a_{kn} \) replaced by \( a_{kn}^+ \) and \( a_{kn}^- \). You should be careful to verify that \( \{a_{kn}\}_{n=1}^\infty \) is summable for each \( k \) and that \( \{S_k = \sum_{n=1}^\infty a_{kn}\}_{k=1}^\infty \) is summable so that \( \sum_{k=1}^\infty \sum_{n=1}^\infty a_{kn} = \sum_{k=1}^\infty S_k \) exists, etc. etc.

**Exercise 1.2 (Prove Fatou’s Lemma in Proposition 1.17).** Suppose that for each \( n \in \mathbb{N} \), \( \{h_n(i)\}_{i=1}^\infty \) is any sequence in \([0, \infty]\), then
\[ \sum_{i=1}^\infty \liminf_{n \to \infty} h_n(i) \leq \liminf_{n \to \infty} \sum_{i=1}^\infty h_n(i). \]

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

**Exercise 1.3 (Prove DCT as in Proposition 1.18).** 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. \( f_n(i) \leq g(i) \) 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) $d(x, z) \leq d(x, y) + d(y, z)$ 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 $d(x, y) = |\tan^{-1}(y) - \tan^{-1}(x)|$.
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 continuous function $d(f, g) := \max_{x \in [0,1]} |f(x) - g(x)|$ 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 $(X, \|\|)$ (see Definition 2.3) is a metric space with $d(x, y) := \|x - y\|$. Thus the space $\ell^p(\mu)$ (as in Theorem 2.1) is a metric space for all $p \in [1, \infty]$.
5. Let $X$ denote the $C^1$ - periodic functions on $\mathbb{R}$. Then $d(f, g) := \max_{x \in \mathbb{R}} |f'(x) - g'(x)|$ is a pseudo-metric on $X$.

Throughout 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 $C_x(\delta) := \{ y \in X : d(x, y) \leq \delta \}$.

**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 $x_n \xrightarrow{d} x$ as $n \to \infty$.

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

$$d(x, y) \leq d(x, x_n) + d(x_n, y) \to 0$$

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 $F$. A set $V \subset X$ is **open** iff $V^c$ is closed. We will write $F \subset X$ to indicate $F$ is a closed subset of $X$ and $V \subset X$ 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 $F$ be a collection of closed subsets of $X$, show $\cap F := \cap_{F \subset X} F$ 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 $V \subset A$ 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
\[
|d_A(x) - d_A(y)| \leq d(x,y) \quad \forall x,y \in X
\]
and in particular if \(x_n \rightarrow x\) in \(X\) then \(d_A(x_n) \rightarrow d_A(x)\) as \(n \rightarrow \infty\). Moreover the set \(F_\varepsilon := \{x \in X : d_A(x) \geq \varepsilon\}\) is closed in \(X\).

**Proof.** Let \(a \in A\) and \(x,y \in X\), then
\[
d_A(x) \leq d(x,a) \leq d(x,y) + d(y,a).
\]
Take the infimum over \(a\) in the above equation shows that
\[
d_A(x) \leq d(x,y) + d_A(y) \quad \forall x,y \in X.
\]
Therefore, \(d_A(x) - d_A(y) \leq d(x,y)\) and by interchanging \(x\) and \(y\) we also have that \(d_A(y) - d_A(x) \leq d(x,y)\) which implies Eq. (2.1). If \(x_n \rightarrow x\) in \(X\), then by Eq. (2.1),
\[
|d_A(x) - d_A(x_n)| \leq d(x,x_n) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty
\]
so that \(\lim_{n \rightarrow \infty} d_A (x_n) = d_A (x)\). Now suppose that \(\{x_n\}_{n=1}^\infty \subset FC\) and \(x_n \rightarrow x\) in \(X\), then
\[
d_A(x) = \lim_{n \rightarrow \infty} d_A(x_n) \geq \varepsilon
\]
since \(d_A(x_n) \geq \varepsilon\) for all \(n\). This shows that \(x \in FC\) and hence \(FC\) is closed. \(\blacksquare\)

**Corollary 2.8.** The function \(d\) satisfies,
\[
|d(x,y) - d(x',y')| \leq d(y,y') + d(x,x').
\]
In particular \(d : X \times X \rightarrow [0,\infty)\) 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,
\[
|d(x,y) - d(x',y')| \leq |d(x,y) - d(x,y')| + |d(x,y') - d(x',y')| \\
\leq d(y,y') + d(x,x').
\]
\(\blacksquare\)

**Example 2.9.** Let \(x \in X\) and \(\delta > 0\), then \(C_x(\delta)\) and \(B_x(\delta)\) are closed subsets of \(X\). For example if \(\{y_n\}_{n=1}^\infty \subset C_x(\delta)\) and \(y_n \rightarrow y\) in \(X\), then \(d(y_n,x) \leq \delta\) for all \(n\) and using Corollary 2.8 it follows \(d(y,x) \leq \delta\), i.e. \(y \in C_x(\delta)\). A similar proof shows \(B_x(\delta)\) 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 \(F_\varepsilon := \{x \in X : d_A(x) \geq \varepsilon\} \subset X\) 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. \(d(x,y) \geq \varepsilon\) 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 \(\varepsilon' \leq \varepsilon\).

**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
\[
\bar{A} := \{x \in X : \exists \{x_n\} \subset A \Rightarrow x = \lim_{n \rightarrow \infty} x_n\}.
\]
That is to say \(\bar{A}\) contains all limit points of \(A\). We say \(A\) is dense in \(X\) if \(A = X\), 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
\[
\bar{A} = \cap \{F : A \subset F \subset X \text{ with } F \text{ closed}\}.
\]
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 \(x = \lim_{n \rightarrow \infty} x_n\), 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 \(A \subset X\) with the property that there exists \(\varepsilon > 0\) such that \(d(a,b) \geq \varepsilon\) for all \(a,b \in A\) with \(a \neq b\). 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 $V \subset_c X$ to indicate that $V \subset X$ and $V \in \tau$ and similarly $F \subset_c X$ 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.
   \[ \bar{A} := \cap \{ F : A \subset F \subset X \} . \]
   (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
   \[ \text{acc}(A) = \{ x \in X : V \cap [A \setminus \{x\}] \neq \emptyset \text{ for all } V \in \tau_x \} . \]

5. The **boundary** of $A$ is the set $\text{bd}(A) := \bar{A} \setminus A^o$.

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:

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

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

\[ \text{bd}(A) := \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = A \cap \overline{A^c}, \]

which is to say $\text{bd}(A)$ 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

\[ d(f(x),f(x')) < \varepsilon \text{ provided that } \rho(x,x') < \delta. \]  \hspace{1cm} (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 $\varepsilon > 0$, $B_{f(x)}(\varepsilon) \subset B_{f(x)}(\varepsilon)$ is a neighborhood of $f(x)$ and so $f^{-1}(B_{f(x)}(\varepsilon))$ is a neighborhood of $x$ which must contain $B_x(\delta)$ for some $\delta > 0$. Because $x_n \to x$, it follows that $x_n \in B_x(\delta) \subset f^{-1}(B_{f(x)}(\varepsilon))$ 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 $\rho(x_n,x) < \frac{\varepsilon}{3} \text{ yet } d(f(x_n),f(x)) \geq \varepsilon$. 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 \(f^{-1}(A^c) = (f^{-1}(A))^c\), 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 \(V \subset Y\) 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

\[f^{-1}(\tau_Y) = \{f^{-1}(V) : V \in \tau_Y\} \subset \tau_X\].

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

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
\[
f(x) = \frac{d_A(x)}{d_A(x) + d_B(x)} \text{ for } x \in X
\] (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 \(A \cap B = \emptyset\) and \(d_A(x) + d_B(x) > 0\) 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 \(F \subset X\) is open and \(d\) is a pseudo metric on \(X\) such that \(f = 1\) on \(F\) while \(f = 0\) on \(\overline{F}\). This of course follows from Lemma 2.22 by taking \(A = F\) and \(B = \overline{F}\).

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 \(f \in C(X, [0, 1])\) such that \(f = 1\) on \(A\) and \(f = 0\) on \(B\). Then set \(U = \{f > \frac{1}{2}\}\) 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 \(d(x, y) = \|x - y\|_2 = \left(\sum_{t=1}^n (x_t - y_t)^2\right)^{1/2}\).
3. \(X = C([a, b])\) 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 \((\overline{X}, \overline{d})\) and an isometric map \(i : X \to \overline{X}\) such that \(i(X)\) is dense in \(\overline{X}\), see Definition 2.11.

1. Let \(C\) denote the collection of Cauchy sequences \(a = \{a_n\}_{n=1}^\infty \subset X\). Given two element \(a, b \in C\) show \(d_C(a, b) := \lim_{n \to \infty} d(a_n, b_n)\) exists, \(d_C(a, b) \geq 0\) for all \(a, b \in C\) and \(d_C\) satisfies the triangle inequality,
\[
d_C(a, c) \leq d_C(a, b) + d_C(b, c) \quad \text{for all } a, b, c \in C.
\]

Thus \((C, d_C)\) would be a metric space if it were true that \(d_C(a, b) = 0\) iff \(a = b\). This however is false, for example if \(a_n = b_n\) for all \(n \geq 100\), then \(d_C(a, b) = 0\) while a need not equal \(b\).

2. Define two elements \(a, b \in C\) to be equivalent (write \(a \sim b\)) whenever \(d_C(a, b) = 0\). Show "\(\sim\)" is an equivalence relation on \(C\) and that \(d_C(a', b') = d_C(a, b)\) if \(a \sim a'\) and \(b \sim b'\). (Hint: see Corollary ??)
3. Given $a \in C$ let $\bar{a} := \{b \in C : b \sim a\}$ denote the equivalence class containing $a$ and let $X := \{\bar{a} : a \in C\}$ denote the collection of such equivalence classes. Show that $d(\bar{a}, \bar{b}) := d_C(a, b)$ is well defined on $X \times X$ and verify $(X, d)$ 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 $i : X \rightarrow X$ is an isometric map and that $i(X)$ is dense in $X$.

5. Verify $(X, d)$ is complete. Hint: if $\{\bar{a}(m)\}^\infty_{m=1}$ is a Cauchy sequence in $X$ choose $b_m \in X$ such that $d(i(b_m)), \bar{a}(m)) \leq 1/m$. Then show $\bar{a}(m) \rightarrow \bar{b}$ where $\bar{b} = \{b_m\}^\infty_{m=1}$.

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 $f : U \rightarrow C$ is Lip-K on $U$ if

$$|f(y) - f(x)| \leq Kd(x, y) \quad \text{for all } x, y \in U.$$ 

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

Lemma 2.28. Let $(X, d)$ be a pseudo-metric space, $U$ be a non-empty subset of $X$, $\bar{U}$ denote its closure in $X$, and $f : U \rightarrow C$ be a Lip-K function. Then there exists a unique Lip-K function, $\bar{f} : \bar{U} \rightarrow C$ such that $f = \bar{f}|_U$. [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 $U$.]

Example 2.29. Let $\varepsilon > 0$, $X = [\varepsilon, \infty)$, $d(x, y) := |y - x|$ for $x, y \in X$, $U := X \cap \mathbb{Q}$, and $f(\frac{m}{n}) = \frac{n}{m} = (\frac{m}{n})^{-1}$ for $m, n \in \mathbb{Q}$. For $x, y \in U$ we have

$$|f(x) - f(y)| = \frac{1}{|x|} \frac{|y - x|}{|y|} \leq \frac{1}{\varepsilon^2} \frac{|y - x|}{|y|}.$$

Thus $f$ is Lip-$\varepsilon^{-2}$ and therefore extends to a Lip-$\varepsilon^{-2}$ function on $\bar{U} = X$. This shows how to construct the inverse of a real number from knowing the how to compute inverses of numbers in $\mathbb{Q}$. Note that for $x \in X$, we will have $x^{-1} = \lim_{n \rightarrow \infty} x_n$ where $x_n \in U$ such that $x_n \rightarrow x$ as $n \rightarrow \infty$. Thus it follows that

$$x \cdot x^{-1} = \lim_{n \rightarrow \infty} x_n \cdot x^{-1} = \lim_{n \rightarrow \infty} [x_n \cdot x_n^{-1}] = 1.$$

2.3 Sequential compactness

Definition 2.30. A metric space, $(Y, \rho)$, is (sequentially) compact iff every sequence, $\{y_n\}^\infty_{n=1} \subset Y$ has a convergent subsequence.

The method of proof of the next proposition typically goes under the name of Cantor’s diagonalization technique.

Proposition 2.31 (A “little” Tychonov theorem). Suppose that $\Lambda$ is a countable set (like $\mathbb{N}$) and for each $\lambda \in \Lambda$, $(Y_\lambda, \rho_\lambda)$ is a compact metric space. Given sequences, $\{y_\lambda(n) : n \in \mathbb{N}\} \subset Y_\lambda$, for each $\lambda \in \Lambda$, there exists an increasing sequence, $\{n_\lambda\}^\infty_{k=1} \subset \mathbb{N}$ such that $y_\lambda := \lim_{k \rightarrow \infty} y_\lambda(n_\lambda)$ exists in $Y_\lambda$ for each $\lambda \in \Lambda$.

Proof. Since $\Lambda$ is countable we may as well assume that $\Lambda = \mathbb{N}$. By assumption and induction there exists infinite subsets $\{\Gamma_\lambda\}^\infty_{\lambda=0}$ of $\mathbb{N}$ such that $\Gamma_1 \supset \Gamma_2 \supset \Gamma_3 \supset \ldots$ and $\lim_{k \rightarrow \infty} y_\lambda(n) = y_\lambda \in Y_\lambda$ exists for all $\lambda \in \Lambda$. We now define $n_\lambda$ inductively by $n_1 = \min \{\Gamma_1\}$, $n_2 = \min \{\Gamma_2 \cap (n_1, \infty)\}$, $n_3 = \min \{\Gamma_3 \cap (n_2, \infty)\}$, etc. It then follows that $n_k \in \Gamma_\lambda$ for all $k \geq \lambda$, $n_k \geq k$ and $n_k \uparrow \infty$ as $k \rightarrow \infty$. Since subsequences of convergent sequences are still convergent we find,

$$\lim_{k \rightarrow \infty} y_\lambda(n_k) = \lim_{k \rightarrow \infty} y_\lambda(n) = y_\lambda \in Y_\lambda$$

for all $\lambda \in \Lambda$.

2.4 Supplementary Remarks

2.4.1 Word of Caution

Example 2.32. 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

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

with the usually Euclidean metric coming from the plane. Then

$$\overline{B_{(0,0)}(1)} = \{(0, y) \in \mathbb{R}^2 : |y| < 1\},$$

$$\overline{B_{(0,0)}(1)} = \{(0, y) \in \mathbb{R}^2 : |y| \leq 1\},$$

while

$$C_{(0,0)}(1) = B_{(0,0)}(1) \cup \{(1, 0)\}.$$
Lemma 2.33. Suppose that $(X,|.|)$ is a normed vector space and $d$ is the metric on $X$ defined by $d(x,y) = |x-y|$. Then
\[
\overline{B_x}(\varepsilon) = C_x(\varepsilon) \text{ and } \text{bd}(B_x(\varepsilon)) = \{y \in X : d(x,y) = \varepsilon\}.
\]
where the boundary operation, \text{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
\[
|v| = |y-x| = d(x,y) \leq \varepsilon.
\]
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 $y \in \bar{B}$. \hfill \blacksquare

2.4.2 Riemannian Metrics

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

Lemma 2.34. 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
\[
\overline{B_x}(\varepsilon) = C_x(\varepsilon) \text{ and } \text{bd}(B_x(\varepsilon)) = \{y \in X : d(x,y) = \varepsilon\}
\]
where the boundary operation, \text{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 $B^c \subset C^c$. Suppose that $y \in B^c$ 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 $d(x,y) \geq \varepsilon + \delta > \varepsilon$ and hence that $y \in C^c$. This will be accomplished by showing: if $d(x,y) < \varepsilon + \delta$ then $B_y(\delta) \cap B_x(\varepsilon) = \emptyset$. If $d(x,y) < \max(\varepsilon,\delta)$ then either $x \in B_y(\delta)$ or $y \in B_x(\varepsilon)$. In either case $B_y(\delta) \cap B_x(\varepsilon) = \emptyset$. Hence we may assume that $\max(\varepsilon,\delta) \leq d(x,y) < \varepsilon + \delta$. Let $\alpha > 0$ be a number such that
\[
\max(\varepsilon,\delta) \leq d(x,y) < \varepsilon + \delta
\]
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 $z \in B_x(\varepsilon) \cap B_x(\delta) \neq \emptyset$. \hfill \blacksquare

Remark 2.35. 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 $0 \leq u < v \leq 1$,
\[
d(x,y) + \varepsilon = \ell(\sigma) = \ell(\sigma|_{[0,u]}) + \ell(\sigma|_{[u,v]}) + \ell(\sigma|_{[v,1]})
\]
\[
\geq d(x,\sigma(u)) + \ell(\sigma|_{[u,v]}) + d(\sigma(v),y)
\]
and therefore, using the triangle inequality,
\[
\ell(\sigma|_{[u,v]}) \leq d(x,y) + \varepsilon - d(x,\sigma(u)) - d(\sigma(v),y)
\]
\[
\leq d(\sigma(u),\sigma(v)) + \varepsilon.
\]
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 $\sigma(v)$. 

![Fig. 2.2. An almost length minimizing curve joining $x$ to $y$.](image)
2.5 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 if 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
\[
d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}.
\]
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
\[
d(f(a), f(b)) < \varepsilon
d\]
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 \(\mathbb{R}\).
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,
   \[ \Omega = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_i \in \{0, 1\}\} = \{0, 1\}^\mathbb{N}. \]

3. If we have a roulette wheel with 38 entries, then we might take \( \Omega = \{00, 0, 1, 2, \ldots, 36\} \) for one spin,
   \( \Omega = \{00, 0, 1, 2, \ldots, 36\}^N \) for \( N \) spins, and
   \( \Omega = \{00, 0, 1, 2, \ldots, 36\}^\mathbb{N} \) 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
   \( \Omega = \mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\} = \{e^{i\theta} : -\pi \leq \theta \leq \pi\}. \)

5. If we throw darts at a board of radius \( R \), we may take
   \( \Omega = D_R := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\} \)
   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, \( 0 \leq t < \infty \). In this case, we might take,
   \( \Omega = \{\omega \in C([0, \infty), \mathbb{R}^3) : \omega(0) = x\}. \)

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 \( \Omega = \{0, 1\}^N \) 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 \( n \)th – toss, while \( \omega_n = 0 \) represents a tail on the \( n \)th – toss.

1. \( A = \{\omega \in \Omega : \omega_3 = 1\} \) represents the event that the third toss was a head.
2. \( A = \bigcup_{i=1}^\infty \{\omega \in \Omega : \omega_i = \omega_{i+1} = 1\} \) represents the event that (at least) two heads are tossed twice in a row at some time.
3. \( A = \bigcap_{N=1}^\infty \bigcup_{n \geq N} \{\omega \in \Omega : \omega_n = 1\} \) is the event where there are infinitely many heads tossed in the sequence.
4. \( A = \bigcup_{N=1}^\infty \bigcap_{n \geq N} \{\omega \in \Omega : \omega_n = 1\} \) 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 \( n \geq N \).

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
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

$$P_N(\bigcup_{j=1}^\infty A_j) = \sum_{j=1}^\infty P_N(A_j).$$

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, $1_{\bigcup_{j=1}^\infty A_j} = \sum_{j=1}^\infty 1_{A_j}$ and therefore

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

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.

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

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, $1_{\bigcup_{j=1}^\infty A_j} = \sum_{j=1}^\infty 1_{A_j}$ and therefore we might hope that (by another leap of faith) that

$$P\left(\bigcup_{j=1}^\infty A_j\right) = \lim_{N \to \infty} P_N\left(\bigcup_{j=1}^\infty A_j\right) = \lim_{N \to \infty} \sum_{j=1}^\infty P_N(A_j) = \sum_{j=1}^\infty \lim_{N \to \infty} P_N(A_j) = \sum_{j=1}^\infty P(A_j).$$

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
\[ P(\{\omega : (\omega_1, \ldots, \omega_k) = \sigma\}) = \frac{1}{2^k}. \]

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
\[
A_n = \{\omega \in \Omega : \omega_n = 1\} = \{\text{heads at time } n\}
\]
\[
B_N := \bigcup_{n \geq N} A_n = \{\text{at least one heads at time } N \text{ or later}\}
\]
and
\[
B = \cap_{N=1}^\infty B_N = \{A_n \text{ i.o.} \} = \cap_{N=1}^\infty \bigcup_{n \geq N} A_n.
\]

Since
\[
B_N^c = \cap_{n \geq N} A_n^c \subseteq \cap_{M \geq n \geq N} A_n^c = \{\omega \in \Omega : \omega_N = \omega_{N+1} = \cdots = \omega_M = 0\},
\]
we see that
\[ P(B_N^c) \leq \frac{1}{2^{M-N}} \to 0 \quad \text{as } M \to \infty. \]

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 6.3 that the continuity condition on \( P \) is equivalent to the \( \sigma \) – additivity of \( P \). For \( z \in \Omega \) and \( N \subset \Omega \) let
\[ zN := \{zn \in \Omega : n \in N\}, \quad (3.5) \]
that is to say \( e^{i\theta}N \) is the set \( N \) rotated counter clockwise by angle \( \theta \). By assumption, we are supposing that
\[ P(zN) = P(N) \quad (3.6) \]
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
\[ \Omega = \sum_{r \in R} (rN) \quad (3.7) \]
where \( \sum_n A_n \) is used to denote the union of pair-wise disjoint sets \( \{A_n\} \). By Eqs. (3.6) and (3.7),
\[ 1 = P(\Omega) = \sum_{r \in R} P(rN) = \sum_{r \in R} P(N). \]

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 \( P(\Omega) = 0 \) or \( P(\Omega) > 0 \).

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 \( \mathbb{R}^d \) by decomposing it into five pieces. To be more precise if \( B \) is the unit ball in \( \mathbb{R}^d \) 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 \( B' \cap B = \emptyset \) and 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, $B \subset 2^\Omega$. 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 ($A$) in $\Omega$, see Proposition 5.8.
2. Verify that $P_0$ has the continuity property of being countably additive – this is a substantial restriction on $P_0$, see Proposition 6.4.
3. Show that continuous $P_0$'s have a unique extension ($P$) to ($B$) – the “closure” of $A$, see Theorem 6.20 and 6.28.
4. Now try to compute as explicitly as possible $P(A)$ for $A \in B$ 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.]
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 \( f : X \rightarrow Y \). 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 \( X = \{1, 2, \ldots, N\} \), we will write \( Y^N \) in place of \( Y^{\{1, 2, \ldots, N\}} \) and denote \( f \in Y^N \) by \( f = (f_1, f_2, \ldots, f_N) \) 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 \( \pi_\alpha : X_A \rightarrow X_\alpha \) be the canonical projection map defined by \( \pi_\alpha(x) = x_\alpha \). 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 \( A \Delta B := (B \setminus A) \cup (A \setminus B) \).

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 \( \bigcup_{\alpha \in I} A_\alpha := \{x \in X : \exists \alpha \in I \exists x \in A_\alpha\} \) and \( \bigcap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \forall \alpha \in I\} \).

**Notation 4.4** We will also write \( \sum_{\alpha \in I} A_\alpha \) for \( \bigcup_{\alpha \in I} A_\alpha \) 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 \( \bigcup \) is closely related to \( \exists \) and \( \bigcap \) is closely related to \( \forall \). For example let \( \{A_n\}_{n=1}^\infty \) be a sequence of subsets from \( X \) and define \( \inf_{k \geq n} A_n := \bigcap_{k \geq n} A_k \), \( \sup_{k \geq n} A_n := \bigcup_{k \geq n} A_k \), \( \limsup_{n \rightarrow \infty} A_n := \bigcup_{n \geq K} \bigcap_{k \geq n} A_k \) and \( \liminf_{n \rightarrow \infty} A_n := \bigcap_{n \geq K} \bigcup_{k \geq n} A_k \) for all \( n \) sufficiently large.

(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 \( \forall N \in \mathbb{N} \exists n \geq N \exists x \in A_n \) and this may be expressed as \( \{A_n \text{ i.o.}\} = \bigcap_{N=1}^\infty \bigcup_{n \geq N} A_n \).

Similarly, \( x \in \{A_n \text{ a.a.}\} \) iff \( \exists N \in \mathbb{N} \exists n \geq N \exists x \in A_n \) which may be written as \( \{A_n \text{ a.a.}\} = \bigcup_{N=1}^\infty \bigcap_{n \geq N} A_n \).
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 functions. Let \( A \) and \( B \) be subsets of \( X \), then
\[
1_{A \cap B} = 1_A \cdot 1_B = \min(1_A, 1_B), \\
1_{A \cup B} = \max(1_A, 1_B), \\
1_{A^c} = 1 - 1_A, \text{ and } 1_{A \Delta B} = |1_A - 1_B|
\]

Lemma 4.7 (Properties of \( \lim \inf \) and \( \lim \sup \)). We have:
1. \( \bigcup_n A_n = \cap_n A_n^c \),
2. \( \{A_n \uparrow, \downarrow\} = \{A_n^c \text{ a.a.}\} \),
3. \( \lim \sup n \to \infty A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) = \infty\} \),
4. \( \lim \inf n \to \infty A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) < \infty\} \),
5. \( \sup_{k \geq n} 1_{A_k}(x) = 1_{\cup_{k \geq n} A_k} = 1_{\sup_{k \geq n} A_k} \),
6. \( \inf_{k \geq n} 1_{A_k}(x) = 1_{\cap_{k \geq n} A_k} = 1_{\inf_{k \geq n} A_k} \),
7. \( \lim \sup_{n \to \infty} A_n = \lim \sup_{n \to \infty} A_n \), and
8. \( \lim \inf_{n \to \infty} A_n = \lim \inf_{n \to \infty} 1_{A_n}. \)

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.) \( \blacksquare \)

Definition 4.8. A set \( X \) is said to be countable if it 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 \( A \subset \mathbb{N} \) is in one to one correspondence with \( \mathbb{N} \).
3. A non-empty set \( X \) is countable iff there exists a surjective function, \( g : \mathbb{N} \to X \).
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 \( f(1) = \min A \) and define \( f \) inductively by
\[
f(n+1) = \min \{A \setminus \{f(1), \ldots, f(n)\}\}.
\]
Since \( A \) is infinite the process continues indefinitely. The function \( f : \mathbb{N} \to A \) 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 \( f : \mathbb{N} \to \mathbb{N} \) is injective which combined with item 2. (taking \( A = f(X) \)) 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
\[
\begin{pmatrix}
(1,1) & (1,2) & (1,3) & \ldots \\
(2,1) & (2,2) & (2,3) & \ldots \\
(3,1) & (3,2) & (3,3) & \ldots \\
& & & \ddots
\end{pmatrix}
\]
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 \( f : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \) by \( f(m,n) := a_m(n) \). The function \( f \) is surjective and hence so is the composition, \( f \circ h : \mathbb{N} \to \mathbb{N} \) 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 \( f(n) = (f_1(n), f_2(n), f_3(n), \ldots) \). 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 \( a \notin f(N) \). 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, 
\[ i : 2^n \to 2^X \] 
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 
\[ f \circ i : 2^n \to \mathbb{N} \] 
is also injective which would imply \( 2^n \) is countable which it is not.

\[ \square \]

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 \( [\cup_{i \in I} A_i] \cap \cup_{i \in I} B_i \subset \cup_{i \in I} (A_i \cap B_i) \) and then use this inclusion twice to show \( [\cup_{i \in I} A_i] \cap \cup_{i \in I} B_i \subset \cup_{i \in I} (A_i \cap B_i) \).

Exercise 4.4 (Triangle inclusion for sets). If \( A, B, C \) are subsets of \( X \), show \( A \setminus B \subset [A \setminus B] \cup [B \setminus C] \) and use this identity twice to show
\[ A \cap C \subset [A \cap B] \cup [B \cap C]. \]  \hspace{1cm} (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
\[ f(C \cap D) \neq f(C) \cap f(D) \] and \( f(C^c) \neq [f(C)]^c \).

4.3 Algebraic sub-structures of sets

Definition 4.10. A collection of subsets \( A \) of a set \( X \) is a \( \pi - \text{system} \) or \( \text{multiplicative system} \) if \( A \) is closed under taking finite intersections.

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

1. \( \emptyset, X \in A \)
2. \( A \in A \) implies that \( A^c \in A \)
3. \( A \) is closed under finite unions, i.e. if \( A_1, \ldots, A_n \in A \) then \( A_1 \cup \cdots \cup A_n \in A \).

In view of conditions 1. and 2., 3. is equivalent to \( \forall A \in A \). \( A \) is closed under finite intersections.

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

Example 4.13. Here are some examples of algebras.

1. \( B = 2^X \), then \( B \) is a \( \sigma - \text{algebra} \).
2. \( B = \{\emptyset, X\} \) is a \( \sigma - \text{algebra} \) called the trivial \( \sigma - \text{field} \).
3. Let \( X = \{1, 2, 3\} \), then \( A = \{\emptyset, X, \{1\}, \{2, 3\}\} \) is an algebra while \( S := \{\emptyset, X, \{2, 3\}\} \) is not an algebra but is a \( \pi - \text{system} \).
4. Suppose that \( S \) is a set (called \text{state space}) and \( \Omega := S^\infty := S^N \) (think of \( N \) as time and \( \Omega \) as \text{path space}). For each \( n \in \mathbb{N} \), let
\[ A_n := \{B \times \Omega : B \subset S^n\} \].

Then \( A_n \) is a \( \sigma\)-algebra while \( A := \cup_{n=1}^\infty A_n \) is an algebra. (When \( S \) is an uncountable set we will typically only used a modified version of this construction.)

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

\textbf{Proof.} Simply take
\[ A(E) := \bigcap\{A : A \text{ is an algebra such that } E \subset A\} \]
and
\[ \sigma(E) := \bigcap\{M : M \text{ is a } \sigma - \text{algebra such that } E \subset M\}. \]

\[ \square \]

Example 4.15. Suppose \( X = \{1, 2, 3\} \) and \( E = \{\emptyset, X, \{1\}, \{2, 3\}\} \), see Figure 4.1. Then
\[ A(E) = \sigma(E) = 2^X. \]

On the other hand if \( E = \{\{1\}\} \), then \( A(E) = \{\emptyset, X, \{1\}, \{3\}\} \).
Exercise 4.8. Suppose that $\mathcal{E}_i \subset 2^X$ for $i = 1, 2$. Show that $A(\mathcal{E}_1) = A(\mathcal{E}_2)$ if $\mathcal{E}_1 \subset A(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset A(\mathcal{E}_1)$. 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 $A(\mathcal{E}_1) = A(\mathcal{E}_2)$ while $\mathcal{E}_1 \neq \mathcal{E}_2$.

In this course we will often be interested in the Borel $\sigma$ – field.

Definition 4.16 (Borel $\sigma$ – field). The Borel $\sigma$ – algebra, $\mathcal{B} = \mathcal{B}_R = \mathcal{B}(\mathbb{R})$, 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, $\mathcal{B}_R$, 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 $\mathcal{E} = \{A_1, \ldots, A_n\}$ where $A_1, \ldots, A_n$ is a partition of $X$. This case $A(\mathcal{E}) = \sigma(\mathcal{E}) = \{\bigcup_{i \in A} A_i : A \subset \{1, 2, \ldots, n\}\}$ where $\bigcup_{i \in A} A_i := \emptyset$ when $A = \emptyset$. Notice that $\#(A(\mathcal{E})) = \#(2^{1, 2, \ldots, n}) = 2^n$.

Example 4.19. Suppose that $X$ is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#(A) < \infty$. 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 orders 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 $x \in A_x \setminus A_y$ 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 $A = \{\bigcup_{i \in A} B_i : A \subset \{1, 2, \ldots, k\}\}$.

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

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 $\mathcal{F}$ partition $\mathcal{F}$ of $X$ such that $\mathcal{F} \subset \mathcal{B}$ and every element $B \in \mathcal{B}$ is of the form $B = \cup \{A \in \mathcal{F} : A \subset B\}$.

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 $A_x \cup A_y = \emptyset$ 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 $a \in \{0, 1\}^N \rightarrow A_a = \cup \{P_n : a_n = 1\} \in \mathcal{B}$ 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. ■
Corollary 4.23. A countable or infinite index set, and \( E = \{ A_i \}_{i \in I} \) be a partition of \( X \). Prove the algebra, \( \mathcal{A}(E) \), and that \( \sigma \) - algebra, \( \sigma(E) \), generated by \( E \) are given by

\[
\mathcal{A}(E) = \left\{ \bigcup_{i \in A} A_i : A \subset I \text{ with } \#(A) < \infty \text{ or } \#(A^c) < \infty \right\}
\] (4.3)

and

\[
\sigma(E) = \left\{ \bigcup_{i \in A} A_i : A \subset I \text{ with } A \text{ countable or } A^c \text{ countable} \right\}
\] (4.4)

respectively. Here we are using the convention that \( \bigcup_{i \in I} A_i := \emptyset \) when \( A = \emptyset \).

Proof. Let \( \mathcal{A} \) denote the right member of Eq. (4.3). From the definition of an algebra, it is clear that \( E \subset A \subset \mathcal{A}(E) \). Hence to show \( \mathcal{A} \) is an algebra. The proof of these assertions are routine except for possibly showing that \( A \) is closed under complementation. To check \( A \) is closed under complementation, let \( Z \in A \) 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

\[
Z^c = \bigcap_{i=1}^{N} \bigcup_{j=1}^{K} B_{ij} = \bigcup_{i=1}^{K} (B_{i1j_1} \cap B_{i2j_2} \cap \cdots \cap B_{iNj_n}) \in \mathcal{A}
\]

wherin we have used the fact that \( B_{i1j_1} \cap B_{i2j_2} \cap \cdots \cap B_{iNj_n} \) is a finite intersection of sets from \( \mathcal{E}_c \).

\[
\blacksquare
\]

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

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

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

\[
A_1 \cap A_2 = \left( \sum_{F \in A_1} F \right) \cap \left( \sum_{F \in A_1} F \right) = \bigcup_{(F_1, F_2) \in A_1 \times A_2} [F_1 \cap F_2]
\]

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

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

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

\[
S := \{(a, b] \cap \mathbb{R} : a, b \in \mathbb{R}\} = \{(a, b) : a \in [-\infty, \infty) \text{ and } a < b < \infty\} \cup \{\emptyset, \mathbb{R}\}
\]

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

\[
A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in A(S).
\]

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

\[
S := A \times B := \{A \times B : A \in A \text{ and } B \in B\}
\]

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 $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ with $E_i \in \mathcal{E}$ for $i = 1, 2, \ldots, n < \infty$.

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
   \[ \mu(E) \leq \sum_{i=1}^{n} \mu(E_i) \]
   whenever $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ with $E_i \in \mathcal{E}$ for $i = 1, 2, \ldots, n < \infty$.

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 $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$, then:

1. ($\mu$ is monotone) $\mu (A) \leq \mu (B)$ 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 subadditive) $\mu (\bigcup_{j=1}^{n} A_j) \leq \sum_{j=1}^{n} \mu(A_j)$.

4. $\mu$ is sub-additive on $\mathcal{A}$ iff
   \[ \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \bigcup_{i=1}^{\infty} A_i \tag{5.3} \]
   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 $A_i \in \mathcal{A}$, then
   \[ \mu \left( \sum_{i=1}^{\infty} A_i \right) \geq \sum_{i=1}^{\infty} \mu (A_i) \tag{5.4} \]
   (See Remark 5.10 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
   \[ \mu (B) = \mu (A) + \mu (B \setminus A) \geq \mu (A). \]

2. Since
   \[ A \cup B = [A \setminus (A \cap B)] \cup [B \setminus (A \cap B)] = \sum A \cap B, \]
   \[ \mu (A \cup B) = \mu (A \cup B \setminus (A \cap B)) + \mu (A \cap B) \]
   \[ = \mu (A \setminus (A \cap B)) + \mu (B \setminus (A \cap B)) + \mu (A \cap B). \]

Adding $\mu (A \cap B)$ to both sides of this equation proves Eq. (5.2).

3. Let $\{E_j\}_{i=1}^{n} \subset \mathcal{A}$ and set $\bar{E}_j = E_j \setminus \bigcup_{j=1}^{n} E_j$ so that the $\bar{E}_j$’s are pair-wise disjoint and $E = \bigcup_{j=1}^{n} E_j$. Since $\bar{E}_j \subset E_j$ it follows from the monotonicity of $\mu$ that
   \[ \mu (E) = \sum_{j=1}^{n} \mu (\bar{E}_j) \leq \sum_{j=1}^{n} \mu (E_j). \]

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 $A_i := B_i \setminus (B_1 \cup \cdots \cup B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of $\mu$ and Eq. (5.3)
   \[ \mu (A) \leq \sum_{i=1}^{\infty} \mu (A_i) \leq \sum_{i=1}^{\infty} \mu (B_i). \]
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 \( \mu \), \( \sum_{i=1}^{n} \mu(A_i) \leq \mu(A) \).

Letting \( n \to \infty \) in this equation shows \( \mu \) is superadditive.

6. This is a simple combination of the previous two items.

\[ \sum_{j=1}^{\infty} \mu_j(A) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mu_j(A_i) \]

\[ = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu_j(A_i) = \sum_{i=1}^{\infty} \mu(A_i). \]

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

\[ P(A) := \sum_{\omega \in A} p(\omega) \]

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.

\[ \mu(A) := \sum_{j=1}^{\infty} \mu_j(A) \]

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 1.11),

\[ \mu(A) = \sum_{j=1}^{\infty} \mu_j(A) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mu_j(A_i) \]

\[ = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu_j(A_i) = \sum_{i=1}^{\infty} \mu(A_i). \]

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,

\[ \mu(A) = \sum_{\omega \in A} \lambda(\omega) := \sup_{A \subset \Omega} \sum_{\omega \in A} \lambda(\omega) \forall A \subset \Omega. \] (5.5)

Here we are using the notation, \( A \subset \subset A \) to indicate that \( A \) is a finite subset of \( A \).

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

\[ A_N := \{ \omega_n : \omega_n \in A \text{ and } 1 \leq n \leq N \}. \]

Then by definition of \( \mu \),

\[ \mu(A) = \sum_{n=1}^{\infty} \lambda(\omega_n) \delta_{\omega_n}(A) = \lim_{N \to \infty} \sum_{n=1}^{N} \lambda(\omega_n) 1_{\omega_n \in A} \]

\[ = \lim_{N \to \infty} \sum_{\omega \in A_N} \lambda(\omega) \leq M. \]

On the other hand if \( A \subset \subset A \), then

\[ \sum_{\omega \in A} \lambda(\omega) = \sum_{n: \omega_n \in A} \lambda(\omega_n) = \mu(A) \leq \mu(\mathcal{B}) \]

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

5.2 Additive probabilities on infinite products of a discrete space

Let \( S \) be a non-empty 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)

\[ A_n := \{ B \times \Omega : B \subset S^n \} \quad \text{for all } n \in \mathbb{N}, \]

and \( \mathcal{A} := \bigcup_{n=1}^{\infty} A_n \). We call the elements, \( A \in \mathcal{A} \), the cylinder subsets of \( \Omega \). Notice that \( A \subset \Omega \) is a cylinder set if and only if there exists \( n \in \mathbb{N} \) and \( B \subset S^n \) such that
\[ A = B \times \Omega := \{ \omega \in \Omega : (\omega_1, \ldots, \omega_n) \in B \} . \]

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

Let us now further suppose that \( P : A \to [0, 1] \) is a finitely additive probability measure such that \( P|_{A_n} \) is \( \sigma \)-additive\(^1\) for each \( n \in \mathbb{N} \) and we further define \( s = (s_1, \ldots, s_n) \in S^n \),

\[
Q_n (B) := P (B \times \Omega) \quad \text{for all} \quad B \subset S^n \quad \text{and} \quad p_n (s) := Q_n (\{s\})
\]

Then \( Q_n \) is easily seen to be countably additive (as you should check) and satisfies,

\[
Q_n (B) = \sum_{s \in B} p_n (s) \quad \text{for all} \quad B \subset S^n .
\]

**Exercise 5.2** (Consistency Conditions). [Look at only, do not hand in.]

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 \( (s_1, \ldots, s_n) \in S^n \),

\[
p_n (s_1, \ldots, s_n) = \sum_{s \in S} p_{n+1} (s_1, \ldots, s_n, s) .
\]

These conditions are basically equivalent to the statements that \( Q_1 (S) = 1 \) and \( Q_{n+1} (B \times S) = Q_n (B) \) for all \( n \in \mathbb{N} \) and \( B \subset S^n \).

**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 \( A \) such that Eq. (5.6) holds for all \( n \in \mathbb{N} \) and \( (s_1, \ldots, s_n) \in S^n \) and such that \( P|_{A_n} \) is a \( \sigma \)-additive measure on \( A_n \) for all \( n \in \mathbb{N} \).

1 We will show a little later that any such \( P \) has a unique extension to a \( \sigma \)-additive measure on \( \sigma (A) \).

**Corollary 5.7.** 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 \( A \) such that \( P|_{A_n} \) is \( \sigma \)-additive for all \( n \in \mathbb{N} \) and

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

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

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

\[
\sum_{s \in S} p_n (s) = \sum_{s \in S} q_1 (s) = 1
\]

and

\[
\sum_{\lambda \in S} p_{n+1} (s_1, \ldots, s_n, \lambda) = \sum_{\lambda \in S} q_1 (s_1) q_2 (s_2) \cdots q_n (s_n) q_{n+1} (\lambda) = q_1 (s_1) q_2 (s_2) \cdots q_n (s_n) \sum_{\lambda \in S} q_{n+1} (\lambda) = q_1 (s_1) q_2 (s_2) \cdots q_n (s_n) = p_n (s_1, \ldots, s_n) .
\]

Hence the result follows from Exercise 5.3.

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 \( 2^\Omega \).

### 5.2 Finitely additive measures on \( \mathbb{R} \)

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

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

\[
\mu (A) = \sum_i \mu (E_i) .
\]
To prove existence, the main point is to show that $\mu(A)$ in Eq. (5.7) is well defined; i.e. if we also have $A = \sum_{j} F_j$ with $F_j \in \mathcal{S}$, then we must show

$$
\sum_{i} \mu(E_i) = \sum_{j} \mu(F_j).
$$

(5.8)

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

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

Similarly,

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

which combined with the previous equation shows that Eq. (5.8) holds. It is now easy to verify that $\mu$ extended to $\mathcal{A}$ as in Eq. (5.7) is an additive measure on $\mathcal{A}$.

Proposition 5.9. Let $\Omega = \mathbb{R}$, $\mathcal{S}$ be the semi-algebra,

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

(5.9)

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 $F : \mathbb{R} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and

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

(5.10)

Conversely, given an increasing function $F : \mathbb{R} \to [0, 1]$ 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. (5.10) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a \to -\infty} F(a)$ and $F(\infty) = \lim_{a \to \infty} F(b)$.)

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

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

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 $F : \mathbb{R} \to [0, 1]$ as in the statement of the theorem is given. Define $\mu$ on $\mathcal{S}$ using the formula in Eq. (5.10). The argument will be completed by showing $\mu$ is additive on $\mathcal{S}$ and hence, by Proposition 5.8 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,

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

Remark 5.10. Suppose that $F : \mathbb{R} \to \mathbb{R}$ is any non-decreasing function such that $F(\mathbb{R}) \subset \mathbb{R}$. Then the same methods used in the proof of Proposition 5.9 shows that there exists a unique finitely additive measure, $\mu = \mu_F$, on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ such that Eq. (5.10) holds. If $F(\infty) > \lim_{b \to \infty} F(b)$ and $A_i = (i, i+1]$ for $i \in \mathbb{N}$, then

$$
\sum_{i=1}^{\infty} \mu_F(A_i) = \sum_{i=1}^{\infty} (F(i+1) - F(i)) = \lim_{N \to \infty} \sum_{i=1}^{N} \left(F(i+1) - F(i)\right)
$$

$$
= \lim_{N \to \infty} \left(F(N+1) - F(1)\right) < F(\infty) - F(1) = \mu_F(\cup_{i=1}^{\infty} A_i).
$$

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 $F(-\infty) < \lim_{a \to -\infty} F(a)$ or if $F$ is not right continuous at some point $a \in \mathbb{R}$. Indeed, in the latter case consider

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

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 $F(\pm \infty) = \lim_{x \to \pm \infty} F(x)$. 
5.3 *Finitely additive measures on product spaces

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

**Definition 5.11.** Let

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

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

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

$$C = \sum_{i=1}^{n} A_i \times B_i \text{ with } (A_i, B_i) \in A \times B. \quad (5.11)$$

**Definition 5.12.** A function, $\rho : A \times B \to \mathbb{C}$ is bi-additive if for each $A \in A$, the function

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

is an additive measure on $B$ and for each $B \in B$, the function

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

is an additive measure on $A$.

**Theorem 5.13.** If $\rho : A \times B \to \mathbb{C}$ is a bi-additive function, then $\rho$ extends uniquely to an additive measure on the product algebra $A \odot B$. [This theorem has an obvious generalization to multiple factors.]

**Proof.** The collection $\mathcal{E} = A \times B$ 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 $A' \subset 2^A$ and $B' \subset 2^B$ generated by $\{A_k\}_{k=1}^{n}$ and $\{B_k\}_{k=1}^{n}$ respectively. Let $\Pi_a \subset A'$ and $\Pi_b \subset B'$ be partitions of $A$ and $B$ which generate $A'$ and $B'$ respectively as described in Proposition 4.20. Then

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

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

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

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

$$\rho(A_k \times B_k) = \rho\left(\sum_{\beta \subset B_k} \sum_{\alpha \subset A_k} \alpha \times \beta\right) = \sum_{\beta \subset B_k} \rho\left(\sum_{\alpha \subset A_k} \alpha \times \beta\right) = \sum_{\alpha \times \beta \subset A_k \times B_k} \rho(\alpha \times \beta). \quad (5.12)$$

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

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

**Example 5.14** (Product Measure). If $A \subset 2^X$ and $B \subset 2^Y$ be are subalgebras and $\mu : A \to [0, \infty]$ and $\nu : B \to [0, \infty]$ are finitely additive measures then there exists a unique finitely additive measure, $\mu \odot \nu : A \odot B \to [0, \infty]$ such that $\mu \odot \nu (A \times B) = \mu(A) \cdot \nu(B)$ for all $A \in A$ and $B \in B$. We refer to $\mu \odot \nu$ as a **product measure**. To verify this assertion one needs only apply Theorem 5.13 with $\rho(A \times B) := \mu(A) \cdot \nu(B)$.

Here is another interesting application of Theorem 5.13.
Proposition 5.15. Suppose that \( \mathcal{A} \subset 2^\Omega \) is an algebra and for each \( t \in \mathbb{R} \), \( \mu_t : \mathcal{A} \rightarrow \mathbb{C} \) is a finitely additive measure. Let \( Y = \{a, b\} \subset \mathbb{R} \) 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 \( \mathcal{E} \) such that
\[
\mu(A \times (a, b)) = \mu(A) - \mu_a(A) \quad \forall (a, b) \in \mathcal{E} \quad \text{and} \quad A \in \mathcal{A}.
\]

Proof. By Proposition 5.9 and Remark 5.10 for each \( A \in \mathcal{A} \), the function \( (a, b) \rightarrow \mu(A \times (a, b)) \) 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 \( A \rightarrow \mu(A \times B) \) is still finitely additive. The proof is complete with an application of Theorem 5.13.

5.4 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, \( f : \Omega \rightarrow Y \) 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 \( f : \Omega \rightarrow Y \) is measurable if \( \{f = y\} := f^{-1}(\{y\}) \in \mathcal{A} \) for all \( y \in Y \). A measurable simple function, \( f : \Omega \rightarrow \mathbb{C} \), is called a simple random variable relative to \( \mathcal{A} \).

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

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

1. \( \mathcal{S}(\mathcal{A}) \) is a sub-algebra of all functions from \( \Omega \) to \( \mathbb{C} \).
2. \( f : \Omega \rightarrow \mathbb{C} \), is a \( \mathcal{A} \)–simple random variable iff there exists \( \alpha_i \in \mathbb{C} \) and \( A_i \in \mathcal{A} \) for \( 1 \leq i \leq n \) for some \( n \in \mathbb{N} \) such that
\[
f = \sum_{i=1}^n \alpha_i 1_{A_i},
\]
3. For any function, \( F : \mathcal{C} \rightarrow \mathbb{C} \), \( F \circ f \in \mathcal{S}(\mathcal{A}) \) for all \( f \in \mathcal{S}(\mathcal{A}) \). In particular, \( |f| \in \mathcal{S}(\mathcal{A}) \) if \( f \in \mathcal{S}(\mathcal{A}) \).

Proof. 1. Let us observe that \( 1_\Omega = 1 \) and \( 1_\emptyset = 0 \) are in \( \mathcal{S}(\mathcal{A}) \). If \( f, g \in \mathcal{S}(\mathcal{A}) \) and \( c \in \mathbb{C} \setminus \{0\} \), then
\[
\{f + cg = \lambda\} = \bigcup_{a,b \in \mathbb{C}, a+cb=\lambda} \{(f = a) \cap (g = b)\} \in \mathcal{A}
\]
and
\[
\{f \cdot g = \lambda\} = \bigcup_{a,b \in \mathbb{C}, a\cdot cb=\lambda} \{(f = a) \cap (g = b)\} \in \mathcal{A}
\]
from which it follows that \( f + cg \) and \( f \cdot g \) are back in \( \mathcal{S}(\mathcal{A}) \).
2. Since \( \mathcal{S}(\mathcal{A}) \) is an algebra, every \( f \) of the form in Eq. (5.13) is in \( \mathcal{S}(\mathcal{A}) \). Conversely if \( f \in \mathcal{S}(\mathcal{A}) \) it follows by definition that \( f = \sum_\alpha f(\omega) \alpha 1_{f(\omega)} \) which is of the form in Eq. (5.13).
3. If \( F : \mathcal{C} \rightarrow \mathbb{C} \), then
\[
F \circ f = \sum_{\alpha \in F(\Omega)} F(\alpha) \cdot 1_{f(\omega)} \in \mathcal{S}(\mathcal{A}).
\]

Exercise 5.4 (\( \mathcal{A} \)–measurable simple functions). As in Example 4.19 let \( \mathcal{A} \subset 2^\Omega \) be a finite algebra and \( \{B_1, \ldots, B_k\} \) be the partition of \( \Omega \) associated to \( \mathcal{A} \). Show that a function, \( f : \Omega \rightarrow \mathbb{C} \), is a \( \mathcal{A} \)–simple function iff \( f \) is constant on \( B_i \) for each \( i \). Thus any \( \mathcal{A} \)–simple function is of the form,
\[
f = \sum_{i=1}^k \alpha_i 1_{B_i},
\]
for some \( \alpha_i \in \mathbb{C} \).

Corollary 5.19. Suppose that \( \Lambda \) is a finite set and \( Z : \Omega \rightarrow \Lambda \) is a function. Let
\[
\Lambda := \mathcal{A}(Z) := Z^{-1}(2^\Lambda) := \{Z^{-1}(E) : E \subset \Lambda\}.
\]
Then \( \Lambda \) is an algebra and \( f : \Omega \rightarrow \mathbb{C} \) is an \( \mathcal{A} \)–simple function iff \( f = F \circ Z \) for some function \( F : \Lambda \rightarrow \mathbb{C} \).

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 \( \Lambda \). 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 \( A_\lambda \), \( F : \Lambda \rightarrow \mathbb{C} \) is a function such that \( f = F \circ Z \).

Conversely if \( F : \Lambda \rightarrow \mathbb{C} \) is a function and \( f = F \circ Z \), then \( f = F(\lambda) \) on \( A_\lambda \), i.e. \( f \) is an \( \mathcal{A} \)–simple function.
5.4.1 The algebraic structure of simple functions*

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

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

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

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

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

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

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

Exercise 5.5. Continuing the notation introduced above:

1. Show $\mathcal{A}(\mathcal{S})$ is an algebra of sets.
2. Show $\mathcal{S}(\mathcal{A})$ is a simple function algebra.
3. Show that the map

$$\mathcal{A} \in \{ \text{Algebras} \subset 2^\Omega \} \rightarrow \mathcal{S}(\mathcal{A}) \in \{ \text{simple function algebras on} \ \Omega \}$$

is bijective and the map, $\mathcal{S} \rightarrow \mathcal{A}(\mathcal{S})$, is the inverse map.

5.5 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 $f \in \mathcal{S}(\mathcal{A})$ the integral or expectation, $\mathbb{E}(f) = \mathbb{E}_P(f)$, is defined by

$$\mathbb{E}_P(f) = \int_{\Omega} f \, dP = \sum_{y \in \mathbb{C}} yP(f = y).$$  

(5.17)

Example 5.23. Suppose that $A \in \mathcal{A}$, then

$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A) .$$  

Remark 5.24. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (5.1) by

$$P(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} 1_A(\omega(k))$$

where $\omega(k) \in \Omega$ was the result of the $k$th “independent” experiment. If we use this interpretation back in Eq. (5.17) we arrive at,

$$\mathbb{E}(f) = \sum_{y \in \mathbb{C}} yP(f = y) = \sum_{y \in \mathbb{C}} y \cdot \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} 1_{f(\omega(k)) = y}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{y \in \mathbb{C}} \sum_{k=1}^{N} 1_{f(\omega(k)) = y}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(\omega(k)) \cdot 1_{f(\omega(k)) = y}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(\omega(k)).$$

Thus informally, $\mathbb{E}f$ should represent the limiting average of the values of $f$ over many “independent” trials. We will later revisit this idea when we discuss 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 $\mathcal{S}_+(\mathcal{A})$ denote the $[0, \infty]$-valued $\mathcal{A}$-simple functions and for $f \in \mathcal{S}_+(\mathcal{A})$ we let

$$\mathbb{E}_\mu(f) = \int_{\Omega} f \, d\mu = \sum_{y \in [0, \infty]} y\mu(f = y).$$  

(5.19)

[For $f \in \mathcal{S}_+(\mathcal{A})$, $\mathbb{E}_\mu(f) = \infty$ is allowed and we use the convention that $0 \cdot \infty = 0 = \infty \cdot 0$.] Further let

$$\mathcal{S}_\mu(\mathcal{A}) := \{ f \in \mathcal{S}(\mathcal{A}) : \mu(f \neq 0) < \infty \}$$

and for $f \in \mathcal{S}_\mu(\mathcal{A})$ 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 \( S_\mu(A) \) is a subspace of \( S(A) \) and that \( S_+(A) \) is closed under addition and scalar multiplication by non-negative constants.

**Proposition 5.26.** The expectation operator, \( E = E_\mu : S_+(A) \to [0, \infty] \), satisfies:

1. If \( f \in S_+(A) \) and \( \lambda \in [0, \infty] \), then
   \[
   E(\lambda f) = \lambda E(f). \tag{5.20}
   \]
2. If \( f, g \in S_+(A) \), then \( f + g \in S_+(A) \) and
   \[
   E(f + g) = E(g) + E(f). \tag{5.21}
   \]
3. If \( f, g \in S_+(A) \) and \( f \leq g \), then \( E(f) \leq E(g) \).

**Proof.**

1. If \( \lambda = 0 \) or \( \lambda = \infty \) we have \( E(0 \cdot f) = 0 = 0 \cdot E(f) \) and \( E(\infty \cdot f) = \infty \) iff \( \mu(f) > 0 \) \( \iff \) \( E(f) > 0 \) \( \iff \) \( \infty \cdot E(f) = \infty \) respectively. If \( 0 < \lambda < \infty \) and \( f \in S_+(A) \), then
   \[
   E(\lambda f) = \sum_{y \in [0, \infty]} y \mu(\lambda f = y) = \sum_{y \in [0, \infty]} y \mu(f = y/\lambda) = \sum_{z \in [0, \infty]} \lambda z \mu(f = z) = \lambda E(f). \tag{5.22}
   \]
2. Writing \( \{f = a, g = b\} \) for \( f^{-1}(\{a\}) \cap g^{-1}(\{b\}) \), we have
   \[
   E(f + g) = \sum_{z \in [0, \infty]} z \mu(f + g = z) = \sum_{z \in [0, \infty]} z \mu \left( \sum_{a+b=z} \{f = a, g = b\} \right) = \sum_{a+b=\infty} z \mu(\{f = a, g = b\}) = \sum_{a+b=\infty} (a+b) \mu(\{f = a, g = b\}) = \sum_{a,b} (a+b) \mu(\{f = a, g = b\}).
   \]
But
   \[
   \sum_{a,b} a \mu(\{f = a, g = b\}) = \sum_{a,b} a \sum_{b} \mu(\{f = a, g = b\}) = \sum_{a} a \mu(\{f = a, g = b\}) = \sum_{a} a \mu(\{f = a\}) = E(f)
   \]
and similarly,
   \[
   \sum_{a,b} b \mu(\{f = a, g = b\}) = E(g).
   \]
   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 \( h \in S_+(A) \). As (is easily verified) \( g = f + h \) it follows that
   \[
   E_\mu g = E_\mu f + E_\mu h \geq E_\mu f.
   \]

**Alternative proof.** If \( \lambda \geq 0 \), then
   \[
   \lambda \cdot \mu(g = \lambda) = \sum_{0 \leq y < \lambda} y \mu(g = \lambda, f = y) \geq \sum_{0 \leq y < \lambda} y \mu(g = \lambda, f = y) = \sum_{0 \leq y } y \mu(g = \lambda, f = y).
   \]
Summing this inequality on \( \lambda \geq 0 \) then gives,
   \[
   E_\mu g \geq \sum_{0 \leq y } y \mu(g = \lambda, f = y) = \sum_{0 \leq y } y \mu(g = \lambda, f = y) = \sum_{0 \leq y } y \mu(f = y) = E_\mu f.
   \]

**Proposition 5.27.** The expectation operator, \( E = E_\mu : S_\mu(A) \to \mathbb{C} \), satisfies:

1. If \( f \in S_\mu(A) \) and \( \lambda \in \mathbb{C} \), then
   \[
   E(\lambda f) = \lambda E(f). \tag{5.22}
   \]
2. If \( f, g \in \mathbb{S}_\mu (A) \), then
\[
\mathbb{E}(f + g) = \mathbb{E}(g) + \mathbb{E}(f).
\] (5.23)

Items 1. and 2. say that \( \mathbb{E}(\cdot) \) is a linear functional on \( \mathbb{S}_\mu (A) \).

3. If \( f = \sum_{j=1}^{N} \lambda_j 1_{A_j} \) for some \( \lambda_j \in \mathbb{C} \) and some \( A_j \in A \) 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 \( 0 \leq f \in \mathbb{S}_\mu (A) \). More generally, if \( f, g \in \mathbb{S}_\mu (A) \) and \( f \leq g \), then \( \mathbb{E}(f) \leq \mathbb{E}(g) \).

5. For all \( f \in \mathbb{S}_\mu (A) \),
\[
|\mathbb{E}f| \leq \mathbb{E}|f|. \tag{5.25}
\]

**Proof.**

1. If \( \lambda \neq 0 \), then
\[
\mathbb{E}(\lambda f) = \sum_{y \in \mathbb{C}} y \mu(\lambda f = y) = \sum_{y \in \mathbb{C}} y \mu(f = y/\lambda) = \sum_{z \in \mathbb{C}} \lambda z \mu(f = z) = \lambda \mathbb{E}(f).
\]

The case \( \lambda = 0 \) is trivial.

2. Writing \( \{ f = a, g = b \} \) for \( f^{-1}(\{a\}) \cap g^{-1}(\{b\}) \), then
\[
\mathbb{E}(f + g) = \sum_{z \in \mathbb{C}} z \mu(f + g = z) = \sum_{z \in \mathbb{C}} \mu \left( \sum_{a+b=z} \{ f = a, g = b \} \right) = \sum_{z \in \mathbb{C}} \sum_{a+b=z} \mu(\{ f = a, g = b \}) = \sum_{a,b} (a+b) \mu(\{ f = a, g = b \}) = \sum_{a,b} (a+b) \mu(\{ f = a, g = b \}).
\]

But
\[
\sum_{a,b} a \mu(\{ f = a, g = b \}) = \sum_{a} a \sum_{b} \mu(\{ f = a, g = b \}) = \sum_{a} a \mu(\cup_{b} \{ f = a, g = b \}) = \sum_{a} a \mu(\{ f = a \}) = \mathbb{E}f
\]

and similarly,
\[
\sum_{a,b} b \mu(\{ f = a, g = b \}) = \mathbb{E}g.
\]

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

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
\[
\mathbb{E}(g) - \mathbb{E}(f) = \mathbb{E}(g - f) \geq 0.
\]

5. By the triangle inequality,
\[
|\mathbb{E}f| = \left| \sum_{\lambda \in \mathbb{C}} \lambda \mu(f = \lambda) \right| \leq \sum_{\lambda \in \mathbb{C}} |\lambda| \mu(f = \lambda) = \mathbb{E}|f|,
\]

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

\[ \blacksquare \]

**Remark 5.28.** If \( \Omega \) is a finite set and \( A = 2^{\Omega} \), then
\[
f(\cdot) = \sum_{\omega \in \Omega} f(\omega) 1_{\{\omega\}}
\]

and hence
\[
\mathbb{E}_P f = \sum_{\omega \in \Omega} f(\omega) P(\{\omega\}).
\]
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 $ho(A, B) := P(A \triangle B)$ where $A \triangle B = (A \setminus B) \cup (B \setminus A).$ Show:

1. $\rho(A, B) = \mathbb{E}|1_A - 1_B|$ and then use this (or not) to show
2. $\rho(A, C) \leq \rho(A, B) + \rho(B, C)$ 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 $\rho(A, B) = 0$ does not imply that $A = B$ but only that $A = B$ modulo a set of probability zero.

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

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

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

the estimates in Eq. (5.26) follow from item 4. of Proposition 5.27. \hspace{1cm} ■

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

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

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

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

$$1 - 1_A = 1 - A = \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M (1 - 1_{A_n})$$

$$= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} 1_{A_{n_1} \cdots A_{n_k}}$$

$$= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}$$

from which it follows that

$$1_{\bigcup_{n=1}^M A_n} 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}.$$ \hspace{1cm} (5.28)

Integrating this identity with respect to $\mu$ gives Eq. (5.27). \hspace{1cm} ■

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

$$\mu\left(\bigcup_{n=1}^M A_n\right) + \sum_{k=2}^M \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}) = \sum_{k=1}^M \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}).$$ \hspace{1cm} (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 $\mu\left(\bigcup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\bigcup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (5.29) are infinite.

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

$$1 - 1_A = (1 - 1_{A_1})(1 - 1_{A_2})(1 - 1_{A_3}) = 1 - (1_{A_1} + 1_{A_2} + 1_{A_3})$$

$$+ 1_{A_1} 1_{A_2} + 1_{A_1} 1_{A_3} + 1_{A_2} 1_{A_3} - 1_{A_1} 1_{A_2} 1_{A_3}$$

so that

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

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 $0 \leq m \leq M.$ (When $m = 0,$ both sides of Eq. (5.28) are zero and so we will only consider the case where $1 \leq m \leq M.$) With this notation we have
\[
\sum_{k=1}^{M} (-1)^{k+1} \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega)
\]
\[
= \sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \leq n_1 < n_2 < \cdots < n_k \leq m} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega)
\]
\[
= \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k}
\]
\[
= 1 - \sum_{k=0}^{m} (-1)^{k} \binom{m}{k}
\]
\[
= 1 - (1 - 1)^{m} = 1.
\]

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), \(\omega : \{1, 2, \ldots, n\} \rightarrow \{1, 2, \ldots, n\}\), 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

\[
P(B) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}).
\]

Since

\[
P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(\{\omega \in \Omega : \omega(i_1) = i_1, \ldots, \omega(i_k) = i_k\}) = \frac{(n-k)!}{n!}
\]

and

\[
\# \{1 \leq i_1 < i_2 < \cdots < i_k \leq n\} = \binom{n}{k},
\]

we find

\[
P(B) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}.
\]

(5.30)

For large \(n\) this gives,

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

**Example 5.33 (Expected number of coincidences).** 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

\[
EN = \sum_{i=1}^{n} E X_i = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = 1.
\]

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

\[
\begin{array}{c|c|c}
\omega & N(\omega) \\
\hline
1 & 3 & 3 \\
2 & 3 & 2 \\
3 & 1 & 2 \\
1 & 2 & 3 \\
2 & 1 & 3 \\
3 & 2 & 1 \\
\end{array}
\]

and so

\[
P(\exists \text{ a fixed point}) = \frac{4}{6} = \frac{2}{3} \approx 0.67 \approx 0.632
\]

while

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

and

\[
EN = \frac{1}{6} (3 + 1 + 1 + 0 + 0 + 1) = 1.
\]

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

1. \((\Omega, A, P)\) is a finitely additive probability space, so \(P(\Omega) = 1\),
2. \(A_i \in A\) 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

\[
S_k := \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})
\]

\[
= \sum_{A \subseteq \{1, 2, \ldots, n\} \, |A| = k} P(\cap_{i \in A} A_i).
\]

Exercise 5.7. For \( 1 \leq k \leq n \), show;

1. (as functions on \( \Omega \)) that

\[
\binom{N}{k} = \sum_{A \subseteq \{1, 2, \ldots, n\} \, |A| = k} 1_{\cap_{i \in A} A_i},
\]

(5.31)

where by definition

\[
\binom{m}{k} = \begin{cases} 
0 & \text{if } k > m \\
\frac{m!}{k!(m-k)!} & \text{if } 1 \leq k \leq m \\
1 & \text{if } k = 0
\end{cases}.
\]

(5.32)

2. Conclude from Eq. (5.31) that for all \( z \in \mathbb{C} \),

\[
(1 + z)^N = 1 + \sum_{k=1}^{n} z^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}}
\]

(5.33)

provided \( (1 + z)^0 = 1 \) even when \( z = -1 \).

3. Conclude from Eq. (5.31) that \( S_k = E_P(\binom{N}{k}) \).

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

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

(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 \( E \left[(1 + z)^N\right] \) explicitly.

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

\[
P(N = m) = \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} S_k
\]

\[
= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})
\]

\[
\text{wherein we have used,}
\]

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

**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:

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

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

\[
S_k = \binom{n}{k} \frac{(n-k)!}{k!} = \frac{1}{k!}
\]

Therefore it follows from Exercise 5.9 that

\[
P(\exists \text{ exactly } m \text{ fixed points}) = P(N = m)
\]

\[
= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!}
\]

\[
= \frac{1}{m!} \sum_{k=m}^{n} (-1)^{k-m} \frac{1}{(k-m)!}.
\]

So if \( n \) is much bigger than \( m \) we may conclude that

\[
P(\exists \text{ exactly } m \text{ fixed points}) \approx \frac{1}{m!} e^{-1}.
\]

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

\[
P(\exists \text{ a fixed point}) = \sum_{m=1}^{n} P(N = m)
\]

\[
= \sum_{m=1}^{n} \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!}
\]

\[
= \sum_{1 \leq m \leq k \leq n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!}
\]

\[
= \frac{n}{k!} \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^{k-m} \binom{k}{m} \frac{1}{k!}
\]

\[
= \sum_{k=1}^{n} \left[ \sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} \right] \frac{1}{k!}
\]

\[
= -\sum_{k=1}^{n} (-1)^{k} \frac{1}{k!}
\]

\[
= 1 - e^{-1}.
\]
5.5.1 * 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 5.36. 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 \( 0 \leq k \leq n \), then

\[
\sum_{l=0}^{k} (-1)^l \binom{n}{l} = (-1)^k \binom{n-1}{k} 1_{n>0} + 1_{n=0}.
\]  \hfill (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,

\[
\sum_{l=0}^{k+1} (-1)^l \binom{n}{l} = (-1)^k \binom{n-1}{k} + \binom{n}{k+1}
\]

\[
= \frac{(-1)^{k+1}}{(k+1)!} [n(n-1)\ldots(n-k) - (k+1)(n-1)\ldots(n-k)]
\]

\[
= \frac{(-1)^{k+1}}{k+1} \left[ (n-1)\ldots(n-k)(n-(k+1)) \right] = (-1)^{k+1} \binom{n-1}{k+1}.
\]

**Second proof.** Let \( \Omega = \{1,2,\ldots,n\} \) and observe that

\[
m_k := \sum_{l=0}^{k} (-1)^l \binom{n}{l} = \sum_{l=0}^{k} (-1)^l \cdot \#(A \in 2^\Omega : \#(A) = l)
\]

\[
= \sum_{A \in 2^\Omega : \#(A) \leq k} (-1)^\#(A).
\]  \hfill (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 vice versa. Moreover, if we let

\[
\Gamma_k := \{ A \in 2^\Omega : \#(A) \leq k \text{ and } 1 \in A \text{ if } \#(A) = k \},
\]

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

\[
\begin{align*}
\sum_{A \in \Gamma_k} (-1)^\#(A) &= \sum_{A \in \Gamma_k} (-1)^\#(T(A)) = \sum_{A \in \Gamma_k} -(-1)^\#(A) \\
\end{align*}
\]

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

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

**Corollary 5.36 (Bonferroni Inequalities).** 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 \( n = 1,2,\ldots,M, N := \sum_{n=1}^M A_n \), and

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

Then for \( 1 \leq k \leq M \),

\[
\mu(\bigcup_{n=1}^M A_n) = \sum_{i=1}^k (-1)^{i+1} S_i + (-1)^k \mathbb{E}_\mu \left[ \binom{N-1}{k} \right].
\]  \hfill (5.37)

This leads to the Bonferroni inequalities;

\[
\mu(\bigcup_{n=1}^M A_n) \leq \sum_{i=1}^k (-1)^{i+1} S_i \text{ if } k \text{ is odd}
\]

and

\[
\mu(\bigcup_{n=1}^M A_n) \geq \sum_{i=1}^k (-1)^{i+1} S_i \text{ if } k \text{ is even}.
\]

**Proof.** By Lemma 5.36

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

Therefore integrating this equation with respect to \( \mu \) gives,

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

and therefore,
The Bonferroni inequalities are a simple consequence of Eq. (5.37) and the fact that
\[
\binom{N-1}{k} \geq 0 \implies \mathbb{E}_\mu \left( \binom{N-1}{k} \right) \geq 0.
\]

5.6 * Finitely Additive Measures on \( \mathbb{R}^d \) and \([0, 1]^d\)

(Riesz Markov theorem is in Section 10.3 below.) Let us begin by describing finite measure on \( \mathcal{A} = \mathcal{A}(\mathbb{R}^d) \) – the subalgebra of \( 2^{\mathbb{R}^d} \) generated by \( \mathcal{E} \times \cdots \times \mathcal{E} \) where \( \mathcal{E} \) is the semi-algebra of subsets of \( \mathbb{R} \) defined by
\[
\mathcal{E} := \{ \mathbb{R} \cap (a, b) : -\infty \leq a < b \leq \infty \} \subset 2^\mathbb{R}.
\]
For \( a, b \in \mathbb{R}^d \) 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,
\[
\mathcal{E} \times \cdots \times \mathcal{E} = \{ \mathbf{a}, \mathbf{b} : \mathbf{a}, \mathbf{b} \in \mathbb{R}^d \text{ with } a \leq b \}.
\]

**Notation 5.37** If \( a, b \in \mathbb{R}^d \) and \( \gamma \subset \{1, 2, \ldots, d\} \), let \( a_\gamma \times b_\gamma \) be the point in \( \mathbb{R}^d \) defined by
\[
(a_\gamma \times b_\gamma)_j := \begin{cases} 
  a_j & \text{if } j \in \gamma \\
  b_j & \text{if } j \in \gamma^c.
\end{cases}
\]

**Lemma 5.38.** For all \( a, b \in \mathbb{R}^d \) with \( a \leq b \) we have,
\[
1_{(a, b)}(x) = \sum_{\gamma \subset \{1, 2, \ldots, d\}} (-1)^{|\gamma|} 1_{(-\infty, a_\gamma \times b_\gamma]}(x).
\]

**Proof.** If \( x \in \mathbb{R}^d \), then
\[
1_{(a, b)}(x) = \prod_{i=1}^d 1_{(a_i, b_i]}(x_i) = \prod_{i=1}^d [1_{(-\infty, b_i]}(x_i) - 1_{(-\infty, a_i]}(x_i)]
\]
\[
= \sum_{\gamma \subset \{1, 2, \ldots, d\}} (-1)^{|\gamma|} \prod_{i \in \gamma} 1_{(-\infty, a_i]}(x_i) \cdot \prod_{j \in \gamma^c} 1_{(-\infty, b_j]}(x_j)
\]
\[
= \sum_{\gamma \subset \{1, 2, \ldots, d\}} (-1)^{|\gamma|} 1_{(-\infty, a_\gamma \times b_\gamma]}(x).
\]

**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 \( b \in \mathbb{R}^d \) let \( F(b) := \mu((-\infty, b]) \). Then for all \( a, b \in \mathbb{R}^d \) with \( a \leq b \),
\[
\mu((a, b]) = \sum_{\gamma \subset \{1, 2, \ldots, d\}} (-1)^{|\gamma|} F(a_\gamma \times b_\gamma).
\]

**Proof.** The result follows directly by integrating Eq. (5.38) relative to \( \mu \) while using \( \mu((-\infty, a_\gamma \times b_\gamma]) = F(a_\gamma \times b_\gamma) \) 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
\[
(a, b] = S_b \setminus [S_{(a_1, b_2)} \cup S_{(b_1, a_2)}]
\]
and hence
\[
\mu((a, b]) = \mu(S_b) - \mu(S_{(a_1, b_2)} \cup S_{(b_1, a_2)})
\]
\[
= \mu(S_b) - [\mu(S_{(a_1, b_2)}) + \mu(S_{(b_1, a_2)}) - \mu(S_{(a_1, b_2)} \cap S_{(b_1, a_2)})]
\]
\[
= \mu(S_b) - [\mu(S_{(a_1, b_2)}) + \mu(S_{(b_1, a_2)}) + \mu(S_0)]
\]
\[
= F(b) - \mu(S_{(a_1, b_2)}) - F(b_1, a_2) + F(a),
\]
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 \( b \in \mathbb{R} \), \( \mu_b(C) := \mu(C \times (-\infty, b]) \) for \( C \in \mathcal{A}(\mathbb{R}^{d-1}) \) defines a finitely additive measure on \( \mathcal{A}(\mathbb{R}^{d-1}) \). Moreover, if \( a, b \in \mathbb{R} \) with \( a \leq b \) we further have \( \mu(C \times (a, b]) = \mu_b(C) - \mu_a(C) \).

**Proposition 5.41.** To every function, \( F : \mathbb{R}^d \to V \), there exists a unique finitely additive measure \( \mu_F \) on \( \mathcal{A}(\mathbb{R}^d) \) such that
\[
\mu_F((a, b]) = \sum_{\gamma \subset \{1, 2, \ldots, d\}} (-1)^{|\gamma|} F(a_\gamma \times b_\gamma)
\]
for all \( a, b \in \mathbb{R}^d \) with \( a \leq b \).

**Proof.** The proof for \( d = 1 \) is completely analogous to the proof of Proposition 5.9 and so will be omitted. To each \( a = (a_1, \ldots, a_d) \in \mathbb{R}^d \), 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.

Suppose that \( d \geq 2 \) and we have proved the proposition when \( d \) is replaced by \( d - 1 \). Then for \( c \in \mathbb{R} \) let \( \mu_c = \mu_{F(c)} \) be the unique finitely additive measure on \( \mathcal{A}(\mathbb{R}^{d-1}) \) (guaranteed to exist by the induction hypothesis) such that for all \( a', b' \in \mathbb{R}^{d-1} \) with \( a' \leq b' \) we have
\[ \mu_c((a', b')] = \sum_{\gamma \in \{1, 2, \ldots, d-1\}} (-1)^{\gamma |} F(a'_\gamma \times b'_{\gamma c}, c). \quad (5.41) \]

By Proposition 5.15 there is a unique finitely additive measure \((\mu_F)\) on \(A(\mathbb{R}^d) = A(\mathbb{R}^{d-1}) \otimes A(\mathbb{R})\) such that

\[
\mu_F(C \times (a_d, b_d)) = \mu_{b_d}(C) - \mu_{a_d}(C) \quad \forall C \in A(\mathbb{R}^{d-1}) \quad \text{and} \quad -\infty < a_d \leq b_d \leq \infty.
\]

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

\[
\mu_F((a, b]) = \sum_{\gamma \in \{1, 2, \ldots, d-1\}} (-1)^{\gamma |} \left[ F(a'_\gamma \times b'_{\gamma c}, b) - F(a'_\gamma \times b'_{\gamma c}, a) \right]
\]

\[
= \sum_{\gamma \in \{1, 2, \ldots, d\}} (-1)^{\gamma |} F(a'_\gamma \times b_{\gamma c}).
\]

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.

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 \(a \leq b\), let

\[ \{a, b]\} = \{[a_1, b_1] \times \cdots \times [a_d, b_d] \}
\]

where \(i = (\text{if } a_i > 0 \text{ and is either } (\text{ or } [\text{ 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 \(E \subset 2^{\{0, 1\}}\) are the sets of the form \([0, b]\) with \(0 \leq b \leq 1\) or \((a, b)\) with \(0 \leq a < b \leq 1\), then \(A([0, 1]^d)\) is generated by the semi-algebra, \(E \times \cdots \times E\), consisting of sets of the form \([a, b]\) with \(0 \leq a \leq b \leq 1\).

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

\[ 1_{\{a, b\}} = \sum_{\gamma \in \{1, 2, \ldots, d\}} w_{\gamma} (-1)^{|\gamma|} 1_{(-\infty, a_{\gamma c}] \times b_{\gamma c}]} \quad (5.42) \]

where

\[ w_{\gamma} = \begin{cases} 1 & \text{if } i = (\text{for all } i \in \gamma) \\ 0 & \text{otherwise.} \end{cases} \]

In particular,

\[ \sum_{|\gamma| \text{ even}} w_{\gamma} 1_{[0, a_{\gamma c}] \times b_{\gamma c}] = \sum_{|\gamma| \text{ odd}} w_{\gamma} 1_{[0, a_{\gamma c}] \times b_{\gamma c}] + 1_{\{a, b\}} \geq \sum_{|\gamma| \text{ odd}} w_{\gamma} 1_{[0, a_{\gamma c}] \times b_{\gamma c}]}. \quad (5.43) \]

**Proof**. Let

\[ w_i := \begin{cases} 1 & \text{if } i = (\text{for all } i \in \gamma) \\ 0 & \text{otherwise.} \end{cases} \]

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

\[ 1_{\{a, b\}} (x) = \prod_{i=1}^d 1_{[0, b_i]} (x_i) = \prod_{i=1}^d 1_{[0, a_i]} (x_i) - \prod_{j \in \gamma'} \prod_{i=1}^d 1_{[-\infty, b_i]} (x_i) \]

\[ = \sum_{\gamma \in \{1, 2, \ldots, d\}} (-1)^{|\gamma|} \prod_{i \in \gamma} 1_{[0, a_i]} (x_i) \prod_{j \in \gamma'} 1_{[-\infty, b_j]} (x_j) \]

\[ = \sum_{\gamma \in \{1, 2, \ldots, d\}} (-1)^{|\gamma|} w_{\gamma} 1_{[0, a_{\gamma c}] \times b_{\gamma c}] (x). \]

**Corollary 5.45**. Suppose that \(V\) is a vector space, \(\mu : A([0, 1]^d) \rightarrow V\) 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 \(a \leq b\),

\[ \mu_{\{a, b\}} = \sum_{\gamma \in \{1, 2, \ldots, d\}} w_{\gamma} (-1)^{|\gamma|} F(a_{\gamma c} \times b_{\gamma c}). \quad (5.44) \]

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

Second proof. We carry this proof out only in the case \(d = 2\). We have

\[ \mu_{\{a, b\}} = \mu_{\{a_1, b_1, b_2\}} = \mu_{\{a_1, b_1\} \times [0, b_2]} - w_2 \mu_{\{a_1, b_1\} \times [0, a_2]} = \mu_{\{a_1, b_1\} \times [0, b_2]} - w_1 \mu_{\{a_1, b_1\} \times [0, a_2]} \]

\[ = w_2 \mu_{\{a_1, b_1\} \times [0, a_2]} - w_1 \mu_{\{a_1, b_1\} \times [0, a_2]} \]

\[ = F(b) - w_1 F(a_1, b_1) - w_2 F(a_1, b_2) - w_1 w_2 F(a). \]

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, \(F : [0, 1]^d \rightarrow V\), there exists a unique finitely additive measure \((\mu_F)\) on \(A([0, 1]^d)\) such that Eq. (5.44) holds for all \(a, b \in [0, 1]^d\) with \(a \leq b\).

**Proof**. The proof for \(d = 1\) is completely analogous to the proof of Proposition 5.9 and so will be omitted. Now suppose that \(d \geq 2\) and we have proved the Proposition when \(d\) is replaced by \(d - 1\). Then for \(c \in [0, 1]\) let \(\mu_c = \mu_{F(c, c)}\) be the unique finitely additive measure on \(A([0, 1]^{d-1})\) guaranteed to exist.
by the induction hypothesis) such that for all \( a', b' \in [0, 1]^{d-1} \) with \( a' \leq b' \) we have
\[
\mu_c(\{a', b'\}) = \sum_{\gamma \in \{1, 2, \ldots, d-1\}} w_{\gamma}(-1)^{\gamma} F(a_{\gamma}' \times b_{\gamma}', c).
\] (5.45)

By Proposition 5.15 there is a unique finitely additive measure \( \mu_F \) on \( \mathcal{A}(\{0, 1\}^{d}) = \mathcal{A} \left( 0, 1 \right)^{d-1} \) such that
\[
\mu_F(\{a, b\}) = \mu_{bd}(C) - w_{\delta} \mu_{\alpha d}(C) \quad \forall \ C \in \mathcal{A} \left( 0, 1 \right)^{d-1} \quad & \ 0 \leq a_d \leq b_d \leq 1.
\]

Letting \( a \leq b \) with \( a, b \in [0, 1]^d \) and using Eq. (5.41) it is not to hard to show,
\[
\mu_F(\{a, b\}) = \sum_{\gamma \in \{1, 2, \ldots, d-1\}} w_{\gamma}(-1)^{\gamma} [F(a_{\gamma}' \times b_{\gamma}', \bar{b}) - w_d F(a_{\gamma}' \times b_{\gamma}', a)]
\]
\[
= \sum_{\gamma \in \{1, 2, \ldots, d\}} w_{\gamma}(-1)^{\gamma} F(a_{\gamma} \times b_{\gamma}') .
\]

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.

\[\square\]

5.7 * Appendix: Riemann Stieljes 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
\[
\mathbb{E}_\mu f := \int_{[0, T]} f d\mu := \sum_{\lambda \in \mathcal{C}} \lambda \mu(f = \lambda) \quad \forall \ f \in \mathbb{S}(\mathcal{A}) .
\] (5.46)

**Notation 5.47** For any function, \( f : \Omega \to \mathbb{C} \) let \( \|f\|_u := \sup_{\omega \in \Omega} |f(\omega)| \).

Further, let \( \mathbb{S} := \mathbb{S}(\mathcal{A}) \) 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 \( f \in \mathbb{S}(\mathcal{A}) \),
\[
|\mathbb{E}_\mu f| \leq \mu(\Omega) \|f\|_u .
\] (5.47)

2. If \( f \in \mathbb{S} \) 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 \( \mathbb{E}_\mu : \mathbb{S} \to \mathbb{C} \) is still linear and still satisfies Eq. (5.47).

4. Show \( |f| \in \mathbb{S} \) if \( f \in \mathbb{S} \) and that Eq. (5.25) is still valid, i.e. \( |\mathbb{E}_\mu f| \leq \mathbb{E}_\mu |f| \) for all \( f \in \mathbb{S} \).

Let us now specialize the above results to the case where \( \Omega = [0, T] \) for some \( T < \infty \). Let \( \mathcal{S} := \{(a, b) : 0 \leq a \leq b \leq T\} \) \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 Stieljes 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 \( 0 \leq a \leq b \leq T \) and \( \mu_F(\{0\}) = 0 \). (In fact one could allow for \( \mu_F(\{0\}) = \lambda \) for any \( \lambda \geq 0 \), but we would then have to write \( \mu_{F,\lambda} \) rather than \( \mu_F \).)

2. Show \( C([0, 1], \mathcal{S}) \subset \overline{\mathbb{S}(\mathcal{A})} \). 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 \( t_{i-1} \leq c_i \leq t_i \) for each \( i \). Then for \( f \in C([0, 1], \mathcal{S}) \), let
\[
f_{\pi,c} := f(0) 1_{(0)} + \sum_{i=1}^n \left( f(c_i) 1_{(t_{i-1}, t_i]} \right) .
\] (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 \( t_{i-1} \leq c_i \leq t_i \).

It is customary to write \( \int_0^T f dF \) for \( \int_{[0, T]} f d\mu_F \). This integral satisfies the estimates,
\[
\left| \int_{[0, T]} f d\mu_F \right| \leq \int_{[0, T]} |f| d\mu_F \leq \|f\|_u (F(T) - F(0)) \quad \forall \ f \in \mathbb{S}(\mathcal{A}) .
\]

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
\[ G(x) = \lambda \cdot 1_{x \geq a} = \begin{cases} \lambda & \text{if } x \geq a \\ 0 & \text{if } x < a \end{cases}. \]

1. Explicitly compute \( \int_{[0,T]} f d\mu_G \) for all \( f \in C([0,1], \mathbb{C}) \).
2. If \( F(x) = x + \lambda \cdot 1_{x \geq a} \) 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 \( F(x) \neq G(x) \) for at most countably many points, \( x \in (0, T) \). Show
\[ \int_{[0,T]} f d\mu_G = \int_{[0,T]} f d\mu_G \text{ for all } f \in C([0,1], \mathbb{C}). \] (5.49)

Note well, given \( F(0) = G(0) \), \( \mu_F = \mu_G \) on \( A \) 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 24.18 So if we only want to integrate continuous functions, we may always assume that \( F : [0, T] \to \mathbb{R} \) is right continuous.

### 5.8 * 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 \( \mu \circ \nu \) be the product measure on \( \mathcal{A} \circ \mathcal{B} \) as described in Example

**Theorem 5.49 (Tonelli’s Theorem).** If \( f \in S_+ (\mathcal{A} \circ \mathcal{B}) \) then for each \( x \in X, \) \( f (x, \cdot) \in S_+ (\mathcal{B}) \) and \( X \ni x \to \int_Y f (x, y) d\nu (y) \) is in \( S_+ (\mathcal{A}) \) and moreover,
\[ \int_{X \times Y} f (x, y) d(\mu \circ \nu) (x, y) = \int_X \left[ \int_Y f (x, y) d\nu (y) \right] d\mu (x). \]

Similarly, for each \( y \in Y, \) \( f (\cdot, y) \in S_+ (\mathcal{A}) \) and \( Y \ni y \to \int_X f (x, y) d\mu (x) \) is in \( S_+ (\mathcal{B}) \) and moreover,
\[ \int_{X \times Y} f (x, y) d(\mu \circ \nu) (x, y) = \int_Y \left[ \int_X f (x, y) d\mu (x) \right] d\nu (y). \]

**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, \( \mu \circ \nu \), on \( \mathcal{A} \circ \mathcal{B} \) such that \( \mu \circ \nu (A \times B) = \mu (A) \nu (B) \) for all \( A \in \mathcal{A} \) and \( B \in \mathcal{B}. \) Moreover if \( f \in S (\mathcal{A} \circ \mathcal{B}) \) then:

1. \( y \to f (x, y) \) is in \( S (\mathcal{B}) \) for all \( x \in X \) and \( x \to f (x, y) \) is in \( S (\mathcal{A}) \) for all \( y \in Y \).
2. \( x \to \int_Y f (x, y) d\nu (y) \) is in \( S (\mathcal{A}) \) and \( y \to \int_X f (x, y) d\mu (x) \) is in \( S (\mathcal{B}) \).
3. We have,
\[ \int_X \left[ \int_Y f (x, y) d\nu (y) \right] d\mu (x) = \int_{X \times Y} f (x, y) d(\mu \circ \nu) (x, y) = \int_Y \left[ \int_X f (x, y) d\mu (x) \right] d\nu (y). \]

We will refer to \( \mu \circ \nu \) as the **product measure** of \( \mu \) and \( \nu \).

**Proof.** According to Eq. (5.11),
\[ 1_C (x, y) = \sum_{i=1}^n 1_{A_i \times B_i} (x, y) = \sum_{i=1}^n 1_{A_i} (x) 1_{B_i} (y), \]
from which it follows that \( 1_C (x, \cdot) \in S (\mathcal{B}) \) 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 \( x \to \int_Y 1_C (x, y) d\nu (y) \in S (\mathcal{A}) \) 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 \( A \odot B \) with the desired properties. Moreover, we have already verified the Theorem in the special case where \( f = 1_C \) with \( C \in A \odot B \). Since the general element, \( f \in \mathbb{S} (A \odot B) \), is a linear combination of such functions, it is easy to verify using the linearity of the integral and the fact that \( \mathbb{S} (A) \) and \( \mathbb{S} (B) \) are vector spaces that the theorem is true in general.

**Example 5.51.** Suppose that \( f \in \mathbb{S} (A) \) and \( g \in \mathbb{S} (B) \). Let \( f \odot g (x, y) := f (x) g (y) \). Since we have,
\[
f \odot g (x, y) = \left( \sum_a a 1_{f=a} (x) \right) \left( \sum_b b 1_{g=b} (y) \right) = \sum_{a,b} ab 1_{\{f=a\} \times \{g=b\}} (x,y),\]

it follows that \( f \odot g \in \mathbb{S} (A \odot B) \). Moreover, using Fubini’s Theorem \( \mu \odot \nu = \int_X \int_Y f \, d\mu \, d\nu \).

**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 5.69 and more generally Section 5.11]

### 5.9 Conditional probabilities and expectations

We begin with the notion of conditional probabilities which tries to capture the dependency of one event on another. Throughout this section let \((\Omega, \mathcal{B}, P)\) be a finitely additive probability space. For motivational purposes we heuristically think that \( P \) may be computed using,
\[
P (B) := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} 1_B (\omega_n) \text{ for all } B \in \mathcal{B},
\]

where \( \{\omega_n\}_{n=1}^\infty \subset \Omega \) are the outcomes of a sequence of identical and independent experiments.

For events \( A \) and \( B \in \mathcal{B} \) 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
\[
P_N (A | B) = \frac{\# \{ k : 1 \leq k \leq N \text{ and } \omega (k) \in A \cap B \}}{\# \{ k : 1 \leq k \leq N \text{ and } \omega (k) \in B \}} = \frac{P_N (A \cap B)}{P_N (B)},
\]

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 5.53.** 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 may now ask a similar question as to what is the expectation of a random variable, \( X : \Omega \to \mathbb{R} \), given an event \( B \). Using the above notation and ideas along with the intuition in Remark 5.24 it is reasonable to write,
\[
E_P [X | B] = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} X (\omega_n) 1_B (\omega_n)
= \lim_{N \to \infty} \frac{N}{\sum_{n=1}^{N} 1_B (\omega_n)} \frac{1}{N} \sum_{n=1}^{N} X (\omega_n) 1_B (\omega_n)
= \frac{1}{P (B)} E [X \cdot 1_B] = E_{P (|B)} [X].
\]

**Definition 5.54 (Conditional expectation 1).** If \( X : \Omega \to \mathbb{R} \) is a \( \mathcal{B} \)-simple function and \( B \in \mathcal{B} \) with \( P (B) > 0 \), then conditional expectation of \( X \) given \( B \) is
\[
E_P [X | B] = \frac{1}{P (B)} E [X \cdot 1_B] = E_{P (|B)} [X].
\]
It turns out that a far reaching generalization of this notion is to defined the conditional expectation of $X$ given a sub-algebra $\mathcal{A} \subset \mathcal{B}$. For this motivational and introductory section we are going to assume that $\mathcal{A}$ is a finite sub-algebra, i.e. $\#(\mathcal{A}) < \infty$, and $X : \Omega \to \mathbb{R}$ is a $\mathcal{B}$ – simple function. Recall from Example 4.19 that for each $\omega \in \Omega$, let $A_\omega := \cap \{A \in \mathcal{A} \mid \omega \in A\}$ and recall that either $A_\omega = A_{\omega'}$ or $A_\omega \cap A_{\omega'} = \emptyset$ for all $\omega, \omega' \in \Omega$. Let $\{B_1, \ldots , B_m\} \subset \mathcal{A}$ be an enumeration of $\{A_\omega \mid \omega \in \Omega\}$ in which case $\{B_1, \ldots , B_m\}$ is a partition of $\Omega$ and $\mathcal{A} = \mathcal{A}(\{B_1, \ldots , B_m\})$.

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 $(i(\omega_n), X(\omega_n))_{n=1}^\infty \subset \{1, \ldots , n\} \times \mathbb{R}$ where $i(\omega) = j$ iff $\omega \in B_j$, i.e. $i := \sum_{j=1}^m 1_{B_j}$. 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 do the best we can and give her the average of the $X(\omega_n)$ for which $\omega_n \in B_j$, i.e. we report that our best guess, $\bar{X}$, to $X$ given we are not able to observe $\omega$ directly but only whether $\omega \in B_j$ for some $j$ is the random variable,

$$\bar{X}(\omega) = \sum_{j=1}^N \mathbb{E}[X|B_j] 1_{B_j}(\omega) \tag{5.50}$$

with the convention (say) that $\mathbb{E}[X|B_j] = 0$ if $P(B_j) = 0$. Noting that $1_{B_j}(\omega) = 1$ iff $\omega \in B_j$ iff $B_j = A_\omega$, we may rewrite this expression as

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)} \mathbb{E}[1_{A_\omega} X] \text{ for all } \omega \in \Omega \tag{5.51}$$

with the convention that $\bar{X}(\omega) = 0$ if $P(A_\omega) = 0$. It should be noted, from Exercise 5.4, that $\bar{X} = \mathbb{E}_A X \in S(A)$.

**Definition 5.55 (Conditional expectation).** Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{B}$ – simple random variable, i.e. $X \in S(\mathcal{B})$ and $\mathcal{A} \subset \mathcal{B}$ as is above, we say that $\bar{X} \in S(A)$ given by either of the previously displayed equations is the **conditional expectation** of $X$ given $\mathcal{A}$ and we denote $\bar{X}$ by $\mathbb{E}[X|A]$ or $\mathbb{E}_A X$. So in summary,

$$\mathbb{E}_A X = \sum_{j=1}^N \mathbb{E}[X|B_j] 1_{B_j} = \sum_{j=1}^m \mathbb{E}_{P(|B_j)} [X] \cdot 1_{B_j} \tag{5.52}$$

with the convention that $\mathbb{E}[X|B_j] = 0$ if $P(B_j) = 0$.

**Exercise 5.13 (Simple conditional expectation).** Let $X \in S(\mathcal{B})$ and, for simplicity, assume all functions are real valued. Prove the following assertions:

1. **(Orthogonal Projection Property 1.)** If $Z \in S(A)$, then

$$\mathbb{E}[XZ] = \mathbb{E}[\bar{X}Z] = \mathbb{E}[\mathbb{E}_A X \cdot Z] \tag{5.53}$$

and

$$\mathbb{E}(\mathbb{E}_A Z)(\omega) = \begin{cases} Z(\omega) & \text{if } P(A_\omega) > 0 \\ 0 & \text{if } P(A_\omega) = 0. \end{cases} \tag{5.54}$$

[Applying Eq. 5.54] with $Z = \mathbb{E}_A X$ then shows $\mathbb{E}_A [\mathbb{E}_A X] = \mathbb{E}_A X$ for all $X \in S(\mathcal{B})$. [1]

2. **(Orthogonal Projection Property 2.)** If $Y \in S(A)$ satisfies, $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all $Z \in S(A)$, then $Y(\omega) = \bar{X}(\omega)$ whenever $P(A_\omega) > 0$. In particular, $P(Y \neq \bar{X}) = 0$. **Hint:** use item 1. to compute $\mathbb{E}[(\bar{X} - Y)^2]$.

3. **(Best Approximation Property.)** For any $Y \in S(A)$,

$$\mathbb{E}[(\bar{X} - Y)^2] \leq \mathbb{E}[(X - Y)^2] \tag{5.55}$$

with equality iff $\bar{X} = Y$ almost surely (a.s. for short), where $\bar{X} = Y$ a.s. iff $P(X \neq Y) = 0$. In words, $\bar{X} = \mathbb{E}_A X$ is the best (“L²”) approximation to $X$ by an $\mathcal{A}$ – measurable random variable.

4. **(Contraction Property.)** $\mathbb{E}[\bar{X}] \leq \mathbb{E}[X]$. (It is typically not true that $|\bar{X}(\omega)| \leq |X(\omega)|$ for all $\omega$.)

5. **(Pull Out Property.)** If $Z \in S(A)$, then

$$\mathbb{E}_A[ZX] = Z \mathbb{E}_A X.$$

**Remark 5.56.** The cleanest way to see that $\mathbb{E}_A$ in Eq. (5.50) is an orthogonal projection is to let

$$(X,Y) := \mathbb{E}[XY] \text{ for all } X,Y \in S(\mathcal{B})$$

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 $\left\{\frac{1_{B_i}}{\sqrt{P(B_i)}}\right\}$ is an orthonormal basis for the subspace $S(A) \subset S(\mathcal{B})$. Therefore orthogonal projection from $S(\mathcal{B})$ onto $S(A)$ is given by

- All in all these results basically show that $\mathbb{E}_A$ is orthogonal projection from $S(\mathcal{B})$ onto $S(A)$ relative to the inner product $$(f,g) = \mathbb{E}[fg] \text{ for all } f,g \in S(\mathcal{B}).$$
Definition 5.58. Suppose that independent provided and therefore, \( \{ t \} \in \Omega \) and therefore,
\[
\sum \frac{1}{P(B_i)} 1_{B_i} \cdot \frac{1}{P(B_i)} = \sum \mathbb{E}[X 1_{B_i}] 1_{B_i},
\]
which is precisely the formula for \( \mathbb{E}_A X \).

5.10 Simple Independence and the Weak Law of Large Numbers

Again we assume throughout this section let \((\Omega, B, P)\) be a finitely additive probability space. 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 dice after the roulette ball has stopped and landed on either black or red. 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 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 now go to the mathematical definitions of independence.

Definition 5.57. Two events, \( A, B \in B \) 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 5.58. Suppose that \( S \) and \( T \) are sets and \( X : \Omega \to S \) and \( Y : \Omega \to T \) are \( B \)-simple functions, i.e. \( X \) and \( Y \) have finite range and \( \{ X = s \}, \{ Y = t \} \in B \) 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 5.59. Suppose that \( S \) and \( T \) are sets and \( X : \Omega \to S \) and \( Y : \Omega \to T \) are \( B \)-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,
\[
f(X)g(Y) = \sum_{s \in S} f(s) g(t) 1_{X = s} 1_{Y = t} = \sum_{s \in S} \sum_{t \in T} f(s) g(t) 1_{X = s, Y = t}
\]
and therefore,
\[
\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].
\]

Exercise 5.14. Suppose that \( S, T, U \) are sets and \( X : \Omega \to S \), \( Y : \Omega \to T \), and \( Z : \Omega \to U \) are \( B \)-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 5.60. 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. \( B \)-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)
\[
\mathbb{E}[f(Z)] = \frac{1}{4} \sum_{x,y \in \{ \pm 1 \}} f(xy) = \frac{1}{2} \sum_{y \in \{ \pm 1 \}} \mathbb{E}[f(X)] = \mathbb{E}f(X),
\]
and
\[
\mathbb{E}[f(Z)g(Y)] = \mathbb{E}[f(XY)]g(Y) = \frac{1}{4} \sum_{x,y \in \{ \pm 1 \}} f(xy)g(y)
\]
\[
= \frac{1}{2} \sum_{y \in \{ \pm 1 \}} [\mathbb{E}f(X)]g(y) = \mathbb{E}f(X) \cdot \mathbb{E}g(Y)
\]
\[
= \mathbb{E}[f(Z)] \cdot \mathbb{E}g(Y).
\]
Exercise 5.16. Suppose that for all (5.15).

Method 2. The reader should verify that for \( \varepsilon, \delta \in \{ \pm 1 \} \) that

\[
P(Z = \delta) = \frac{1}{2},
\]

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

\[
P(Z = 1) \cdot P((X, Y) = (1, -1)) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} \neq 0.
\]

Exercise 5.17. Suppose now that \( S \) is a finite set, \( \Omega = S^n, B = 2^\Omega \), and \( X_i : \Omega \to S \) is projection onto the \( i \)th factor of \( \Omega \), i.e.

\[
X_i(\omega) = \omega_i \in S \quad \forall \omega = (\omega_1, \ldots, \omega_n) \in \Omega.
\]

Let \( P \) be a probability measure on \((\Omega, B)\). Show \( \{X_i\}_{i=1}^n \) are \( P \)-independent if \( P \) there are functions, \( q_i : S \to [0, 1] \) (for \( 1 \leq i \leq n \)) such that \( \sum_{s \in S} q_i(s) = 1 \) and

\[
P(\{s\}) = \prod_{i=1}^n q_i(s_i) \quad \forall s \in \Omega.
\]

and define \( q_i(s) := P(X_i = s) \) for all \( s \in S \) and \( 1 \leq i \leq n \).

**Example 5.64 (Heuristics of independence and conditional expectations).** Let us suppose that we have an experiment consisting of spinning a spinner with values in \( A_1 = \{1, 2, \ldots, 10\} \) and rolling a die with values in \( A_2 = \{1, 2, 3, 4, 5, 6\} \). So the outcome of an experiment is represented by a point, \( \omega = (x, y) \in \Omega = A_1 \times A_2 \), where \( x \) is the outcome of the die and \( y \) is the result of spinning the spinner.

\[
\mathcal{A} = \mathcal{A}(X) = X^{-1}(2^{A_1}) = \{X^{-1}(A) : A \subset A_1\} \subset B,
\]

so that \( \mathcal{A} \) is the smallest algebra of subsets of \( \Omega \) such that \( \{X = x\} \in \mathcal{A} \) for all \( x \in A_1 \). Notice that the partition associated to \( \mathcal{A} \) is precisely

\[
\{\{X = 1\}, \{X = 2\}, \ldots, \{X = 10\}\}.
\]

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 \( g : A_2 \to \mathbb{R} \) is a function, we have (heuristically) that
\[
\mathbb{E}_A [g(Y)](x, y) = \lim_{N \to \infty} \frac{\sum_{n=1}^{N} g(Y(x_n, y_n))}{\sum_{n=1}^{N} 1_{x_n=x}}
\]
As the \( \{y_n\} \) sequence of results are independent of the \( \{x_n\} \) sequence, we should expect by the usual mantra\(^5\) that
\[
\lim_{N \to \infty} \frac{\sum_{n=1}^{N} g(y_n) 1_{x_n=x}}{\sum_{n=1}^{N} 1_{x_n=x}} = \lim_{N \to \infty} \frac{1}{M(N)} \sum_{n=1}^{M(N)} g(y_n) = \mathbb{E}[g(Y)],
\]
where \( M(N) = \sum_{n=1}^{N} 1_{x_n=x} \) and \( \{\bar{y}_1, \bar{y}_2, \ldots\} = \{y_i : 1_{x_i=x}\} \). (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 \( \mathbb{E}_A [g(Y)] : \Omega \to \mathbb{R} \) is the constant function;
\[
\mathbb{E}_A [g(Y)](x, y) = \mathbb{E}[g(Y)] \text{ for all } (x, y) \in \Omega. \tag{5.58}
\]
Let us further observe that if \( f : A_1 \to \mathbb{R} \) is any other function, then \( f(X) \) is an \( A \)– simple function and therefore by Eq. \( (5.58) \) and Exercise \( 5.13 \)
\[
\mathbb{E}[f(X) \cdot g(Y)] = \mathbb{E}[f(X) \cdot \mathbb{E}_A [g(Y)]] = \mathbb{E}[f(X) \cdot \mathbb{E}_A [g(Y)]] = \mathbb{E}[f(X) \cdot g(Y)].
\]

**Theorem 5.66 (Independence result summary).** Let \((\Omega, \mathcal{B}, P)\) be a finitely additive probability space, \( A \) be a finite set, and \( X, Y : \Omega \to A \) be two \( A \)– measurable simple functions, i.e. \( \{X = i\} \in \mathcal{B} \) and \( \{Y = j\} \in \mathcal{B} \) for all \( x, y \in A \). Further let \( A = A(X) := A(\{X = i\} : x \in X) \). Then the following are equivalent;

1. \( P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \) for all \( x \in A \) and \( y \in A \).
2. \( \mathbb{E}[f(X) \cdot g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)] \) for all functions, \( f : A \to \mathbb{R} \) and \( g : A \to \mathbb{R} \).
3. \( \mathbb{E}_A(g(Y)) = \mathbb{E}[g(Y)] \) for all \( g : A \to \mathbb{R} \), and
4. \( \mathbb{E}_A(f(X)) = \mathbb{E}[f(X)] \) for all \( f : A \to \mathbb{R} \).

We say that \( X \) and \( Y \) are \( P \)– independent if any one (and hence all) of the above conditions holds.

**Exercise 5.18 (A Weak Law of Large Numbers).** Suppose that \( A \subset \mathbb{R} \) is a finite set, \( n \in \mathbb{N} \), \( \Omega = A^n \), \( p(\omega) = \prod_{i=1}^{n} q(\omega_i) \) where \( q : A \to [0, 1] \) such that \( \sum_{\lambda \in \mathcal{A}} q(\lambda) = 1 \), and let \( P : 2^\Omega \to [0, 1] \) be the probability measure defined as in Eq. \( (5.57) \) with \( S \) replaced by \( A \). Further let \( X_i(\omega) = \omega_i \) for \( i = 1, 2, \ldots, n \), \( \xi := \mathbb{E}X_1 \), \( \sigma^2 := \mathbb{E}(X_i - \xi)^2 \), and
\[
S_n = \frac{1}{n} (X_1 + \cdots + X_n).
\]

1. Show, \( \xi = \sum_{\lambda \in \mathcal{A}} \lambda q(\lambda) \) and
\[
\sigma^2 = \sum_{\lambda \in \mathcal{A}} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda \in \mathcal{A}} \lambda^2 q(\lambda) - \xi^2. \tag{5.59}
\]
2. Show, \( \mathbb{E}S_n = \xi \).
3. Let \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \). Show
\[
\mathbb{E}[(X_i - \xi)(X_j - \xi)] = \delta_{ij}\sigma^2.
\]
4. Using \( S_n - \xi \) may be expressed as, \( \frac{1}{n} \sum_{i=1}^{n} (X_i - \xi) \), show
\[
\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{5.60}
\]
5. Conclude using Eq. \( (5.60) \) and Chebyshev’s Inequality (Remark \( 5.29 \)) that
\[
P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{5.61}
\]
So for large \( n \), \( S_n \) is concentrated near \( \xi = \mathbb{E}X_1 \) with probability approaching \( 1 \) for \( n \) large. This is a version of the weak law of large numbers.
Suppose that

Theorem 5.69 (Weierstrass Approximation Theorem via Bernstein’s

Let

Exercise 5.19 (Bernoulli Random Variables). Let

\begin{align*}
\t X \& := \mathbb{E} \cdot X \\
\t \text{The variance of } X,
\end{align*}

\[ \text{Var} (X) := \text{Cov} (X, X) = \mathbb{E} [X^2] - (\mathbb{E} X)^2 \]

We say that \( X \) and \( Y \) are **uncorrelated** if \( \text{Cov} (X, Y) = 0 \), i.e. \( \mathbb{E} [XY] = \mathbb{E} X \cdot \mathbb{E} Y \). More generally we say \( \{X_k\}_{k=1}^{\infty} \subset \mathbb{S} (B) \) are uncorrelated iff \( \text{Cov} (X_i, X_j) = 0 \) for all \( i \neq j \).

Remark 5.68. 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 \( \text{Cov} (X, Y) = 0 \).

2. If you look at your proof of the weak law of large numbers in Exercise 5.18, 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 5.19 (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 \( Q = x \delta_1 \) and \( (1-x) \delta_0 \), i.e. \( Q \{0\} = 1 - x \) and \( Q \{1\} = x \). Verify,

\[ \begin{align*}
\xi (x) & := \mathbb{E} Q X = x \\
\sigma^2 (x) & := \mathbb{E} Q (X - x)^2 = (1 - x) x \leq 1/4.
\end{align*} \]

Theorem 5.69 (Weierstrass Approximation Theorem via Bernstein’s

Polynomials.). Suppose that \( f \in C([0, 1], \mathbb{C}) \) and

\[ p_n (x) := \sum_{k=0}^{n} \binom{n}{k} f (\frac{k}{n}) x^k (1 - x)^{n-k}. \]

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

\[ P_x (\{\omega\}) = q (\omega_1) \cdots q (\omega_n) = x^{\sum_{i=1}^{n} \omega_i} \cdot (1 - x)^{1 - \sum_{i=1}^{n} \omega_i}. \]

As above, let \( S_n = \frac{1}{n} (X_1 + \cdots + X_n) \), where \( X_i (\omega) = \omega_i \) and observe that

\[ P_x (S_n = \frac{k}{n}) = \binom{n}{k} x^k (1 - x)^{n-k}. \]

Therefore, writing \( \mathbb{E} \) for \( \mathbb{E} P_x \), we have

\[ \mathbb{E} [f (S_n)] = \sum_{k=0}^{n} \binom{n}{k} \left( \frac{k}{n} \right) x^k (1 - x)^{n-k} = p_n (x). \]

Hence we find

\[ |p_n (x) - f (x)| = |\mathbb{E} f (S_n) - f (x)| = |\mathbb{E} [f (S_n) - f (x)]| \]

\[ \leq \mathbb{E} |f (S_n) - f (x)| \]

\[ = \mathbb{E} \{f (S_n) - f (x) : |S_n - x| \geq \varepsilon\} \]

\[ + \mathbb{E} \{f (S_n) - f (x) : |S_n - x| < \varepsilon\} \]

\[ \leq 2M \cdot P_x (|S_n - x| \geq \varepsilon) + \delta (\varepsilon) \]

where

\[ M := \max_{y \in [0, 1]} |f (y)| \]

\[ \delta (\varepsilon) := \sup \{|f (y) - f (x)| : x, y \in [0, 1] \text{ and } |y - x| \leq \varepsilon\} \]

is the modulus of continuity of \( f \). Now by the above exercises,

\[ P_x (|S_n - x| \geq \varepsilon) \leq \frac{1}{4n \varepsilon^2} \]  

(5.62)

and hence we may conclude that

\[ \max_{x \in [0, 1]} |p_n (x) - f (x)| \leq \frac{M}{2n \varepsilon^2} + \delta (\varepsilon) \]

and therefore, that

\[ \lim_{n \to \infty} \sup_{x \in [0, 1]} |p_n (x) - f (x)| \leq \delta (\varepsilon). \]

This completes the proof, since by uniform continuity of \( f \), \( \delta (\varepsilon) \downarrow 0 \) as \( \varepsilon \downarrow 0 \). 

5.10 Simple Independence and the Weak Law of Large Numbers

The main goal of this subsection is to prove Theorem 5.70 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 5.69. All of the results in this section have natural generalization to higher dimensions as well, see Theorem 5.76.
5 Finitely Additive Measures / Integration

![Plots of $P_x(S = k/n)$ versus $k/n$ for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).]

**Theorem 5.70 (Weierstrass Approximation Theorem).** Suppose that $K = [0, 1]^2$, $f \in C(K, \mathbb{C})$, and

$$p_n(x, y) := \sum_{k,l=0}^\infty f \left( \frac{k}{n}, \frac{l}{n} \right) \binom{n}{k} \binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}.$$  

(5.63)

Then $p_n \to f$ uniformly on $K$.

**Proof.** We are going to follow the argument given in the proof of Theorem 5.69. 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 $1 \leq i \leq n$. 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

$$E[f(S_n, T_n)] = \sum_{k,l=0}^\infty f \left( \frac{k}{n}, \frac{l}{n} \right) P(n \cdot S_n = k, n \cdot T_n = l) = p_n(x, y)$$

where $p_n(x, y)$ is the polynomial given in Eq. (5.63) 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 $M = \sup \{|f(x, y)| : (x, y) \in K\}$, $\varepsilon > 0$,

$$\delta_\varepsilon = \sup \{|f(x', y') - f(x, y)| : (x, y), (x', y') \in K \text{ and } \|(x', y') - (x, y)\| \leq \varepsilon\},$$

and

$$A := \{(S_n, T_n) - (x, y) \| > \varepsilon\},$$

we have,

$$|f(x, y) - p_n(x, y)| = |E(f(x, y) - f((S_n, T_n)))|$$

$$\leq E|f(x, y) - f((S_n, T_n))|$$

$$= E|f(x, y) - f(S_n, T_n) : A|$$

$$+ E|f(x, y) - f(S_n, T_n) : A^c|$$

$$\leq 2M \cdot P(A) + \delta_\varepsilon \cdot P(A^c)$$

(5.64)

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. (5.62) we know

$$P(A) \leq P \left(|S_n - x| > \varepsilon/\sqrt{2}\right) + P \left(|T_n - y| > \varepsilon/\sqrt{2}\right)$$

$$\leq \frac{1}{2n\varepsilon^2} + \frac{1}{2n\varepsilon^2} = \frac{1}{n\varepsilon^2}.$$  

(5.65)

Using this estimate in Eq. (5.64) gives,

$$|f(x, y) - p_n(x, y)| \leq 2M \cdot \frac{1}{n\varepsilon^2} + \delta_\varepsilon$$

and as right is independent of $(x, y) \in K$ we may conclude,

$$\lim_{n \to \infty} \sup_{(x, y) \in K} |f(x, y) - p_n(x, y)| \leq \delta_\varepsilon$$

which completes the proof since $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$ because $f$ is uniformly continuous on $K$.  

**Remark 5.71.** We can easily improve our estimate on $P(A)$ in Eq. (5.65) by a factor of two as follows. As in the proof of Theorem 5.69

$$E \left[\|(S_n, T_n) - (x, y)\|^2\right] = E \left[(S_n - x)^2 + (T_n - y)^2\right]$$

$$= \text{Var}(S_n) + \text{Var}(T_n)$$

$$= \frac{1}{n} x(1-x) + y(1-y) \leq \frac{1}{2n}.$$

Corollary 5.72. 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 5.70 to find a polynomial, $p(x, y)$, such that $\text{sup}_{(x, y) \in [0, 1]^2} |F(x, y) - p(x, y)| \leq \varepsilon$. Letting $\xi = a + x(b - a)$ and $\eta := c + y(d - c)$, it now follows that

$$\text{sup} \{ (\xi, \eta) \in K \} \left| f(\xi, \eta) - p\left(\frac{\xi - a}{b - a}, \frac{\eta - c}{d - c}\right) \right| \leq \varepsilon$$

which completes the proof since $p\left(\frac{\xi - a}{b - a}, \frac{\eta - c}{d - c}\right)$ is a polynomial in $(\xi, \eta)$.

Theorem 5.73 (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 $\text{sup}_{z \in K} \{q_n(z, \bar{z}) - f(z)\} \rightarrow 0$ as $n \rightarrow \infty$ for every $f \in C(K, \mathbb{C})$.

Proof. The mapping $(x, y) \in \mathbb{R} \times \mathbb{R} \rightarrow z = x + iy$ 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 Corollary 5.72.

Example 5.74. Let $K = S^1 = \{ z \in \mathbb{C} : |z| = 1 \}$ and $A$ be the set of polynomials in $(z, \bar{z})$ restricted to $S^1$. Then $A$ 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 5.73 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 5.75 (Density of Trigonometric Polynomials). Any $2\pi$ - periodic continuous function, $f : \mathbb{R} \rightarrow \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

$$p(x) = \sum_{\lambda \in \Lambda} a_{\lambda} e^{i\lambda x},$$

where $\Lambda$ is any 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 $\theta \rightarrow e^{i\theta}$ is a local homeomorphism, i.e. for any $J = (a, b)$ with $b - a < 2\pi$, the map $\theta \in J \rightarrow \hat{J} := \{e^{i\theta} : \theta \in J\} \subset S^1$ is a homeomorphism, it follows that $F(z) = f \circ \phi^{-1}(z)$ for $z \in \hat{J}$. This shows $F$ is continuous when restricted to $\hat{J}$. Since such sets cover $S^1$, it follows that $F$ is continuous.

By Example 5.74, the polynomials in $z$ and $\bar{z} = z^{-1}$ are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z, \bar{z}) = \sum_{0 \leq m, n \leq N} a_{m, n} z^m \bar{z}^n$$

such that $|F(z) - p(z, \bar{z})| \leq \varepsilon$ for all $z \in S^1$. Taking $z = e^{i\theta}$ then implies

$$\text{sup}_{\theta} |f(\theta) - p(e^{i\theta}, e^{-i\theta})| \leq \varepsilon$$

where

$$p(e^{i\theta}, e^{-i\theta}) = \sum_{0 \leq m, n \leq N} a_{m, n} e^{i(m-n)\theta}$$

is the desired trigonometric polynomial.

Exercise 5.20. Use Example 5.74 to show that any $2\pi$ - periodic continuous function, $g : \mathbb{R}^d \rightarrow \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

$$p(x) = \sum_{\lambda \in \Lambda} a_{\lambda} e^{i\lambda x}$$

where $\Lambda$ is any 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, $f : (S^1)^d \rightarrow \mathbb{C}$ such that $f(e^{ix_1}, \ldots, e^{ix_d}) = F(x)$ for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

Exercise 5.21. 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

$$\int_{0}^{2\pi} f(x) e^{inx} dx = 0$$

for all $n \in \mathbb{Z}$, show again that $f \equiv 0$. Hint: Use Exercise 5.20.
5.11 *Appendix: A Multi-dimensional Weierstrass Approximation Theorem

The following theorem is the multi-dimensional generalization of Theorem 5.69.

Theorem 5.76 (Weierstrass Approximation Theorem). Suppose that 
\[ K = [a_1, b_1] \times \cdots [a_d, b_d] \] 
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 \(f \in C([0,1], \mathbb{R})\).

Given \(x \in K\), let \(\{X_n = (X_1^n, \ldots, X_n^n)\}_{n=1}^{\infty}\) be i.i.d. random vectors with values in \(\mathbb{R}^d\) such that
\[ P(X_n = \eta) = \prod_{i=1}^{d} (1 - \eta_i) \eta_i^n \]
for all \(\eta = (\eta_1, \ldots, \eta_d) \in \{0,1\}^d\). Since each \(X^j_n\) is a Bernoulli random variable with \(P(X^j_n = 1) = x_j\), we know that
\[ \mathbb{E}X_n = x \] and \(\text{Var}(X^j_n) = x_j - x_j^2 = x_j(1 - x_j)\).

As usual let \(S_n := X_1 + \cdots + X_n \in \mathbb{R}^d\), then
\[ \mathbb{E}\left[\frac{S_n}{n}\right] = x \] and
\[ \mathbb{E}\left[\left\|\frac{S_n}{n} - x\right\|^2\right] = \frac{1}{n^2} \sum_{j=1}^{d} \mathbb{E}\left[\frac{S^j_n}{n} - x_j\right]^2 \]
\[ = \frac{1}{n^2} \sum_{j=1}^{d} \text{Var}\left(\frac{S^j_n}{n}\right) \]
\[ = \frac{1}{n^2} \sum_{j=1}^{d} x_j(1 - x_j) \leq \frac{d}{4n}. \]

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 \(\varepsilon > 0\),
\[ P\left(\left\|\frac{S_n}{n} - x\right\| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{d}{4n}. \]

Observe that
\[ p_n(x) := \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \]
\[ = \sum_{\eta \in \{0,1\}^d} f\left(\frac{\eta(1) + \cdots + \eta(n)}{n}\right) P(X_1 = \eta(1), \ldots, X_n = \eta(n)) \]
\[ = \sum_{\eta \in \{0,1\}^d} f\left(\frac{\eta(1) + \cdots + \eta(n)}{n}\right) \prod_{k=1}^{d} (1 - x_k)^{1 - \eta_k} x_k^{\eta_k} \]
(5.66)
is a polynomial in \(x\) of degree \(nd\) at most. If we further let \(M = \sup\{|f(x)| : x \in K\}\), and
\[ \delta_x = \sup\{|f(y) - f(x)| : x, y \in K \text{ and } \|y - x\| \leq \varepsilon\}, \]
then
\[ |f(x) - p_n(x)| \leq \mathbb{E}\left[|f(x) - f\left(\frac{S_n}{n}\right)|\right] \leq \mathbb{E}\left[|f(x) - f\left(\frac{S_n}{n}\right)| : \|S_n - x\| > \varepsilon\right] \]
\[ + \mathbb{E}\left[|f(x) - f\left(\frac{S_n}{n}\right)| : \|S_n - x\| \leq \varepsilon\right] \]
\[ \leq 2M P(\|S_n - x\| > \varepsilon) + \delta_x \leq \frac{2dM}{n\varepsilon^2} + \delta_x. \] (5.67)

Therefore,
\[ \limsup_{n \to \infty} \max_{x \in K} |f(x) - p_n(x)| \leq \delta_x \to 0 \text{ as } \varepsilon \downarrow 0 \]
wherein we have used the uniform continuity of \(f\) to guarantee \(\delta_x \downarrow 0\) as \(\varepsilon \downarrow 0\).

Remark 5.77. We can write out the expression in Eq. (5.66) more explicitly by observing that \(\{S_n^1, \ldots, S_n^d\}\) are independent random variables with
\[ \mathbb{P}_x\left(S_n^i = k\right) = \binom{n}{k} x_i^k (1 - x_i)^{n-k} \text{ for } 0 \leq k \leq n \]
and therefore
\[ p_n(x) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \mathbb{E}\left[f\left(\frac{S_n^1, \ldots, S_n^d}{n}\right)\right] \]
\[ = \sum_{k \in \{0, \ldots, n\}^d} f\left(\frac{k}{n}\right) \prod_{i=1}^{d} \binom{n}{k_i} x_i^{k_i} (1 - x_i)^{n-k_i}. \]
Theorem 5.73 has the following multi-dimensional generalization.

**Theorem 5.78 (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

$$q(z, \bar{z}) = p\left(\frac{z + \bar{z}}{2}, \frac{z - \bar{z}}{2i}\right).$$

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 5.76.

**Example 5.79.** Let

$$(S^1)^d = \{z = (z_1, \ldots, z_d) \in \mathbb{C}^d : |z_j| = 1 \; 1 \leq j \leq d\} \subset \mathbb{C}^d$$

$A$ be the set of polynomials in $(z, \bar{z})$ restricted to $(S^1)^d$. Then $A$ is dense in $C((S^1)^d)$. To prove this first observe if $f \in C((S^1)^d)$ then

$$F(z) = |z_1| \ldots |z_d| f\left(\frac{z_1}{|z_1|}, \ldots, \frac{z_d}{|z_d|}\right) \text{ when } z_i \neq 0 \text{ for all } i$$
and $F(z) = 0$ if $z_i = 0$ for some $i$ defines a continuous function on $\mathbb{C}^d$ such that $F|_K = f$. By applying Theorem 5.78 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 $z$ and $z^{-1} := (z_1^{-1}, \ldots, z_d^{-1})$ are dense in $C((S^1)^d)$. 


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 \((\Omega, \mathcal{A}, \mu)\) 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.

The main construction theorem of this chapter is Theorem 6.20 which in part states that if \((\Omega, \mathcal{A}, \mu)\) is a premeasure space such that \( \mu(\Omega) < \infty \), then there exists a unique \( \sigma \)-additive measure, \( \bar{\mu} : \sigma(\mathcal{A}) \to [0, \mu(\Omega)] \), which extends \( \mu \). Before going on to Theorem 6.20 and its proof let us first give some equivalent conditions for checking when a finitely additive measure is in fact a premeasure.

**Proposition 6.3 (Equivalent premeasure conditions).** Suppose that \( \mu \) is a finitely additive measure on an algebra, \( \mathcal{A} \subset 2^\Omega \). Then the following are equivalent:

1. \( \mu \) is \( \sigma \)-additive 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, \( A_n = \bigcup_{k=1}^n A_k' \) and \( A = \bigcup_{k=1}^\infty A_k' \).

Therefore,

\[
\mu(A) = \sum_{k=1}^\infty \mu(A_k') = \lim_{n \to \infty} \sum_{k=1}^n \mu(A_k') = \lim_{n \to \infty} \mu(\bigcup_{k=1}^n A_k') = \lim_{n \to \infty} \mu(A_n).
\]

3. \( \implies \) 2. If \( \{A_n\}_{n=1}^\infty \) are disjoint and \( A := \bigcup_{n=1}^\infty A_n \in \mathcal{A} \), then \( \bigcup_{n=1}^N A_n \uparrow A \). Therefore,

\[
\mu(A) = \lim_{N \to \infty} \mu(\bigcup_{n=1}^N A_n) = \lim_{N \to \infty} \sum_{n=1}^N \mu(A_n) = \sum_{n=1}^\infty \mu(A_n).
\]

We now further assume that \( \mu(\Omega) < \infty \) which allows us to use \( \mu(B \setminus A) = \mu(B) - \mu(A) \) for all \( A, B \in \mathcal{A} \) with \( A \subseteq B \).

3. \( \implies \) 4. If \( A_n \in \mathcal{A} \) such that \( A_n \downarrow A \in \mathcal{A} \), then \( A_n' \uparrow A' \) and therefore,

\[
\lim_{n \to \infty} (\mu(\Omega) - \mu(A_n)) = \lim_{n \to \infty} \mu(A_n') = \mu(A') - \mu(A).
\]

4. \( \implies \) 3. If \( A_n \in \mathcal{A} \) such that \( A_n \uparrow A \in \mathcal{A} \), then \( A_n' \downarrow A' \) and therefore we again have

\[
\lim_{n \to \infty} (\mu(\Omega) - \mu(A_n)) = \lim_{n \to \infty} \mu(A_n') = \mu(A') = \mu(\Omega) - \mu(A).
\]

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

\[
\lim_{n \to \infty} [\mu(A) - \mu(A_n)] = \lim_{n \to \infty} \mu(A \setminus A_n) = 0.
\]

The following result is another useful result for verifying a finitely additive measure is a premeasure.

**Proposition 6.4.** 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
\[
E_j = A \cap E_j = \bigcap_{n=1}^{\infty} A_n \cap E_j = \bigcup_{i=1}^{\infty} E_{n,i} \cap E_j.
\]
By the assumed sub-additivity of \( \mu \) on \( \mathcal{S} \),
\[
\mu(E_j) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j).
\]
Summing this equation on \( j \) and using the finite additivity of \( \mu \) shows
\[
\mu(A) = \sum_{j=1}^{k} \mu(E_j) \leq \sum_{j=1}^{k} \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j)
= \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \sum_{j=1}^{k} \mu(E_{n,i} \cap E_j) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i}) = \sum_{n=1}^{\infty} \mu(A_n).
\]

We will give explicit examples of premeasure spaces and hence measure spaces after we have developed Theorem 4.20. The notion of an “outer measure” (which is the topic of the next section) is an intermediate step in proving the main measure construction theorems.

### 6.1 Outer Measures

**Definition 6.5 (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. \( \nu(A) \leq \nu(B) \) whenever \( A \subset B \)) and
3. \( \nu \) is countably sub-additive (i.e. \( \nu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \nu(A_n) \) for all \( \{A_n\}_{n=1}^{\infty} \subset 2^\Omega \)).

**Proposition 6.6 (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
\[
\rho^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \rho(E_k) : A \subset \bigcup_{k=1}^{\infty} E_k \text{ with } E_k \in \mathcal{E} \right\}.
\]
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
\[
\sum_{k=1}^{\infty} \rho(E_{nk}) \leq \rho^*(A_n) + 2^{-n} \varepsilon.
\]
Using these inequalities and the fact that \( \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n,k=1}^{\infty} E_{nk} \), it follows that
\[
\rho^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \rho(E_{nk}) \leq \sum_{n=1}^{\infty} \left[ \rho^*(A_n) + 2^{-n} \varepsilon \right] = \sum_{n=1}^{\infty} \rho^*(A_n) + \varepsilon.
\]

The sub-additivity of \( \rho^* \) then follows from this inequality as \( \varepsilon > 0 \) was arbitrary.

It is important that we allowed for countable \( \mathcal{E} \)-covers of \( A \) in the definition of \( \rho^*(A) \). If we had defined
\[
\rho^*(A) = \inf \left\{ \sum_{k=1}^{N} \rho(E_k) : A \subset \bigcup_{k=1}^{N} E_k \text{ with } E_k \in \mathcal{E} \text{ and } N \in \mathbb{N} \right\}
\]
we could not in general show that \( \rho^* \) is countably subadditive but only finitely subadditive. Countable subadditivity is key to all of the development to follow.

**Definition 6.7 (The outer metric).** If \( \nu \) is an outer measure on \( \Omega \), the associate outer pseudo-metric \( (d_\nu) \) on \( 2^\Omega \) is defined by
\[
d_\nu(A,B) := \nu(A \triangle B) \in [0, \infty] \text{ } \forall A,B \in 2^\Omega.
\]

**Proposition 6.8 (Properties of the outer metric).** If \( \nu : 2^\Omega \to [0, \infty] \) is an outer measure and \( d = d_\nu \) as in Definition 6.7, then;
1. \( d \) is a pseudo metric, i.e. is symmetric and satisfies the triangle inequality.
2. \( (A^c,C^c) = d(A,C) \) for all \( A,B \in 2^\Omega \).
3. If \( \{A_n\}_{n=1}^{\infty}, \{B_n\}_{n=1}^{\infty} \subset 2^\Omega \), then
\[
d(\bigcup_{n=1}^{\infty} A_n, \bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} d(A_n, B_n) \text{ and } \quad (6.2)
\]
\[
d(\bigcap_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} d(A_n, B_n). \quad (6.3)
\]

**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 $A, B, C \in \mathcal{P}_0$, then

$$1_{A \triangle C} = |1_A - 1_C| = |1_A - 1_B| + |1_B - 1_C| = 1_{A \triangle B} + 1_{B \triangle C}.$$ 

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,

$$d(A, C) = \nu(A \triangle C) \leq \nu([A \triangle B] \cup [B \triangle C])$$

$$\leq \nu(A \triangle B) + \nu(B \triangle C) = d(A, B) + d(B, C).$$

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 verified 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

$$\bigcup_{n=1}^{\infty} A_n \triangle \bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} [A_n \triangle B_n].$$ (6.4)

Hence using the monotonicity and finite subadditivity of $\nu$ then gives;

$$d(\bigcup_{n=1}^{\infty} A_n, \bigcup_{n=1}^{\infty} B_n) = \nu(\bigcup_{n=1}^{\infty} A_n \triangle \bigcup_{n=1}^{\infty} B_n) \leq \nu(\bigcup_{n=1}^{\infty} [A_n \triangle B_n])$$

$$\leq \sum_{n=1}^{\infty} \nu(A_n \triangle B_n) = \sum_{n=1}^{\infty} d(A_n, B_n),$$

which proves Eq. (6.2). Equation (6.3) may be proved similarly or by combining item 2. with Eq. (6.2) as follows;

$$d(\bigcap_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} B_n)$$

$$= d(\bigcap_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} B_n)$$

$$= d(\bigcap_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} d(A_n, B_n).$$


6.2 Construction of $\sigma$–Additive Finite Measures

The most important special case of the construction in Proposition 6.6 is when $\mathcal{E} = \mathcal{A}$ is an algebra and $\rho = \mu$ is a premeasure on $\mathcal{A}$. The next two relatively easy results will play an important role in this section.

**Lemma 6.9 (Outer measures dominate).** Suppose that $(\Omega, \mathcal{A}, \mu)$ is a premeasure space and $\nu: \sigma(\mathcal{A}) \to [0, \infty]$ is a $\sigma$–additive measure such that $\nu = \mu$ on $\mathcal{A}$, then

$$\nu(B) \leq \mu^*(B) \text{ for all } B \in \sigma(\mathcal{A}).$$

**Proof.** If $B \in \sigma(\mathcal{A})$ and $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ is a cover of $B$ (i.e. $B \subset \bigcup_{n=1}^{\infty} A_n$), then

$$\nu(B) \leq \sum_{n=1}^{\infty} \nu(A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

As this inequality holds for every $\mathcal{A}$-cover of $B$, it follows that $\nu(B) \leq \mu^*(B).$

**Proposition 6.10.** If $(\Omega, \mathcal{A}, \mu)$ is a finitely additive measure space then $\mu^* \leq \mu$ on $\mathcal{A}$. Moreover $\mu^* = \mu$ on $\mathcal{A}$ iff $\mu$ is a premeasure.

**Proof.** If $A \in \mathcal{A}$, then $A \subset \bigcup_{n=1}^{\infty} A_n$ where $A_1 = A$ and $A_k = \emptyset$ for all $k \geq 2$ and therefore,

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \mu(A_n) = \mu(A).$$

Hence $\mu^* \leq \mu$ on $\mathcal{A}$.

Now suppose $\mu$ is a premeasure on $\mathcal{A}$ and that $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ is a given cover of $A$, i.e. $A = \bigcup_{n=1}^{\infty} A_n$. Let $A_0 = \emptyset$ and defined

$$\tilde{A}_n := A \cap A_n \setminus [A_1 \cup \cdots \cup A_{n-1}] \in \mathcal{A} \text{ for all } n \in \mathbb{N}.$$ We then have $\tilde{A}_n \subset A_n$ and $A = \sum_{n=1}^{\infty} \tilde{A}_n = \bigcup_{n=1}^{\infty} A_n$ and therefore

$$\mu(A) = \sum_{n=1}^{\infty} \mu(\tilde{A}_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

As $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ was an arbitrary cover of $A$, it follows from the previous equation that $\mu(A) \leq \mu^*(A)$ and hence $\mu = \mu^*$ on $\mathcal{A}$.

Lastly let us suppose that $\mu = \mu^*$ on $\mathcal{A}$. If $\{A\} \cup \{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ are such that $A = \sum_{n=1}^{\infty} A_n$, then by definition of $\mu^*$ and the assumption that $\mu = \mu^*$ on $\mathcal{A}$ we find,

$$\mu(A) = \mu^*(A) \leq \sum_{n=1}^{\infty} \mu(A_n) = \lim_{N \to \infty} \sum_{n=1}^{N} \mu(A_n) = \lim_{N \to \infty} \mu \left( \sum_{n=1}^{N} A_n \right) \leq \mu(A),$$

$$\therefore$$
wherein we have used the finite additivity and monotonicity of \( \mu \). The previous equation clearly implies \( \mu (A) = \sum_{n=1}^{\infty} \mu (A_n) \) which shows \( \mu \) is a premeasure on \( \mathcal{A} \).

**Corollary 6.11.** Suppose that \((\Omega, \mathcal{A}, \mu)\) is a premeasure space such that \( \mu (\Omega) < \infty \) and let \( d(A, B) = \mu^* (A \triangle B) \) for all \( A, B \in 2^\Omega \). Then \( \mu \) is Lip-1 relative to \( d \). Hence, Lemma 2.28 \( \mu \) has a unique Lip-1 extension, \( \bar{\mu} \), to \( \mathcal{B} = \mathcal{B} (\mu) := \mathcal{A}^d \).

**Summary:** A subset, \( B \in \mathcal{A} \), is in \( \mathcal{B} \) iff there exists \( \{A_n\} \subset \mathcal{A} \) such that \( \lim_{n \to \infty} \mu^* (A_n \triangle B) = 0 \) and in which case \( \bar{\mu} (B) := \lim_{n \to \infty} \mu (A_n) = \mu^* (B) \).

**Proof.** Let \( A, B \in \mathcal{A} \). By the basic properties of the simple integral and the fact that \( \mu = \mu^* \) on \( \mathcal{A} \) we find,

\[
|\mu (B) - \mu (A)| = \left| \int_{\Omega} (1_B - 1_A) \, d\mu \right| \\
\leq \int_{\Omega} |1_B - 1_A| \, d\mu = \mu (B \triangle A) = \mu^* (B \Delta A) = d(A, B).
\]

**Lemma 6.13.** 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.** By construction \( \mathcal{A}_\sigma \) is closed under countable unions. Moreover if \( A = \bigcup_{i=1}^{\infty} A_i \) and \( B = \bigcup_{j=1}^{\infty} B_j \) with \( A_i, B_j \in \mathcal{A} \), then

\[
A \cap B = \bigcup_{i,j=1}^{\infty} A_i \cap B_j \in \mathcal{A}_\sigma,
\]

which shows that \( \mathcal{A}_\sigma \) is also closed under finite intersections.

2. If \( A_i \in \mathcal{A} \) and \( A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}_\sigma \), then \( A^c = \bigcap_{i=1}^{\infty} A_i^c \in \mathcal{A}_\delta \) and visa versa.

3. This item follows directly from items 1. and 2. or may be proved directly.

**Proposition 6.14.** Suppose that \((\Omega, \mathcal{A}, \mu)\) is a premeasure space. If \( A \in \mathcal{A}_\sigma \) and \( \{A_n\}_{n=1}^{\infty} \subset \mathcal{A} \) are such that \( A_n \uparrow A \), then

\[
\mu^* (A) = \lim_{n \to \infty} \mu (A_n) = \sum_{n=1}^{\infty} \mu (A_n \setminus A_{n-1}).
\] (6.5)

Moreover, if we further assume that \( \mu (\Omega) < \infty \), then

\[
d(A_n, A) = \mu^* (A) - \mu (A_n) \to 0 \text{ as } n \to \infty
\]

and in particular \( \mathcal{B} = \mathcal{A} = \mathcal{A}_\sigma \) and

\[
\bar{\mu} (A) = \lim_{n \to \infty} \mu (A_n) = \mu^* (A).
\] (6.6)

**Proof.** Since \( A = \bigcup_{n=1}^{\infty} [A_n \setminus A_{n-1}] \), by the definition of \( \mu^* (A) \) it follows that

\[
\mu^* (A) \leq \sum_{n=1}^{\infty} \mu (A_n \setminus A_{n-1}) = \lim_{N \to \infty} \sum_{n=1}^{N} \mu (A_n \setminus A_{n-1})
\]

\[
= \lim_{N \to \infty} \mu \left( \sum_{n=1}^{N} [A_n \setminus A_{n-1}] \right) = \lim_{N \to \infty} \mu (A_N) \leq \mu^* (A),
\]

wherein the last inequality we have used \( \mu (A_N) = \mu^* (A) \leq \mu^* (A) \) since that \( \mathcal{A} \ni A_N \subset \mathcal{A} \). For the second assertion observe that for each \( n \in \mathbb{N} \),

\[
A \ni A_N \setminus A \uparrow A \setminus A_n \text{ as } N \uparrow \infty.
\]

So by what we have just proved while using \( \mu (\Omega) < \infty \) so that \( \mu (A_N) < \infty \) we find,

\[
d(A_n, A) = \mu^* (A \setminus A_n) = \lim_{N \to \infty} \mu (A_N \setminus A_n)
\]

\[
= \lim_{N \to \infty} \mu (A_N) \setminus \mu (A_n) = \mu^* (A) - \mu (A_n)
\]

where the last expression goes to zero as \( n \to \infty \) since the first assertion implies, \( \mu (A_n) \uparrow \mu^* (A) \) as \( n \to \infty \). This in particular shows that \( \mathcal{A}_\sigma \subset \mathcal{A} \) and as \( \mathcal{A} \subset \mathcal{A}_\sigma \) we may conclude that \( \mathcal{A} \subset \mathcal{A}_\sigma \subset \mathcal{A} \), i.e. \( \mathcal{B} = \mathcal{A} = \mathcal{A}_\sigma \).

**Proposition 6.15.** If \((\Omega, \mathcal{A}, \mu)\) is a finite (\( \mu (\Omega) < \infty \)) premeasure space, then \( \mathcal{B} = \mathcal{A} \subset 2^\Omega \) is a \( \sigma \)-algebra.

**Proof.** Since \( A \subset B \), we have \( 0, \Omega \in B \). If \( B \in \mathcal{B} \), there exists \( A_n \in \mathcal{A} \) so that \( \lim_{n \to \infty} d(A_n, B) = 0 \) and so by Proposition 6.8

\[
\lim_{n \to \infty} d(A_n^c, B^c) = \lim_{n \to \infty} d(A_n, B) = 0
\]

which shows \( B^c \in \mathcal{B} \).
Now suppose that \( \{B_n\}_{n=1}^{\infty} \subset B \) and \( B := \bigcup_{n=1}^{\infty} B_n \). Given \( \varepsilon > 0 \) we may find \( A_n \in \mathcal{A} \) so that \( d(A_n, B_n) \leq \varepsilon 2^{-n} \) for all \( n \). With \( A := \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_\sigma \), it follows (using Proposition 6.8) that
\[
d(A, B) = d(\bigcup_{n=1}^{\infty} A_n, \bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} d(A_n, B_n) \leq \varepsilon.
\] (6.7)

As \( \varepsilon > 0 \), this shows \( B \subset \bar{\mathcal{A}_\sigma} = \bar{\mathcal{A}}^d = B \).

**Remark 6.16.** If \( (\Omega, \mathcal{A}, \mu) \) is a premeasure space, \( 0 < M < \infty \), and \( B \subset \Omega \), then \( \mu^*(B) < M \) iff there exists \( D := \sum_{j=1}^{\infty} D_j \) such that \( D_j \in \mathcal{A}, B \subset D \), and
\[
\mu^*(D) = \sum_{j=1}^{\infty} \mu(D_j) < M.
\]

**Proposition 6.17.** Let \( (\Omega, \mathcal{A}, \mu) \) be a premeasure space with \( \mu(\Omega) < \infty \) and \( B \in \mathcal{A}^d \). Then the following are equivalent;

1. 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 \( \mu^*(C \setminus A) < \varepsilon \).
2. For all \( \varepsilon > 0 \) there exists \( A \in \mathcal{A}_\sigma \) such that \( B \subset C \) and \( \mu^*(C \setminus B) < \varepsilon \).
3. The set \( B \) is in \( \mathcal{A} = \mathcal{B}(\mu) \).

**Proof.** (1 \( \implies \) 2) This implication simply follows by monotonicity of \( \mu^* \) which implies \( \mu^*(C \setminus B) \leq \mu^*(C \setminus A) < \varepsilon \).

(2 \( \implies \) 3) Assuming 2 we may find \( C_n \in \mathcal{A}_\sigma \) so that \( B \subset C_n \) and \( d(B, C_n) = \mu^*(C_n \setminus B) < 1/n \to 0 \) as \( n \to \infty \). This shows that \( B \subset \bar{\mathcal{A}_\sigma} = B \).

(3 \( \implies \) 1) Given if \( B \in \mathcal{B} \), then for all \( \varepsilon > 0 \) there exists \( G_\varepsilon \in \mathcal{A}_\sigma \) so that \( \varepsilon > d(G_\varepsilon, B) = \mu^*(B \Delta G_\varepsilon) \) and so by Remark 6.16 there exists \( D_\varepsilon \subset \mathcal{A}_\sigma \) such that \( \mu^*(D_\varepsilon) < \varepsilon \) and \( B \Delta G_\varepsilon \subset D_\varepsilon \). We now let
\[
\mathcal{A}_\delta \ni A_\varepsilon := G_\varepsilon \setminus D_\varepsilon \subset C_\varepsilon \setminus [B \Delta G_\varepsilon] = B \cap G_\varepsilon \subset B
\]
and
\[
\mathcal{A}_\sigma \ni C_\varepsilon := G_\varepsilon \cup D_\varepsilon \supset G_\varepsilon \cup [B \Delta G_\varepsilon] = B \cup G_\varepsilon \supset B.
\]
Thus \( A_\varepsilon \subset B \subset C_\varepsilon \) and the proof is complete since \( \mu^*(C_\varepsilon \setminus B_\varepsilon) = \mu^*(D_\varepsilon) < \varepsilon \) because,
\[
C_\varepsilon \setminus B_\varepsilon = [G_\varepsilon \cup D_\varepsilon] \setminus [G_\varepsilon \setminus D_\varepsilon] = [(G_\varepsilon \setminus D_\varepsilon) \cup D_\varepsilon] \setminus [G_\varepsilon \setminus D_\varepsilon] = D_\varepsilon.
\]

**Proposition 6.18.** Let \( (\Omega, \mathcal{A}, \mu) \) be a premeasure space with \( \mu(\Omega) < \infty \), then \( \bar{\mu} = \mu^* \) on \( B \).

**Proof.** Let \( B \in \mathcal{B} \) and \( \varepsilon > 0 \) be given and choose \( C \in \mathcal{A}_\sigma \) so that \( B \subset C \) and \( d(B, C) = \mu^*(C \setminus B) < \varepsilon \). Using the fact that \( \bar{\mu} \) is Lip-1 on \( B \) we have
\[
\bar{\mu}(B) = |\bar{\mu}(B) - \bar{\mu}(\emptyset)| \leq d(B, \emptyset) = \mu^*(B), \quad \text{and}
\]
\[
|\bar{\mu}(B) - \bar{\mu}(C)| \leq d(B, C) < \varepsilon.
\]

These inequalities along with Eq. (6.6) then imply,
\[
\bar{\mu}(B) \leq \mu^*(B) \leq \mu^*(C) = \bar{\mu}(C) < \bar{\mu}(B) + \varepsilon.
\]

As \( \varepsilon > 0 \) was arbitrary we conclude that \( \bar{\mu}(B) = \mu^*(B) \).

**Proposition 6.19.** If \( (\Omega, \mathcal{A}, \mu) \) is a finite premeasure space, then \( \bar{\mu} \) defined in Theorem 6.20 is a \( \sigma \)-additive measure on the \( \sigma \)-algebra, \( \mathcal{B} = \bar{\mathcal{A}}^d \).

**Proof.** For \( A, B \in \mathcal{B} \), choose \( A_n, B_n \in \mathcal{A} \) so that \( A_n \xrightarrow{d} A \) and \( B_n \xrightarrow{d} B \) as \( n \to \infty \). By Proposition 6.8 it follows that
\[
A_n \cup B_n \xrightarrow{d} A \cup B \quad \text{and} \quad A_n \cap B_n \xrightarrow{d} A \cap B \quad \text{as} \quad n \to \infty.
\]

Therefore using the definition of \( \bar{\mu} \) and the finite additivity of \( \mu \), it follows that
\[
\bar{\mu}(A \cup B) + \bar{\mu}(A \cap B) = \lim_{n \to \infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)]
\]
\[
= \lim_{n \to \infty} [\mu(A_n) + \mu(B_n)]
\]
\[
= \bar{\mu}(A) + \bar{\mu}(B).
\]

This shows that \( \bar{\mu} \) is finitely additive. Since, by Propositions 6.18 \( \bar{\mu} = \mu^* \) on \( \mathcal{B} \), it follows that \( \bar{\mu} \) is sub-additive on \( \mathcal{B} \) and hence \( \bar{\mu} \) is countably additive, see Proposition 5.2.

The next theorem summarizes most of the results we have proved in this section.

**Theorem 6.20 (Finite premeasure extension theorem).** If \( (\Omega, \mathcal{A}, \mu) \) is a finite premeasure space, \( \mathcal{B} = \bar{\mathcal{A}}^d \subset \mathcal{A}^d \), and \( \bar{\mu} \) is the unique \( d \)-continuous extension of \( \mu \) from \( \mathcal{A} \) to \( \mathcal{B} \), then;

1. \( |\bar{\mu}(B) - \bar{\mu}(A)| \leq d(A, B) = \mu^*(A \Delta B) \) for all \( A, B \in \mathcal{B} \).
2. \( \mathcal{B} = \bar{\mathcal{A}}^d \) is a \( \sigma \)-algebra which necessarily contains \( \mathcal{A} \) and hence \( \sigma(\mathcal{A}) \).
3. For \( B \in \mathcal{A}^d \), the following are equivalent;
   a) 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 \( \mu^*(C \setminus A) < \varepsilon \).
   b) For all \( \varepsilon > 0 \) there exists \( A \in \mathcal{A}_\sigma \) such that \( B \subset C \) and \( \mu^*(C \setminus B) < \varepsilon \).
6 Countably Additive Measures

c) The set \( B \) is in \( \mathcal{B} = \mathcal{B}(\mu) \), i.e. for \( \varepsilon > 0 \) there exists \( A \in \mathcal{A} \) such that \( \mu^*(B \Delta A) < \varepsilon \).

4. \( \bar{\mu} = \mu^* \) on \( \mathcal{B} \) so that

\[
\bar{\mu}(B) = \mu^*(B)
= \inf \left\{ \mu^*(C) : B \subset C \subset A \right\}.
\]

5. \( \bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)] \) is a \( \sigma \)-additive measure, i.e. \((\Omega, \mathcal{B}, \bar{\mu})\) is a measure space!

6. If \( \nu \) is any measure on \( \sigma(\mathcal{A}) \) such that \( \nu = \mu \) on \( \mathcal{A} \), then \( \nu = \bar{\mu} \) on \( \sigma(\mathcal{A}) \).

**Proof.** The only thing left to prove is the uniqueness assertion. If \( \nu \) is a measure on \( \sigma(\mathcal{A}) \) such that \( \nu = \mu \) on \( \mathcal{A} \), then from Lemma 6.9 we know that \( |\nu(B_1) - \nu(B_2)| \leq \mu^*(B) \) for all \( B \in \sigma(\mathcal{A}) \). Consequently if \( B_1, B_2 \in \sigma(\mathcal{A}) \), then

\[
|\nu(B_1) - \nu(B_2)| = |\mathbb{E}_n[1_{B_1} - 1_{B_2}]| \leq \mathbb{E}_n[1_{B_1} - 1_{B_2}] = \mathbb{E}_n[1_{B_1} \Delta B_2] = \mu^*(B_1 \Delta B_2) = d_{\mu^*}(B_1, B_2).
\]

Thus \( \nu \) is a Lip -1 continuous and hence \( \nu = \bar{\mu} \) on \( \sigma(\mathcal{A}) \subset \mathcal{B} \) as \( \mathcal{A} \) is \( d_{\mu^*} \) - dense in \( \sigma(\mathcal{A}) \).

**Exercise 6.1.** Suppose \( \mathcal{A} \subset 2^\Omega \) is an algebra and \( \mu \) and \( \nu \) are two finite \( \sigma \)-additive measures on \( \mathcal{B} = \sigma(\mathcal{A}) \) such that \( \mu = \nu \) on \( \mathcal{A} \). Show \( \mu = \nu \) on \( \mathcal{B} \).

**Exercise 6.2.** Let \( \mu, \bar{\mu}, \mathcal{A} \) and \( B := \mathcal{B}(\mu) \) be as in Theorem 6.20. Further suppose that \( B_0 \subset 2^\Omega \) is a \( \sigma \)-algebra such that \( \mathcal{A} \subset B_0 \subset \mathcal{B} \) and \( \nu : B_0 \to [0, \mu(\Omega)] \) is a \( \sigma \)-additive measure on \( B_0 \) such that \( \nu = \mu \) on \( \mathcal{A} \). Show that \( \nu = \bar{\mu} \) on \( B_0 \) as well.

6.3 A Discrete Kolmogorov’s Extension Theorem

In this section we continue the setup in Section 3.2. In detail, let \( S \) be a non-empty finite or countable state space, \( \Omega := S^\infty := S^N \) (path or configuration space),

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

and \( \mathcal{A} := \bigcup_{n=1}^\infty \mathcal{A}_n \).

The next result is a key compactness result we will use in this section.

**Theorem 6.21 (A finite intersection property I).** Let \( S \) be an arbitrary set, \( \Omega := S^\infty \), \( \{N_n\}_{n=1}^\infty \subset \mathbb{N} \) be an increasing sequence such that \( \lim_{n \to \infty} N_n = \infty \), and \( \tilde{B}_n = B_n \times \Omega \) with \( B_n \subset S^{N_n} \) for all \( n \in \mathbb{N} \). If \( \{\tilde{B}_n\}_{n=1}^\infty \) has the finite intersection property, i.e. \( \cap_{n=1}^M \tilde{B}_n \neq \emptyset \) for all \( M < \infty \), then \( \cap_{n=1}^\infty \tilde{B}_n \neq \emptyset \).

**Proof.** Let \( \pi_n : \Omega \to S^{N_n} \) be the projection map onto the first \( N_n \) - coordinates,

\[
\pi_n(\omega) = (\omega_1, \omega_2, \ldots, \omega_{N_n}) \in S^{N_n} \forall \omega \in \Omega.
\]

We will abuse notation and also use \( \pi_n \) to denote the projection from \( S^M \to S^{N_n} \) for all \( M \geq N_n \). By replacing \( \tilde{B}_n \) by \( \tilde{B}_1 \cap \cdots \cap \tilde{B}_n \) if necessary, we may assume that \( \{\tilde{B}_n = B_n \times \Omega\}_{n=1}^\infty \) are non-increasing non empty sets where again \( \emptyset \neq B_n \subset S^{N_n} \).

For each \( n \in \mathbb{N} \) choose \( \omega(n) \in \tilde{B}_n = \pi_n^{-1}(\tilde{B}_n) \). Then using \( \tilde{B}_n \subset \tilde{B}_1 \) for all \( n \) we have,

\[
\pi_1(\omega(n)) = \pi_1(\tilde{B}_n) \subset \pi_1(\tilde{B}_1) = \tilde{B}_1 \forall n \in \mathbb{N}.
\]

Let \( \Gamma_0 := \mathbb{N} \). As \( B_1 \) is a finite set, there exists \( \lambda_1 \in B_1 \) such that

\[
\Gamma_1 = \{ n \in \mathbb{N} = \leq_{\mathbb{N}} : \pi_1(\omega(n)) = \lambda_1 \}
\]

is an infinite set. Similarly for \( n \in \Gamma_1 \) with \( n \geq 2 \),

\[
\pi_2(\omega(n)) = \pi_2(\tilde{B}_n) \subset \pi_2(\tilde{B}_2) = B_2
\]

and since \( B_2 \) is a finite set there exists \( \lambda_2 \in B_2 \) such that

\[
\Gamma_2 = \{ n \in \Gamma_1 : n \geq 2 \text{ and } \pi_2(\omega(n)) = \lambda_2 \}
\]

is an infinite set. Continuing this way inductively we produce for all \( k \in \mathbb{N} \), an \( \lambda_k \in B_k \) such that

\[
\Gamma_k = \{ n \in \Gamma_{k-1} : n \geq k \text{ and } \pi_k(\omega(n)) = \lambda_k \} \subset \Gamma_{k-1}
\]

is an infinite set. By construction, if \( k > l \geq 1 \) and \( n \in \Gamma_k \subset \Gamma_l \), then

\[
\lambda_l = \pi_l(\omega(n)) = \pi_l \circ \pi_k(\omega(n)) = \pi_l(\lambda_k).
\]

From this “consistency relation,” it follows that there exists a unique point \( \lambda \in \Omega \) such that \( \pi_l(\lambda) = \lambda_l \) for all \( l \in \mathbb{N} \). Since

\[
\pi_l(\lambda) = \lambda_l = \pi_l(\omega(n)) \in B_l \text{ for all } n \in \Gamma_l
\]

it follows that \( \lambda \in \tilde{B}_l \) for all \( l \) and therefore \( \lambda \in \cap_{l=1}^{\infty} \tilde{B}_l \) which shows \( \cap_{n=1}^\infty \tilde{B}_n \neq \emptyset \).

**For later purposes, let us record a jazzed up version of the last result.**

**Theorem 6.22 (A finite intersection property II).** Let \((S, \rho)\) be a metric space, \( \Omega := S^N \), \( \{N_n\}_{n=1}^\infty \subset \mathbb{N} \) be an increasing sequence such that \( \lim_{n \to \infty} N_n = \infty \), and \( \tilde{B}_n = B_n \times \Omega \) with \( B_n \) is a compact subset of \( S^{N_n} \) for all \( n \in \mathbb{N} \). If \( \{\tilde{B}_n\}_{n=1}^\infty \) has the finite intersection property, i.e. \( \cap_{n=1}^M \tilde{B}_n \neq \emptyset \) for all \( M < \infty \), then \( \cap_{n=1}^\infty \tilde{B}_n \neq \emptyset \).
Proof. The proof is very similar to what we have just done. We again let \( \pi_n : \Omega \rightarrow S_n \) be the projection map onto the first \( N_n \) – coordinates as above and again replace \( B_n \) by \( B_1 \cap \cdots \cap B_n \) if necessary so that we may assume that \( \{ B_n = B_n \times \Omega \}_{n=1}^\infty \) are non-increasing non-empty sets with \( B_n \) being a non-empty compact subset of \( S_n \). [As \( B_1 \cap \cdots \cap B_n = K \times \Omega \) where \( K = \cap_{k=1}^n \{ B_k \times S_n - N_n \} \) is a closed subset of \( B_n \times S_n \) and \( B_n \) was compact it follows that \( K \) is still compact.]

For each \( n \in \mathbb{N} \) choose \( \omega(n) \in B_n = \pi_n^{-1}(B_n) \). Then using \( \tilde{B}_n \subset B_n \) for all \( n \) we have,

\[
\pi_1(\omega(n)) \in \pi_1(\tilde{B}_n) \subset \pi_1(\tilde{B}_1) = B_1 \ \forall \ n \in \mathbb{N}.
\]

Let \( \Gamma_0 := \mathbb{N} \). As \( B_1 \) is a compact set, there exists \( \lambda_1 \in B_1 \) and an infinite subset, \( \Gamma_1 \subset N_n \), such that

\[
\pi_1(\omega(n)) \rightarrow \lambda_1 \text{ as } \Gamma_1 \ni n \rightarrow \infty.
\]

Similarly for \( n \in \Gamma_1 \) with \( n \geq 2 \),

\[
\pi_2(\omega(n)) \in \pi_2(\tilde{B}_n) \subset \pi_2(\tilde{B}_2) = B_2
\]

and since \( B_2 \) is a compact set there exists \( \lambda_2 \in B_2 \) and an infinite subset, \( \Gamma_2 \subset \Gamma_1 \), such that

\[
\pi_2(\omega(n)) \rightarrow \lambda_2 \text{ as } \Gamma_2 \ni n \rightarrow \infty.
\]

Continuing this way inductively we produce for all \( k \in \mathbb{N} \), a \( \lambda_k \in B_k \) and an infinite subset, \( \Gamma_k \subset \Gamma_{k-1} \), such that

\[
\pi_k(\omega(n)) \rightarrow \lambda_k \text{ as } \Gamma_k \ni n \rightarrow \infty.
\]

By construction, if \( k > l \geq 1 \) then \( \Gamma_k \subset \Gamma_l \) and

\[
\lambda_l = \lim_{l \ni n \rightarrow \infty} \pi_l(\omega(n)) = \lim_{l \ni n \rightarrow \infty} \pi_l(\omega(n))
\]

\[
= \lim_{k \ni n \rightarrow \infty} \pi_l \circ \pi_k(\omega(n)) = \pi_l(\lambda_k).
\]

From this “consistency relation,” it follows that there exists a unique point \( \lambda \in \Omega \) such that \( \pi_l(\lambda) = \lambda_l \) for all \( l \in \mathbb{N} \). Since

\[
\pi_l(\lambda) = \lambda_l = \pi_l(\omega(n)) \in B_l \text{ for all } n \in \Gamma_l
\]

it follows that \( \lambda \in \tilde{B}_l \) for all \( l \) and therefore \( \lambda \in \cap_{l=1}^\infty \tilde{B}_l \) which shows \( \cap_{l=1}^\infty \tilde{B}_l \neq \emptyset \).

Exercise 6.3. Suppose \( S \) is a finite set, \( \Omega = S^N \), and \( B := \sigma(A) \) as above. Show every sequence of probability measures \( \{P_n\}_{n=1}^\infty \) possesses a subsequence \( \{P_{n_k}\}_{k=1}^\infty \) such that \( \lim_{n_k \rightarrow \infty} P_{n_k}(A) = P(A) \) for all \( A \in A \) where \( P \) is a probability measure on \( (\Omega, B) \). Hints: 1) note that \( A \) is a countable collection of subsets of \( \Omega \), and 2) use Cantor’s diagonalization argument.

The next theorem extends this result to the setting where \( S \) is countably additive.

Theorem 6.24 (Kolmogorov’s Extension Theorem II). If \( P : A \rightarrow [0,1] \) is a finitely additive probability measure such that \( P|_{A_n} \) is a \( \sigma \)-additive measure for all \( n \in \mathbb{N} \) then \( P \) is a premeasure and hence extends uniquely to a \( \sigma \)-additive probability measure on \( \sigma(A) \), still denoted by \( P \).

Proof. Suppose that \( A_n \in A \) with \( A_n \downarrow \) and \( P(A_n) \geq \varepsilon \) for all \( n \in \mathbb{N} \). Choose \( N_n \in \mathbb{N} \) such that \( N_n \uparrow \infty \) as \( n \rightarrow \infty \) such that \( A_n = C_n \times \Omega \) for all \( n \in \mathbb{N} \) with \( C_n \subset S^{N_n} \) for all \( n \in \mathbb{N} \). Then choose finite sets \( B_n \subset C_n \) so that \( P(A_n \setminus B_n) \leq \frac{1}{2^2 - n} \) for all \( n \). We then have

\[
P(A_n \setminus B_1 \cap \cdots \cap B_n) \leq \sum_{j=1}^n P(A_j \setminus B_j) \leq \frac{1}{2^2} \sum_{j=1}^n 2^{-j}
\]

\[
\leq \frac{1}{4} \varepsilon \left( 2^{-1} \sum_{j=1}^n 2^{-j} \right) = \frac{1}{2} \varepsilon
\]

from which it follows that

\[
\varepsilon \leq P(A_n) = P(B_1 \cap \cdots \cap B_n) + P(A_n \setminus B_1 \cap \cdots \cap B_n)
\]

\[
\leq P(B_1 \cap \cdots \cap B_n) + \varepsilon/2
\]

so that \( P(B_1 \cap \cdots \cap B_n) \geq \varepsilon/2 \). In particular \( B_1 \cap \cdots \cap B_n \neq \emptyset \) for all \( n \in \mathbb{N} \) and therefore by Theorem 6.21

\[
\emptyset \neq \cap_{j=1}^\infty B_j \subset \cap_{j=1}^\infty A_j
\]

and the result is proved.
Corollary 6.25 (Kolmogorov’s Extension Theorem IIa). Probability measures, \( P \), on \((\Omega, \sigma(A))\) are in one to one correspondence with sequences of functions, \( p_n : S^n \to [0, 1] \) for \( n \in \mathbb{N} \) which satisfy the consistency conditions,

\[
\sum_{t \in S} p_n (s_1, \ldots, s_{n-1}, t) = p_{n-1} (s_1, \ldots, s_{n-1}) \quad \text{for all} \ n \in \mathbb{N} \quad (6.8)
\]

where by convention, \( p_0 := 1 \). The correspondence is determined by

\[
p_n (s_1, \ldots, s_n) := P \left( \{ \omega \in \Omega : \omega_1 = s_1, \ldots, \omega_n = s_n \} \right). \quad (6.9)
\]

**Proof.** If \( P \) is a probability measure on \((\Omega, \sigma(A))\) and \( p_n : S^n \to [0, 1] \) is defined as in Eq. (6.9), it is a simple matter to verify the consistency conditions in Eq. (6.8) hold using the countable additivity of \( P \). Conversely if \( p_n : S^n \to [0, 1] \) for \( n \in \mathbb{N} \) which satisfy the consistency conditions in Eq. (6.8), then you have shown in Exercise 5.3 that there exists a unique finitely additive probability measure, \( P, \) on \((\Omega, A)\) such that \( P|_{A_n} \) is \( \sigma \)-additive for each \( n \) and

\[
p_n (s_1, \ldots, s_n) := P \left( \{ \omega \in \Omega : \omega_1 = s_1, \ldots, \omega_n = s_n \} \right).
\]

By Theorem 6.24 this finitely additive measure extends uniquely to as \( \sigma \)-additive measure, still denote by \( P, \) on \( \sigma(A) \).

**Example 6.26 (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 \( \sigma(A) \) such that, for all \( n \in \mathbb{N} \) and \( (s_1, \ldots, s_n) \in S^n \), we have

\[
P \left( \{ \omega \in \Omega : \omega_1 = s_1, \ldots, \omega_n = s_n \} \right) = q(s_1) \cdots q(s_n).
\]

One need only define \( p_n (s_1, \ldots, s_n) := q(s_1) \cdots q(s_n) \) and note that

\[
\sum_{t \in S} p_n (s_1, \ldots, s_{n-1}, t) = q(s_1) \cdots q(s_{n-1}) \sum_{t \in S} q(t) = p_{n-1} (s_1, \ldots, s_{n-1}).
\]

Thus the claim follows directly from Corollary 6.25.

**Example 6.27 (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.

\[
\sum_{y \in S} p(x, y) = 1 \quad \text{for all} \ x \in S. \quad (6.10)
\]

Also let \( \pi : S \to [0, 1] \) be a probability function, i.e. \( \sum_{x \in S} \pi(x) = 1 \). We now take

\[
\Omega := S^{\mathbb{N}_0} = \{ \omega = (s_0, s_1, \ldots) : s_j \in S \}
\]

and let \( X_n : \Omega \to S \) be given by

\[
X_n (s_0, s_1, \ldots) = s_n \quad \text{for all} \ n \in \mathbb{N}_0.
\]

Then there exists a unique probability measure, \( P_\pi \), on \( \sigma(A) \) such that

\[
P_\pi (X_0 = x_0, \ldots, X_n = x_n) = \pi (x_0) p(x_0, x_1) \cdots p(x_{n-1}, x_n)
\]

for all \( n \in \mathbb{N}_0 \) and \( x_0, x_1, \ldots, x_n \in S \). To see such a measure exists, we need only verify that

\[
p_n (x_0, \ldots, x_n) := \pi (x_0) p(x_0, x_1) \cdots p(x_{n-1}, x_n)
\]

verifies the hypothesis of Exercise 5.3 taking into account a shift of the \( n \) − index.

### 6.4 Construction of \( \sigma \)-Finite Measures

The goal of this section is to generalize Theorem 6.20 and Proposition 6.17 to \( \sigma \)-finite measures.

**Theorem 6.28 (\( \sigma \)-Finite Premeasure Extension Theorem).** If \((\Omega, A, \mu)\) is a premeasure space such that \( \mu \) is \( \sigma \)-finite on \( A \) (i.e. there exists \( A_n \in A \) with \( \mu(A_n) < \infty \) such that \( A_n \uparrow \Omega \)), then there exists a unique measure, \( \bar{\mu} \), on \( \sigma(A) \) such that \( \bar{\mu} = \mu \) on \( A \).

**Proof.** Existence of an extension. Let \( \{\Omega_n\}_{n=1}^{\infty} \subset A \) be a partition of \( \Omega \) such that \( \mu(\Omega_n) < \infty \) for all \( n \). Take \( \Omega_n := A_n \setminus A_{n-1} \) with the convention that \( A_0 = \emptyset \). To each \( n \in \mathbb{N} \) let \( \mu_n \) be the finite premeasure on \( A \) defined by

\[
\mu_n (A) := \mu_n (A \cap \Omega_n) \quad \text{for all} \ A \in A
\]

and let \( \bar{\mu}_n \) be an extension of \( \mu_n \) to \( \sigma(A) \) guaranteed by Theorem 6.20. Then the measure, \( \bar{\mu} := \sum_{n=1}^{\infty} \bar{\mu}_n \), is then an extension of \( \mu \) to \( \sigma(A) \).

**Uniqueness of the extension.** Suppose that \( \nu \) is another measure on \( \sigma(A) \) such that \( \nu = \mu \) on \( A \). Given \( n \in \mathbb{N} \), let \( \nu_n \) be the measure on \( \sigma(A) \) defined by \( \nu_n (B) := \nu (B \cap \Omega_n) \) for all \( B \in \sigma(A) \). As \( \nu_n = \mu_n = \bar{\mu}_n \) on \( A \), it follows by Exercise 6.1 that \( \nu_n = \bar{\mu}_n \) on \( \sigma(A) \) and therefore

\[
\nu = \sum_{n=1}^{\infty} \nu_n = \sum_{n=1}^{\infty} \bar{\mu}_n = \bar{\mu}.
\]

**Corollary 6.29.** Suppose that \( \mu \) is a \( \sigma \)-finite premeasure on an algebra \( A \).

If \( B \in \sigma(A) \) and \( \varepsilon > 0 \) is given, there exists \( A \in A_\delta \) and \( C \in A_\sigma \) such that \( A \subset B \subset C \) and \( \bar{\mu}(C \setminus A) < \varepsilon \). Moreover if \( \bar{\mu}(B) < \infty \) and \( \varepsilon > 0 \) is given, there exists \( A \in A \) so that \( \bar{\mu}(B \triangle A) < \varepsilon \).
Proof. Let \( (\Omega_n)_{n=1}^{\infty} \subset A \) be a partition of \( \Omega \) such that \( \mu (\Omega_n) < \infty \) for all \( n \) and let \( \bar{\mu}_n \) be the measure on \( \sigma (A) \) defined by \( \bar{\mu}_n (B) := \bar{\mu} (B \cap \Omega_n) \) for all \( B \in \sigma (A) \). According to Proposition 6.17, to each \( B \in \sigma (A) \) and \( \varepsilon > 0 \), there exists (for all \( n \in \mathbb{N} \)) \( C_n \in A_\sigma \) such that \( B \subset C_n \) and
\[
\varepsilon 2^{-n} > \bar{\mu}_n (C_n \setminus B) = \bar{\mu} (\{B \cap \Omega_n \} \setminus B) + \bar{\mu} (\Omega_n \setminus C_n \setminus \Omega_n) = \bar{\mu} (\{C_n \cap \Omega_n \} \setminus B).
\]
Taking \( C := \bigcap_{n=1}^{\infty} [C_n \cap \Omega_n] \in A_\sigma \) it follows that \( B \subset C \) and
\[
\bar{\mu} (C \setminus B) = \sum_{n=1}^{\infty} \bar{\mu} (C_n \cap \Omega_n \setminus B) \leq \sum_{n=1}^{\infty} \varepsilon 2^{-n} < \varepsilon.
\]
Applying this same result to \( B^c \in \sigma (A) \), there exists \( A \in A_\delta \) so that \( B^c \subset A^c \in A_\sigma \) and
\[
\bar{\mu} (B \setminus A) = \bar{\mu} (A^c \setminus B^c) < \varepsilon.
\]
From this we conclude that \( \bar{\mu} (C \setminus A) < 2\varepsilon \) 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 there exists a \( C \in A_\sigma \) such that \( B \subset C \) with \( \bar{\mu} (C \setminus B) < \varepsilon \) and further, by the definition of \( A_\sigma \), there exists \( A_n \in A \) such that \( A_n \uparrow C \in A_\sigma \). Hence it follows that,
\[
\bar{\mu} (A_n \triangle B) = \bar{\mu} ([A_n \setminus B] \cup [B \setminus A_n]) \leq \bar{\mu} (A_n \setminus B) + \bar{\mu} (B \setminus A_n) \leq \bar{\mu} (C \setminus B) + \bar{\mu} (B \setminus A_n) < \varepsilon + \bar{\mu} (B \setminus A_n).
\]
As \( \bar{\mu} (B) < \infty \) and \( B \setminus A_n \subset C \setminus A_n \downarrow \emptyset \), we know that \( \bar{\mu} (B \setminus A_n) \downarrow 0 \) and therefore for \( n \) sufficiently large we see that \( \bar{\mu} (A_n \triangle B) < \varepsilon \).

**Corollary 6.30.** Suppose that \( \mu \) is a \( \sigma \) – finite premeasure on an algebra \( A \) and \( \bar{\mu} \) is its unique extension to a measure on \( \sigma (A) \), then again \( \bar{\mu} = \mu^* \) on \( \sigma (A) \).

**Proof.** From Lemma 6.29 we know that \( \bar{\mu} \leq \mu^* \) on \( \sigma (A) \). Let \( B \in \sigma (A) \) and \( \varepsilon > 0 \) be given and then use Corollary 6.29 to find \( C \in A_\sigma \) such that \( B \subset C \) and
\[
\bar{\mu} (C \setminus B) < \varepsilon.
\]
If \( A_n \in A \) with \( A_n \uparrow C \), it follows from Eq. 6.5 of Proposition 6.14 and the continuity of \( \bar{\mu} \) under increasing limits that
\[
\lim_{n \to \infty} \mu (A_n) = \lim_{n \to \infty} \bar{\mu} (A_n) = \bar{\mu} (C).
\]
Combining these observations implies,
\[
\bar{\mu} (B) \leq \mu^* (B) \leq \mu^* (C) = \bar{\mu} (C) = \bar{\mu} (C \setminus B) + \bar{\mu} (B) + \varepsilon,
\]
i.e.
\[
\bar{\mu} (B) \leq \mu^* (B) \leq \bar{\mu} (B) + \varepsilon \text{ for all } \varepsilon > 0.
\]
As this is true for all \( \varepsilon > 0 \), we have shown \( \bar{\mu} (B) = \mu^* (B) \) and the proof is complete.

The following slight reformation of Theorem 6.28 can be useful.

**Corollary 6.31 (\(*\)).** Let \( A \) be an algebra of sets, \( \{\Omega_n\}_{n=1}^{\infty} \subset A \) is a given sequence of sets such that \( \Omega_m \uparrow \Omega \) as \( m \to \infty \). Let
\[
A_f := \{A \in A : A \subset \Omega_m \text{ for some } m \in \mathbb{N}\}.
\]
Notice that \( A_f \) is a ring, i.e. closed under differences, intersections and unions and contains the empty set. Further suppose that \( \mu : A_f \to [0, \infty) \) is an additive set function such that \( \mu (A_n) \downarrow 0 \) for any sequence, \( \{A_n\} \subset A \) such that \( A_n \downarrow \emptyset \) as \( n \to \infty \). Then \( \mu \) extends uniquely to a \( \sigma \) – finite measure on \( A \).

**Proof. Existence.** By assumption, \( \mu_m := \mu |_{\Omega_m} : A_{\Omega_m} \to [0, \infty) \) is a premeasure on \( (\Omega_m, A_{\Omega_m}) \) and hence by Theorem 6.28 extends to a measure \( \mu_m' \) on \( (\Omega_m, \sigma (A_{\Omega_m}) = B_{\Omega_m}) \). Let \( \bar{\mu}_m (B) := \mu_m' (B \cap \Omega_m) \) for all \( B \in B \). Then \( \{\mu_m\}_{m=1}^{\infty} \) is an increasing sequence of measure on \( (\Omega, B) \) and hence \( \bar{\mu} := \lim_{m \to \infty} \mu_m \) defines a measure on \( (\Omega, B) \) such that \( \bar{\mu} |_{A_f} = \mu \).

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 6.5 or Exercise 1.4 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. \( \mu_1 = \mu_2 \).

**6.5 Radon Measures on \( \mathbb{R} \)**

We say that a measure, \( \mu \), on \( (\mathbb{R}, B_\mathbb{R}) \) 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 \( \mathbb{R} \). Throughout this section let
\[
S := \{(a, b) : -\infty \leq a \leq b \leq \infty\},
\]
be the semi-algebra of half open intervals and \( A = A (S) \) be the algebra generated by \( S \) consisting of those subsets, \( A \subset \mathbb{R} \), which may be written as finite disjoint unions of sets from \( S \), see Example 4.28. Recall that \( B_\mathbb{R} = \sigma (A) = \sigma (S) \).

If \( \mu \) is a Radon measure on \( (\mathbb{R}, B_\mathbb{R}) \), then we can always find a function, \( F : \mathbb{R} \to \mathbb{R} \), such that
\[
\mu ([a, b]) = F (b) - F (a) \quad \text{for all } -\infty < a \leq b < \infty.
\]
(6.11)

For example if \( \mu (\mathbb{R}) < \infty \) we can take \( F (x) = \mu ((-\infty, x]) \) while if \( \mu (\mathbb{R}) = \infty \) we might take
\[
F (x) = \begin{cases} \mu ([0, x]) & \text{if } x \geq 0 \\ -\mu ((x, 0]) & \text{if } x \leq 0 \end{cases}.
\]
(6.12)
Remark 6.32. It is a simple matter to see that Eq. (6.11) only uniquely determines $F$ modulo an additive constant. We can fix this constant by requiring $F$ take on a specific value at a specific point in $\mathbb{R}$ (or possibly $\mathbb{R}$). For example by replacing $F$ by $F - F(0)$ if necessary we may normalize $F$ so that $F(0) = 0$. This normalization then leads to the formula for $F$ in Eq. (6.12). If $\mu(\mathbb{R}) < \infty$ then we may replace $F$ by $F - \lim_{x \to -\infty} F(x)$. In this case the resulting $F$ satisfies $F(-\infty) := \lim_{x \to -\infty} F(x) = 0$ and we find $F(x) = \mu ((-\infty, x])$.

Lemma 6.33. If $\mu$ is a Radon measure on $(\mathbb{R}, \mathcal{B})$ 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 $\mu ((a, b]) < \infty$ 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 6.34. Suppose $F: \mathbb{R} \to \mathbb{R}$ is a right continuous increasing function. Then there exists a unique Radon measure, $\mu = \mu_F$, on $(\mathbb{R}, \mathcal{B})$ such that Eq. (6.11) holds.

Proof. Let us 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 5.29 and Remark 5.10. To finish the proof it suffices by Theorem 6.28 to show that $\mu$ is a premeasure on $\mathcal{A} = \mathcal{A}(\mathcal{S})$. In light of Proposition 6.4, to finish the proof it suffices to show $\mu$ is sub-additive on $\mathcal{S}$. i.e. we must show

$$\mu(J) \leq \sum_{n=1}^\infty \mu(J_n),$$

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

$$\sum_{n=1}^\infty \mu(J_n) \leq \mu(J).$$

We begin with the special case where $-\infty < a < b < \infty$. We now give two proofs of Eq. (6.14).

First proof of Eq. (6.14) when $-\infty < a < b < \infty$. Our proof will be by “continuous induction.” The strategy is to show $a \in A$ where

$$A := \left\{ \alpha \in [a, b] : \mu(J \cap (\alpha, b]) \leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b]) \right\}.$$  (6.15)

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 $J_k \subset A$ so that $A$ is not empty. To finish the proof we are going to show $\bar{a} := \inf A \in A$ and that $\bar{a} = a$.

- Let $\alpha_m \in A$ such that $\alpha_m \downarrow \bar{a}$, i.e.

$$\mu(J \cap (\alpha_m, b]) \leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b]).$$  (6.16)

The right continuity of $F$ implies $\alpha \to \mu(J \cap (\alpha, b])$ is right continuous. So by the dominated convergence theorem for sums,

$$\mu(J \cap (\bar{a}, b]) = \lim_{m \to \infty} \mu(J \cap (\alpha_m, b])$$

$$\leq \lim_{m \to \infty} \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b])$$

$$= \sum_{n=1}^\infty \lim_{m \to \infty} \mu(J_n \cap (\alpha_m, b]) = \sum_{n=1}^\infty \mu(J_n \cap (\bar{a}, b]),$$

i.e. $\bar{a} \in A$.

- 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,

$$\mu(J \cap (\alpha, b]) = \mu(J \cap (\alpha, \bar{a})] + \mu(J \cap (\bar{a}, b])$$

$$\leq \mu(J_l \cap (\alpha, \bar{a})] + \sum_{n \neq l} \mu(J_n \cap (\bar{a}, b])$$

$$= \mu(J_l \cap (\alpha, \bar{a})] + \mu(J_l \cap (\bar{a}, b]) + \sum_{n \neq l} \mu(J_n \cap (\bar{a}, b])$$

$$= \mu(J_l \cap (\alpha, b]) + \sum_{n \neq l} \mu(J_n \cap (\bar{a}, b])$$

$$\leq \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b]).$$

This shows $\alpha \in A$ and $\alpha < \bar{a}$ which violates the definition of $\bar{a}$. Thus we must conclude that $\bar{a} = a$.  

---

1 DCT applies as $\mu(J_n \cap (\alpha_m, b]) \leq \mu(J_n)$ and $\sum_{n=1}^\infty \mu(J_n) \leq \mu(J) < \infty$ by Eq. (6.16).
Second proof of Eq. (6.14) when \(-\infty < a < b < \infty\). Choose numbers \(\tilde{a} > a, \tilde{b}_n > b_n\) in which case \(\tilde{I} := (\tilde{a}, \tilde{b}] \subset J\).

\[
\tilde{J}_n := (a_n, \tilde{b}_n] \supset \tilde{J}_n := (a_n, \tilde{b}_n) \supset J_n.
\]

Since \(\tilde{I} = [\tilde{a}, b]\) is compact and \(\tilde{I} \subset J \subset \bigcup_{n=1}^{\infty} \tilde{J}_n\) there exists \(N < \infty\) such that

\[
I \subset \tilde{I} \subset \bigcup_{n=1}^{N} \tilde{J}_n \subset \bigcup_{n=1}^{N} J_n.
\]

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,

\[
\mu(J) = \mu([a, b]) = F(b) - F(\tilde{a}) \leq \sum_{n=1}^{\infty} \mu(\tilde{J}_n).
\]

\[
= \sum_{n=1}^{\infty} \mu(J_n) + \sum_{n=1}^{\infty} \mu(\tilde{J}_n \setminus J_n).
\]

Given \(\varepsilon > 0\), we may use the right continuity of \(F\) to choose \(\tilde{b}_n\) so that

\[
\mu(\tilde{J}_n \setminus J_n) = F(\tilde{b}_n) - F(b_n) \leq \varepsilon 2^{-n} \quad \forall \ n \in \mathbb{N}.
\]

Using this in Eq. (6.17) shows

\[
\mu(J) = \mu([a, b]) \leq \sum_{n=1}^{\infty} \mu(J_n) + \varepsilon
\]

which verifies Eq. (6.13) since \(\varepsilon > 0\) was arbitrary.

Using either of the two proofs above, 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 \(c := \sup \{x \leq b : [\tilde{a}, x]\) is finitely covered by \(\left\{\tilde{J}_n\right\}_{n=1}^{\infty}\). If \(c < b\), then \(c \in \tilde{J}_m\) for some \(m\) and there exists \(x \in \tilde{J}_m\) such that \([\tilde{a}, x]\) is finitely covered by \(\left\{\tilde{J}_n\right\}_{n=1}^{\infty}\), say by \(\left\{\tilde{J}_n\right\}_{n=1}^{N}\). We would then have that \(\left\{\tilde{J}_n\right\}_{n=1}^{\max(m, N)}\) finitely covers \([\tilde{a}, \tilde{c}]\) for all \(\tilde{c} \in \tilde{J}_m\). But this contradicts the definition of \(c\).
Proof. Write $A = \sum_{k=1}^{n} (a_k, b_k)$ and then take $B := \sum_{k=1}^{n} (\bar{a}_k, b_k)$ where $a_k < \bar{a}_k < b_k$. Then

$$B \subset \bar{B} = \sum_{k=1}^{n} [\bar{a}_k, b_k] \subset A$$

and

$$\mu (A \setminus B) = \mu \left( \sum_{k=1}^{n} (a_k, \bar{a}_k) \right) = \sum_{k=1}^{n} [F(\bar{a}_k) - F(a_k)].$$

Using the right continuity of $F$ we may now choose $\bar{a}_k$ sufficiently close to $a_k$ so that the latter sum is less than $\delta$. □

Alternate proof of Theorem 6.34 For $J = (\alpha, \beta]$ with $-\infty < \alpha < \beta < \infty$, let $\mu_J (A) := \mu (J \cap A)$ for all $A \in \mathcal{A}$ so that $\mu_J$ is a finitely additive measure on $\mathcal{A}$ such that

$$\mu_J (\mathbb{R}) = \mu (J) = F(\beta) - F(\alpha) < \infty.$$

Claim: $\mu_J$ is a premeasure and hence extends to a measure, $\tilde{\mu}_J$, on $\mathcal{A} = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

To prove this claim, suppose that $A_n \in \mathcal{A}$ such that $A_n \downarrow$ and

$$\varepsilon := \lim_{n \to \infty} \mu_J (A_n) = \inf_n \mu_J (A_n) > 0.$$

We need to show $\cap_{n=1}^{\infty} A_n \neq \emptyset$. Since $\mu_J (A_n) = \mu (A_n \cap J)$ we may replace $A_n$ by $A_n \cap J$ if necessary so that $A_n \subset J$ for all $n$ and particular $A_n$ is bounded for all $n$. By the Lemma 6.36 we may choose $B_n \in \mathcal{A}$ such that $B_n \subset \bar{B}_n \subset A_n$ and $\mu (A_n \setminus B_n) < \varepsilon 2^{-(n+1)}$. It then follows that

$$\varepsilon - \mu (B_1 \cap \cdots \cap B_n) \leq \mu_J (A_n) - \mu_J (B_1 \cap \cdots \cap B_n)$$

$$= \mu_J (A_n \setminus B_1 \cap \cdots \cap B_n) \leq \frac{1}{2} \sum_{k=1}^{n} \varepsilon 2^{-k} \leq \varepsilon / 2$$

and hence $\mu_J (B_1 \cap \cdots \cap B_n) \geq \varepsilon / 2$ for all $n$. In particular, this shows

$$\emptyset \neq B_1 \cap \cdots \cap B_n \subset B_1 \cap \cdots \cap \bar{B}_n \text{ for all } n \in \mathbb{N}.$$

Since the $\{B_n\}_{n=1}^{\infty}$ are all closed subsets of the compact set, $J$, they satisfy the finite intersection property and we may conclude that

$$\emptyset \neq \cap_{n=1}^{\infty} B_n \subset \cap_{n=1}^{\infty} A_n$$

and the claim is proved.

To finish the proof of existence, let $J_n := (n, n + 1]$ and then define $\bar{\mu} := \sum_{n \in \mathbb{Z}} \mu_J (J_n)$, which is a measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Let us note that $\bar{\mu}_J (J_n) = \mu_J (J_n) = \mu (\emptyset) = 0$. So given $-\infty < a < b < \infty$ we choose $M \in \mathbb{N}$ so that $(a, b] \subset (-M, M]$, then

$$\bar{\mu} ((a, b]) = \sum_{|n| \leq M} \bar{\mu}_J ((a, b]) = \sum_{|n| \leq M} \mu_J ((a, b])$$

$$= \sum_{|n| \leq M} \mu (J_n \cap (a, b]) = \mu ((-M, M + 1] \cap (a, b])$$

$$= \mu ((a, b]) = F(b) - F(a).$$

Uniqueness. Suppose that $\mu$ and $\nu$ are two $\sigma$–additive measures Radon measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which agree on $\mathcal{A}$. For $J = (\alpha, \beta]$ as above, let $\mu_J (B) = \mu (B \cap J)$ and $\nu_J (B) = \nu (B \cap J)$ for all $B \in \mathcal{B}_{\mathbb{R}}$. Then $\mu_J$ and $\nu_J$ are bounded measures $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which agree on $\mathcal{A}$ and hence $\mu_J = \nu_J$. It then follows for any $B \in \mathcal{B}_{\mathbb{R}}$ that,

$$\mu (B) = \lim_{n \to \infty} \mu ((-n, n] \cap B) = \lim_{n \to \infty} \nu ((-n, n] \cap B) = \nu (B).$$

Exercise 6.4. Let $\Omega = \mathbb{R}$, $\mathcal{A} \subset \mathcal{P}^{\mathbb{R}}$ be the algebra generated by half open intervals, $\mu$ be the length measure on $\mathcal{A}$, and set $B = \cup_{n=-\infty}^{\infty} (2n, 2n + 1] \subset \mathcal{A}$. Show $d (B, A) = \infty$ for all $A \in \mathcal{A}$ and use this to conclude that $\mathcal{A}_\sigma$ is not contained in $\mathcal{A}$ and that $B = \emptyset$ is not a $\sigma$–algebra.

Definition 6.37. 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

$$F(x) := P ((-\infty, x]).$$

Example 6.38. 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,

$$F(x) = \begin{cases} 0 & \text{for } x < -1 \\ p & \text{for } -1 \leq x < 1 \\ p + q & \text{for } 1 \leq x < \pi \\ 1 & \text{for } \pi \leq x < \infty \end{cases}.$$
Lemma 6.39. 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. (6.18) 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 \( F(b+) := \lim_{y \uparrow b} F(y) = F(b) \). This show that \( F \) is right continuous. Similar arguments show that \( F(\infty) = 1 \) and \( F(\infty) = 0 \).

The converse of Lemma 6.39 now follows directly from Theorem 6.34.

Theorem 6.40. To each function \( F : \mathbb{R} \to [0, 1] \) satisfying properties 1. – 3., in Lemma 6.39, there exists a unique probability measure, \( P_F \), on \( \mathcal{B}_\mathbb{R} \) such that

\[
P_F((a, b]) = F(b) - F(a) \text{ for all } -\infty < a \leq b < \infty.
\]

Example 6.41 (Uniform Distribution). The function,

\[
F(x) := \begin{cases} 
0 & x \leq 0 \\
x & 0 \leq x < 1 \\
1 & 1 \leq x < \infty
\end{cases}
\]

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.

6.6 Lebesgue Measure

Definition 6.42 (Lebesgue Measure). If \( F(x) = x \) for all \( x \in \mathbb{R} \), we denote \( \mu_F \) by \( m \) and call \( m \) Lebesgue measure on \((\mathbb{R}, \mathcal{B}_\mathbb{R}) \), i.e. Lebesgue measure is the unique \( \sigma \)-additive measure, \( m \), on \((\mathbb{R}, \mathcal{B}_\mathbb{R}) \) such that

\[
m((a, b]) = b - a \text{ for all } -\infty < a < b < \infty.
\]

Notation 6.43 Given \( x \in \mathbb{R} \) and \( B \subset \mathbb{R} \), let

\[
x + B := \{ x + y : y \in B \} \text{ and } x \cdot B := \{ xy : y \in B \}.
\]

In Exercise 9.2 below are asked to show that for any \( B \in \mathcal{B}_\mathbb{R} \) and any \( x \in \mathbb{R} \) that \( x + B \in \mathcal{B}_\mathbb{R} \) and \( x \cdot B \in \mathcal{B}_\mathbb{R} \).

Theorem 6.44. Lebesgue measure \( m \) is invariant under translations, i.e. for \( B \in \mathcal{B}_\mathbb{R} \) and \( x \in \mathbb{R} \),

\[
m(x + B) = m(B). \tag{6.19}
\]

Lebesgue measure, \( m \), is the unique measure on \( \mathcal{B}_\mathbb{R} \) such that \( m((0,1]) = 1 \) and Eq. (6.19) holds for \( B \in \mathcal{B}_\mathbb{R} \) and \( x \in \mathbb{R} \). Moreover, \( m \) has the scaling property

\[
m(\lambda B) = |\lambda| m(B) \tag{6.20}
\]

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, \( m_x = m \) on \( \mathcal{A}(S) \)
where $S$ is the semi-algebra of half open intervals and so by the uniqueness assertion in Theorem 6.28 or more directly using Theorem 6.34.

For the converse, suppose that $m$ is translation invariant and $m((0, 1]) = 1$. Given $n \in \mathbb{N}$, we have

$$
(0, 1] = \bigcup_{k=1}^{n} \left( \frac{k-1}{n}, \frac{k}{n} \right) = \bigcup_{k=1}^{n} \left( \frac{k-1}{n} + (0, \frac{1}{n}] \right).
$$

Therefore,

$$
1 = m((0, 1]) = \sum_{k=1}^{n} m \left( \frac{k-1}{n} + (0, \frac{1}{n}] \right)
= \sum_{k=1}^{n} m \left( 0, \frac{1}{n} \right) = n \cdot m \left( 0, \frac{1}{n} \right).
$$

That is to say

$$
m \left( 0, \frac{1}{n} \right) = 1/n.
$$

Similarly, $m((0, \frac{1}{n}]) = l/n$ 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. (6.20) 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 $B_\mathbb{R}$ 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 $m_\lambda = m$.

### 6.7 Metric-Measure Space Regularity Results

This section is a self study guide to the “approximating” Borel sets in a metric space by closed and open subsets of the metric space. We begin with some basic properties of metric spaces. Throughout this section we will assume that $(X, \rho)$ is a metric space and $B_X$ denotes the Borel $\sigma$-algebra on $X$.

**Lemma 6.45.** For any non empty subset $A \subset X$, let $\rho_A (x) := \inf \{ \rho(x, a) | a \in A \}$, then

$$
|\rho_A (x) - \rho_A (y)| \leq \rho (x, y) \quad \forall x, y \in X
$$

which shows $\rho_A : X \to [0, \infty)$ is continuous.

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

$$
\rho_A (x) \leq \rho (x, a) \leq \rho (x, y) + \rho (y, a).
$$

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

$$
\rho_A (x) \leq \rho (x, y) + \rho_A (y) \quad \forall x, y \in X.
$$

Therefore, $\rho_A (x) - \rho_A (y) \leq \rho (x, y)$ and by interchanging $x$ and $y$ we also have that $\rho_A (y) - \rho_A (x) \leq \rho (x, y)$ which implies Eq. (6.21).

**Corollary 6.46.** The function $\rho$ satisfies,

$$
|\rho (x, y) - \rho (x', y')| \leq \rho (y, y') + \rho (x, x').
$$

In particular $\rho : X \times X \to [0, \infty)$ is continuous.

**Proof.** By Lemma 6.45 for single point sets and the triangle inequality for the absolute value of real numbers,

$$
|\rho (x, y) - \rho (x', y')| \leq |\rho (x, y) - \rho (x, y')| + |\rho (x, y') - \rho (x', y')| \leq \rho (y, y') + \rho (x, x').
$$

**Corollary 6.47.** Given any set $A \subset X$ and $\varepsilon > 0$, then

$$
A_\varepsilon := \{ \rho_A < \varepsilon \} := \{ x \in X : \rho_A (x) < \varepsilon \}
$$

is an open set containing $A$ and $A_\varepsilon \downarrow \bar{A}$ as $\varepsilon \downarrow 0$ where $\bar{A}$ is the closure of $A$. Similarly,

$$
F_\varepsilon := \{ \rho_A \geq \varepsilon \} := \{ x \in X : \rho_A (x) \geq \varepsilon \}
$$

is a closed set and $F_\varepsilon \uparrow (A^c)^o$ as $\varepsilon \downarrow 0$ where $(A^c)^o$ is the interior of $A^c := X \setminus A$.

**Proof.** Because of the continuity of $\rho_A$ and the facts that $(-\infty, \varepsilon)$ is open in $\mathbb{R}$ and $[\varepsilon, \infty)$ is closed in $\mathbb{R}$, it follows that $A_\varepsilon = \rho_A^{-1} ((-\infty, \varepsilon))$ is open and $F_\varepsilon = \rho_A^{-1} ([\varepsilon, \infty))$ is closed. We have $x \in \cap_{\varepsilon > 0} A_\varepsilon$ iff $\rho_A (x) < \varepsilon$ for all $\varepsilon > 0$ iff $\rho_A (x) = 0$ and hence

$$
A \subset \{ \rho_A = 0 \} = \cap_{\varepsilon > 0} A_\varepsilon.
$$
Since \( \{\rho_A = 0\} \) is closed it follows that \( \tilde{A} \subseteq \{\rho_A = 0\} \). Conversely if \( x \in \{\rho_A = 0\} \) then there exists \( \{x_n\} \subset A \) such that \( \lim_{n \to \infty} \rho(x,x_n) = 0 \), i.e. \( x_n \to x \) and therefore \( x \in A \).

To finish the proof observe that
\[
\bigcup_{\varepsilon > 0} F_\varepsilon^c = \bigcap_{\varepsilon > 0} F_\varepsilon = \bigcap_{\varepsilon > 0} \{\rho_A < \varepsilon\} = \tilde{A}
\]
and therefore
\[
\bigcup_{\varepsilon > 0} F_\varepsilon = \tilde{A}^c = (A^c)^c.
\]

\begin{lemma}[Urysohn’s Lemma for Metric Spaces] \( \)\( \text{Let} \ (X,d) \text{be a metric space and suppose that} \ A \text{and} \ B \text{are two disjoint closed subsets of} \ X. \text{Then}
\end{lemma}

\[
f \left( x \right) = \frac{d_B \left( x \right)}{d_A \left( x \right) + d_B \left( x \right)} \quad \text{for} \ x \in X
\]
defines a continuous function, \( f : X \to [0,1] \), such that \( f \left( x \right) = 1 \) for \( x \in A \) and \( f \left( x \right) = 0 \) if \( x \in B \).

**Proof.** By Lemma [6.4.5] \( d_A \) and \( d_B \) are continuous functions on \( X \). Since \( A \) and \( B \) are closed, \( d_A \left( x \right) > 0 \) if \( x \notin A \) and \( d_B \left( x \right) > 0 \) if \( x \notin B \). Since \( A \cap B = \emptyset \), \( d_A \left( x \right) + d_B \left( x \right) > 0 \) 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 used in the following form. Suppose \( F \subseteq V \subset X \) with \( F \) being closed and \( V \) being open, then there exists \( f \in C \left( X, [0,1] \right) \) such that \( f = 1 \) on \( F \) while \( f = 0 \) on \( V^c \). This of course follows from Lemma [6.4.8] by taking \( A = F \) and \( B = V^c \).

\begin{corollary} \( \) \( \text{If} \ A \text{and} B \text{are two disjoint closed subsets of a metric space,} \ (X,d), \text{then there exists disjoint open subsets} \ U \text{and} V \text{of} \ X \text{such that} \ A \subseteq U \text{and} B \subseteq V. \end{corollary}

**Proof.** Let \( f \) be as in Lemma [6.4.8] so that \( f \in C \left( X, [0,1] \right) \) such that \( f = 1 \) on \( A \) and \( f = 0 \) on \( B \). Then set \( U = \{ f > \frac{1}{2} \} \) and \( V = \{ f < 1/2 \} \).

We now (as we did with outer measures) associate a pseudo metric to any measure space.

\begin{proposition}[The measure pseudo metric] \( \) \( \text{Let} \ (\Omega,\mathcal{B},\mu) \text{be a measure space and define}
\end{proposition}

\[
d_\mu \left( A,B \right) := \mu \left( A \triangle B \right) \in [0,\infty] \ \forall \ A, B \in \mathcal{B}.
\]

Then \( d = d_\mu \) satisfies:

1. \( d \) is a pseudo metric, i.e. \( d \left( A,B \right) = d \left( B,A \right) \) and \( d \left( A,C \right) \leq d \left( A,B \right) + d \left( B,C \right) \) for all \( A,B,C \in \mathcal{B} \).
2. \( d \left( A^c,C^c \right) = d \left( A,C \right) \) for all \( A,B \in \mathcal{B} \).
3. If \( \{A_n\}_{n=1}^\infty, \{B_n\}_{n=1}^\infty \subseteq \mathcal{B} \), then
\[
d \left( \bigcup_{n=1}^\infty A_n, \bigcup_{n=1}^\infty B_n \right) \leq \sum_{n=1}^\infty d \left( A_n, B_n \right) \quad \text{and}
\]
\[
d \left( \bigcap_{n=1}^\infty A_n, \bigcap_{n=1}^\infty B_n \right) \leq \sum_{n=1}^\infty d \left( A_n, B_n \right).
\]

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

1. The fact that \( d \) is a pseudo metric easily follows from the fact that \( 1_{A \triangle C} = \left| 1_A - 1_C \right| \) and therefore,
\[
d \left( A,C \right) = \mathbb{E} \left| 1_A - 1_C \right|.
\]

2. Item 2. follows from the fact that
\[
A^c \triangle C^c = \left[ A^c \cap C \right] \cup \left[ C^c \cap A \right] = \left[ C \setminus A \right] \cup \left[ A \setminus C \right] = A \triangle C
\]
which is also seen via,
\[
1_{A \triangle C^c} = \left| 1_A - 1_C \right| = \left| 1_A - 1 \right| \left| 1_C - 1 \right| = \left| 1_A - 1_C \right| = 1_{A \triangle C}.
\]

3. It is a simple exercise to verify,
\[
\left[ \bigcup_{n=1}^\infty A_n \right] \triangle \left[ \bigcup_{n=1}^\infty B_n \right] \subseteq \bigcup_{n=1}^\infty \left[ A_n \triangle B_n \right]
\]
and hence
\[
d \left( \bigcup_{n=1}^\infty A_n, \bigcup_{n=1}^\infty B_n \right) = \mu \left( \bigcup_{n=1}^\infty \left[ A_n \triangle B_n \right] \right) \leq \sum_{n=1}^\infty d \left( A_n, B_n \right),
\]
which proves Eq. (6.23). Equation (6.24) may be proved similarly or by combining item 2. with Eq. (6.23) as follows;
6 Countably Additive Measures

\[ d (\bigcap_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} B_n) = d (\bigcup_{n=1}^{\infty} A_n^c, \bigcup_{n=1}^{\infty} B_n^c) \leq \sum_{n=1}^{\infty} d (A_n^c, B_n^c) = \sum_{n=1}^{\infty} d (A_n, B_n). \]

6.7.1 Regularity Exercises for Borel measures on \((X, B_X)\)

Exercise 6.5. If \((X, \mu)\) is a metric space and \(\mu\) is a finite measure on \((X, B_X)\), then for all \(A \in B_X\) and \(\epsilon > 0\) there exists a closed set \(F\) and open set \(V\) such that \(F \subset A \subset V\) and \(\mu (V \setminus F) = \mu (F \setminus V) < \epsilon\). Here are some suggestions:

1. Let \(B_0\) denote those \(A \subset X\) such that for all \(\epsilon > 0\) there exists a closed set \(F\) and open set \(V\) such that \(F \subset A \subset V\) and \(\mu (V \setminus F) = \mu (F \setminus V) < \epsilon\).
2. Show \(B_0\) contains all closed (or open if you like) sets using Corollary 6.47.
3. Show \(B_0\) is a \(\sigma\)-algebra. [You may find Proposition 6.50 to be helpful in this step.]
4. Explain why this proves the result.

Exercise 6.6. Let \((X, \mu)\) be a metric space and \(\mu\) be a measure on \((X, B_X)\). If there exists open sets, \(\{V_n\}_{n=1}^{\infty}\) of \(X\) such that \(V_n \uparrow X\) and \(\mu (V_n) < \infty\) for all \(n\), then for all \(A \in B_X\) and \(\epsilon > 0\) there exists a closed set \(F\) and open set \(V\) such that \(F \subset A \subset V\) and \(\mu (V \setminus F) = \mu (F \setminus V) < \epsilon\). Hints:

1. Show it suffices to prove; for all \(\epsilon > 0\) and \(A \in B_X\), there exists an open set \(V\) such that \(V \subset A \subset \mu (V) < \epsilon\).
2. Now you must verify the assertion above holds. For this, you may find it useful to apply Exercise 6.5 to the measures, \(\mu_n : B_X \to [0, \mu (V_n)]\), defined by \(\mu_n (A) := \mu (A \cap V_n)\) for all \(A \in B_X\). The \(\epsilon > 0\) in Exercise 6.5 should be replaced by judiciously chosen small quantities \(\epsilon_n > 0\) depending on \(n\), for example \(\epsilon_n = \epsilon 2^{-n}\) will work.

Exercise 6.7 (Special case of above). Let \(B = B_{\mathbb{R}^n} = \sigma (\{\text{open subsets of } \mathbb{R}^n\})\) be the Borel \(\sigma\)-algebra on \(\mathbb{R}^n\) and \(\mu\) be a probability measure on \(B\). Further, let \(B_0\) denote those sets \(B \in B\) such that for every \(\epsilon > 0\) there exists \(F \subset B \subset V\) such that \(F\) is closed, \(V\) is open, and \(\mu (V \setminus F) < \epsilon\). Show:

1. \(B_0\) contains all closed subsets of \(B\). [Given a closed subset, \(F \subset \mathbb{R}^n\) and \(k \in \mathbb{N}\), let \(V_k := \bigcup_{x \in F} B(x, 1/k)\), where \(B(x, \delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}\). Show, \(V_k \uparrow F\) as \(k \to \infty\).
2. Show \(B_0\) is a \(\sigma\)-algebra and use this along with the first part of this exercise to conclude \(B = B_0\). [Follow closely the method used in the first step of the proof of Theorem 7.7.]
3. Show for every \(\epsilon > 0\) and \(B \in B\), there exist a compact subset, \(K \subset \mathbb{R}^n\), such that \(K \subset B\) and \(\mu (B \setminus K) < \epsilon\). [Take \(K := F \cap \{x \in \mathbb{R}^n : |x| \leq n\}\) for some sufficiently large \(n\).

Go to Chapter 9

6.8 * The \(\pi - \lambda\) theorem

This section may and probably should be omitted on first reading. Later in Chapter 11 we will come back to a function theoretic variant of the results in this chapter which we will tend to use throughout the book.

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 6.51 (\(\lambda\)-system). A collection of sets, \(\mathcal{L} \subset 2^\Omega\), is a \(\lambda\)-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 6.52. 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 6.53 (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.

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 \( A \subseteq B \), 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 \( \bigcup_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} B_n \in \mathcal{L} \).

**Theorem 6.54 (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 \( A \subseteq B \) with \( A, B \in \mathcal{L}^C \), then \( A \cap C, B \cap C \in \mathcal{L} \) with \( A \cap C \subseteq B \cap C \) and therefore,

\[
(D \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.
\]

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 \( A_n \cap A \cap C \subseteq B \cap C \), 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 6.52 \( \mathcal{L} \) 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 6.55.** Suppose that \( \mathcal{P} \subset 2^\Omega \) is a \( \pi \) – system. If \( P \) and \( Q \) are two probability 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

\[
\lambda \in \mathcal{L} \quad \text{and hence} \quad \lambda \in \mathcal{L}.
\]

More generally, \( P \) and \( Q \) could be two measures such that \( P(\Omega) = Q(\Omega) < \infty \).

\[
P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)
\]

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 6.56.** 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 \( \nu(\{a\}) = \frac{1}{5} \) and \( \nu(\{b\}) = \nu(\{c\}) = \frac{3}{8} \). In this example,

\[
\mathcal{L} := \{ A \in 2^\Omega : P(A) = Q(A) \}
\]

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 6.8.** 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, \( \mu_n(\Omega) := \mu(\Omega \cap \Omega_n) \) and \( \nu_n(\Omega) = \nu(\Omega \cap \Omega_n) \).

**Corollary 6.57.** 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 6.55 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 6.58.** Corollary 6.57 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,

\[
F(x) := P((\infty, x]) \quad \text{for all } x \in \mathbb{R}^n
\]

where now

\[
(\infty, x] := (\infty, x_1] \times (\infty, x_2] \times \cdots \times (\infty, x_n].
\]

**Exercise 6.9.** Given \( x \in \mathbb{R} \setminus \{0\} \) let

\[
x + B := \{ x + y : y \in B \} \quad \text{and} \quad x \cdot B := \{ xy : y \in B \}.
\]

Use the \( \pi - \lambda \) Theorem 6.54 to show \( x + B \) and \( x \cdot B \) are in \( \mathcal{B}_\mathbb{R} \) for all \( B \in \mathcal{B}_\mathbb{R} \).
Exercise 6.10 (* Density of $A$ in $\sigma (A)$). Suppose that $A \subset 2^\Omega$ is an algebra, $B := \sigma (A)$, and $P$ is a probability measure on $B$. Let $\rho (A, B) := P (A \triangle B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $A$ is dense in $B$ relative to the “metric,” $\rho$. More precisely you are to show using the following outline that for every $B \in B$ there exists $A \in A$ such that that $P (A \triangle B) < \varepsilon$. [Some of the assertions to be proved below have already been covered in the Exercises in subsection 4.2]

1. Recall from Exercise 5.6 that $\rho (a, B) = P (A \triangle B) = \mathbb{E} |1_A - 1_B|$. 
2. Observe; if $B = \bigcup B_i$ and $A = \bigcup A_i$, then

$$B \setminus A = \bigcup_i \{B_i \setminus A_i\} \subset \bigcup_i A_i \triangle B_i$$

so that

$$A \triangle B \supset \bigcup_i (A_i \triangle B_i).$$

3. We also have

$$B_2 \setminus B_1 \setminus A_2 \setminus A_1 = B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c$$

and similarly,

$$A_1 \setminus B_2 \setminus A_2 \setminus B_1 \supset (A_2 \setminus B_2) \cup (B_1 \setminus A_1).$$

4. Observe that $A_n \subset B$ and $A_n \uparrow A$.

Thus

$$P (B \triangle A_n) = P (B \setminus A_n) + P (A_n \setminus B)$$

and

$$P (A \triangle B) := P (A \setminus B) + P (B \setminus A) = P (A \triangle B).$$

5. Let $\mathcal{L}$ be the collection of sets $B \in B$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a $\lambda -$ system which contains $A$.

6.9 * Supplement: Generalizations of Theorem 6.34 to $\mathbb{R}^n$

Notation 6.59 Let $a, b \in \mathbb{R}^n$ 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 6.60. 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 $b \in [a, \delta)$. 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 $a < y < b$.

Definition 6.61. Let $A = \mathcal{A} (\mathbb{R}^n)$ be the sub-algebra of $\mathbb{R}^n$ generated by sets of the form $(a, b] \cap \mathbb{R}^n$ with $a, b \in \mathbb{R}^n$. [Note that $(a, b] = \emptyset$ if $a_i \geq b_i$ for some $i$.]

Theorem 6.62. Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is a right continuous function and $\mu : \mathcal{A} (\mathbb{R}^n) \to \mathbb{R}$ is the associated finitely additive measure as given in Eq. (5.40) of Proposition 5.41. If $\mu_F$ is positive in the sense that $\mu_F ((a, b]) \geq 0$ for all $a \leq b$, then there exists a unique measure $\mu = \mu_F$ on $\mathcal{B}_\mathbb{R}$ such that

$$\mu ((a, b]) = \mu_F ((a, b]) = \sum_{A \subset \{1, \ldots, n\},} (-1)^{|A|} F (a_A \times b_{\Lambda^c})$$

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 $b_n > b_n$ for all $n \in \mathbb{N}$, we have

\[ It is not sufficient to assume that $F$ is non-decreasing in each of its variables. For example $F (x, y) := x + y - xy$ on $[0, 1]^2$ satisfies, $F_x = 1 - y \geq 0$ and $F_y = 1 - x \geq 0$ while $F_{x,y} = -1 < 0$. In this case $\mu_F ([0, 1]^2) = 0$ and more generally $\mu_F ((a, b]) = (b_1 - a_1)(b_2 - a_2) < 0$ for all $a < b$. Moreover if $F$ is sufficiently smooth, then non-decreasing in each of its variables means $\partial_i F \geq 0$ for all $i$ whereas $\partial_i F = \partial_1 \cdots \partial_i F$ will be a positive measure iff $\partial_1 \cdots \partial_i F \geq 0$.\]
implies, such that Also using the right continuity of \( \mu \), By finite sub-additivity and monotonicity of \( \mu \), we have

Using the right continuity of \( F \) it follows that

Also using the right continuity of \( F \), for every \( \varepsilon > 0 \) we may choose \( \tilde{b}_n > b_n \) such that

\[
\mu \left( (a, b] \right) \leq \sum_{n=1}^{\infty} \mu \left( (a_n, \tilde{b}_n] \right) \leq \sum_{n=1}^{\infty} \mu \left( (a_n, b_n] \right) + \varepsilon 2^{-n}
\]

which combined with Eq. (6.27) implies,

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 6.34. ■

**Lemma 6.63 (Right Continuous Versions).** 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 \( F \). ■

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.

### 6.10 Appendix: Completions of Measure Spaces

**Definition 6.64.** A set \( E \subseteq \Omega \) is a *null* set if \( E \in \mathcal{B} \) and \( \mu(E) = 0 \). If \( F \) 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 6.65.** A measure space \((\Omega, \mathcal{B}, \mu)\) is *complete* if every subset of a null set is in \( \mathcal{B} \), *i.e.* for all \( F \subseteq \Omega \) such that \( F \subseteq \mathcal{B} \) with \( \mu(E) = 0 \) implies that \( F \in \mathcal{B} \).

**Proposition 6.66 (Completion of a Measure).** Let \((\Omega, \mathcal{B}, \mu)\) be a measure space. Set

\[
\mathcal{N} := \{ N : N \subseteq \Omega \text{ such that } N \subseteq F \text{ and } \mu(F) = 0 \},
\]

\[
\mathcal{B} := \{ \{ A \cap N : A \in \mathcal{B} \text{ and } N \in \mathcal{N} \} \text{ and } \}
\]

\[
\mu(A \cap N) := \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N},
\]

see Fig. 6.2. Then \( \mathcal{B} \) is a \( \sigma \)–algebra, \( \bar{\mu} \) is a well defined measure on \( \mathcal{B} \), \( \bar{\mu} \) is the unique measure on \( \mathcal{B} \) which extends \( \mu \) on \( \mathcal{B} \), and \((\Omega, \mathcal{B}, \bar{\mu})\) is complete measure space. The \( \sigma \)-algebra, \( \mathcal{B} \), is called the completion of \( \mathcal{B} \) relative to \( \mu \) and \( \bar{\mu} \), is called the completion of \( \mu \).

**Proof.** Clearly \( \Omega, \emptyset \in \mathcal{B} \). Let \( A \in \mathcal{B} \) and \( N \in \mathcal{N} \) and choose \( F \in \mathcal{B} \) such that

\[
N = \bigcap_{n=1}^{\infty} (a_n, b_n] \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n] \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n].
\]

Fig. 6.2. Completing a \( \sigma \)–algebra.

that \( N \subseteq F \) and \( \mu(F) = 0 \). Since \( N^c = (F \setminus N) \cup F^c \),
\[(A \cup N)^c = A^c \cap N^c = A^c \cap (F \setminus N \cup F^c)\]
\[= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]\]

where \([A^c \cap (F \setminus N)] \in \mathcal{N}\) and \([A^c \cap F^c] \in \mathcal{B}\). Thus \(\mathcal{B}\) 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
\[\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \mathcal{B}\] 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, \(\mathcal{B}\) 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.

\[\blacksquare\]
Measure construction extras

The reader is advised that this chapter may and should be ignored.

7.1 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 ability measures with desirable properties. The main results of this chapter is the existence assertion of the theorem is contained in Theorem 6.20 (also restated in Theorem 7.4 below).

Theorem 7.1. A finitely additive finite measure \( \mu \) on an algebra, \( A \subset 2^\Omega \), extends to \( \sigma - \) additive measure on \( \sigma (A) \) iff \( \mu \) is a premeasure on \( A \). If the extension exists it is unique.

Proof. The uniqueness assertion is proved Proposition 6.55 below. The existence assertion of the theorem is contained in Theorem 6.20 (also restated in Theorem 7.4 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.

7.1.1 An Extension of Measure Strategy

Let us end this overview by briefly explaining the strategy we will use below for extending measures.

Example 7.2. It is easy to verify that to every finitely additive measure, \( \mu : A \to [0, \infty) \), the function \( d_\mu : A \times A \to [0, \infty) \) defined by

\[ d_\mu (A, B) := \mu (A \triangle B) = \int_\Omega |1_B - 1_A| \, d\mu \]

is a pseudo metric on \( A \). In general \( d_\mu (A, B) = 0 \) does not imply \( A = B \) but only that \( A = B \) modulo sets of \( \mu - \) measure zero.

Lemma 7.3. If \( \mu : A \to [0, \infty) \) is a finitely additive measure, then \( \mu \) is Lip-1 relative to \( d_\mu \), i.e.

\[ |\mu (B) - \mu (A)| \leq d_\mu (A, B) \quad \forall A, B \in A. \]

Proof. First proof. By the basic properties of the simple integral we find,

\[ |\mu (B) - \mu (A)| = \left| \int_\Omega (1_B - 1_A) \, d\mu \right| \leq \int_\Omega |1_B - 1_A| \, d\mu = \mu (B \triangle A) = d_\mu (A, B). \]

Second proof. Using the basic properties of measures we have,

\[ \mu (B) - \mu (A) = \mu (B \setminus A) + \mu (A \cap B) - [\mu (A \setminus B) + \mu (B \cap A)] = \mu (B \setminus A) - \mu (A \setminus B). \]

Basic inequalities then give

\[ |\mu (B) - \mu (A)| \leq \mu (B \setminus A) + \mu (A \setminus B) = \mu (B \triangle A) = d_\mu (A, B). \]

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 \( A \) and then 2) we extend \( \mu \) to a Lip-1 function \( (\bar{\mu}) \) on \( \bar{A}^d \) by continuity using Lemma 2.28. It will turn out that if we choose \( d \) sufficiently carefully (i.e. sufficiently “small”), then \( B := \bar{A}^d \) will be a \( \sigma - \) algebra and \( \bar{\mu} \) will be a measure on \( B \). The outcome of this strategy is summarized in the next theorem.

Theorem 7.4 (Finite premeasure extension theorem II.). Let \( (\Omega, A, \mu) \) be a premeasure space with \( \mu (\Omega) < \infty \) and define \( \mu^* : 2^\Omega \to [0, \mu (\Omega)] \) by

\[ \mu^* (B) = \inf \left\{ \sum_{n=1}^\infty \mu (A_n) : A_n \in A \text{ with } B \subset \bigcup_{n=1}^\infty A_n \right\}. \]

Further let \( B \subset 2^\Omega \) consist of those \( B \subset \Omega \) such that there exists \( \{A_n\} \subset A \) such that \( \lim_{n \to \infty} \mu^* (A_n \triangle B) = 0 \). Then;

1. \( B \) is a \( \sigma - \) algebra containing \( A \).
7.2 Monotone Class Theorems

In this section we record a theorem which is a cousin of the $\pi - \lambda$ theorem.

Definition 7.5 (Montone Class). $C \subset 2^{\Omega}$ is a monotone class if it is closed under countable increasing unions and countable decreasing intersections.

Lemma 7.6 (Monotone Class Theorem*). Suppose $A \subset 2^{\Omega}$ is an algebra and $C$ is the smallest monotone class containing $A$. Then $C = \sigma(A)$.

Proof. For $C \in C$ let

$$C(C) = \{B \in C : C \cap B, C \cap B^c, B \cap C^c \in C\},$$

then $C(C)$ is a monotone class. Indeed, if $B_n \in C(C)$ and $B_n \uparrow B$, then $B_n^c \downarrow B^c$ and so

$$C \ni C \cap B_n \uparrow C \cap B$$

then $C(C)$ is closed under increasing limits and a similar argument shows that $C(C)$ is closed under decreasing limits. Thus we have shown that $C(C)$ is a monotone class for all $C \in C$. If $A \subset C \subset A$, then $A \subset C \cap B_n, A \subset B_n \subset A$ and hence it follows that $C \subset C(A) \subset C$. Since $C$ is a monotone class containing $A$ and $C(A)$ is a monotone class containing $C$, we conclude that $C(A) = C$ for any $A \subset C$. Let $B \in C$ and notice that $A \subset C(B)$ happens iff $B \subset C(A)$. This observation and the fact that $C(A) = C$ for all $A \subset C \subset A$ implies $A \subset C(B) \subset C$ for all $B \in C$. Again since $C$ is the smallest monotone class containing $A$ and $C(B)$ is a monotone class we conclude that $C(B) = C$ for all $B \in C$. That is to say, if $A, B \subset C$ then $A \subset C = C(B)$ and hence $A \subset C \cap B, A \subset B \subset C, A^c \subset C \cap B$. So $C$ is closed under complements (since $\Omega \in C \subset C$) and finite intersections and increasing unions from which it easily follows that $C$ is a $\sigma$-algebra.

7.3 $\sigma$ - Algebra Regularity and Uniqueness Results

The goal of this appendix is 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 Carathéodory’s existence of measures proof. Nevertheless, we state these results again and give another somewhat independent proof.

Theorem 7.7 (Finite Regularity Result). Suppose $A \subset 2^{\Omega}$ is an algebra, $B = \sigma(A)$ and $\mu : B \rightarrow [0, \infty)$ is a finite measure, i.e. $\mu(\emptyset) < \infty$. Then for every $\varepsilon > 0$ and $B \in B$ there exists $A \in A_\varepsilon$ and $C \in A_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$.

Proof. Let $B_0$ denote the collection of $B \in B$ such that for every $\varepsilon > 0$ there here exists $A \in A_\varepsilon$ and $C \in A_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$. It is now clear that $A \subset B_0$ and that $B_0$ is closed under complementation. Now suppose that $B_i \in B_0$ for $i = 1, 2, \ldots$ and $\varepsilon > 0$ is given. By assumption there exists $A \in A_\varepsilon$ and $C \in A_\sigma$ such that $A \subset B_i \subset C$ and $\mu(C \setminus A_i) < 2^{-i} \varepsilon$.

Let $A := \cup_{n=1}^{\infty} A_i$, $A^N := \cup_{i=1}^{\infty} A_i \in A_\delta$, $B := \cup_{n=1}^{\infty} B_i$, and $C := \cup_{i=1}^{\infty} C_i \in A_\sigma$. Then $A^{-N} \subset A \subset B \subset C$ and

$$C \setminus A = \left[\cup_{i=1}^{\infty} C_i \setminus A \right] \subset \cup_{i=1}^{\infty} [C \setminus A_i] \subset \cup_{i=1}^{\infty} [C_i \setminus A_i].$$

Therefore,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^{\infty} [C_i \setminus A]) \leq \sum_{i=1}^{\infty} \mu(C_i \setminus A_i) \leq \sum_{i=1}^{\infty} \mu(C_i \setminus A_i) < \varepsilon.$$

Since $C \setminus A \subset C \setminus A_i$, it also follows that $\mu(C \setminus A \setminus A_i) < \varepsilon$ for sufficiently large $N$ and this shows $A = (\cup_{i=1}^{\infty} B_i \setminus B_0)$. Hence $B_0$ is a sub-$\sigma$-algebra of $B = \sigma(A)$ which contains $A$ which shows $B_0 = B$.

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 7.8. Suppose $\Omega$ is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^{\Omega}$ and $\mu : \mathcal{B} \rightarrow [0, \infty]$ 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 $(\Omega, \mu, \mathcal{B})$ 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 7.9 (σ – Finite Regularity Result). Theorem 7.7 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 \( \bigcup_{n=1}^{\infty} \Omega_n = \Omega \) and \( \mu(\Omega_n) < \infty \) for all \( n \). Since \( A \in \mathcal{B} \Rightarrow \mu_n(A) := \mu(\Omega_n \cap A) \) is a finite measure on \( A \in \mathcal{B} \) for each \( n \), by Theorem 7.7 for every \( B \in \mathcal{B} \) there exists \( C_n \in \mathcal{A}_n \) such that \( B \subseteq C_n \) and \( \mu(\Omega_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n} \varepsilon \). Now let \( C := \bigcup_{n=1}^{\infty} [\Omega_n \cap C_n] \in \mathcal{A}_\sigma \) and observe that \( B \cap C \) and

\[
\mu(C \setminus B) = \mu(\bigcup_{n=1}^{\infty} ([\Omega_n \cap C_n] \setminus B)) \\
\leq \sum_{n=1}^{\infty} \mu([\Omega_n \cap C_n] \setminus B) = \sum_{n=1}^{\infty} \mu(\Omega_n \cap [C_n \setminus B]) < \varepsilon.
\]

Applying this result to \( B^c \) shows there exists \( D \in \mathcal{A}_\sigma \) such that \( B^c \subseteq D \) and

\[
\mu(B \setminus D^c) = \mu(D \setminus B^c) < \varepsilon.
\]

So if we let \( A := D^c \in \mathcal{A}_\delta \), then \( A \subseteq B \subseteq C \) and

\[
\mu(C \setminus A) = \mu((B \setminus A) \cup [(C \setminus B) \setminus A]) \leq \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon
\]

and the result is proved.

Exercise 7.1. 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 7.10. 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 \( A \in \mathcal{A}_\delta \) and \( C \in \mathcal{A}_\sigma \) such that \( A \subseteq B \subseteq C \) and \( \mu(C \setminus A) = 0 \).

Proof. By Theorem 7.7 given \( B \in \mathcal{B} \), we may choose \( A_n \in \mathcal{A}_\delta \) and \( C_n \in \mathcal{A}_\sigma \) such that \( A_n \subseteq B \subseteq C_n \) and \( \mu(C_n \setminus B) \leq 1/n \) and \( \mu(B \setminus A_n) \leq 1/n \). By replacing \( A_N \) by \( \bigcup_{n=1}^{N} A_n \) 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 \( A = \bigcup A_n \in \mathcal{A}_\delta \) and \( C = \bigcap C_n \in \mathcal{A}_\sigma \), then \( A \subseteq B \subseteq C \) and

\[
\mu(C \setminus A) = \mu(C \setminus B) + \mu(B \setminus A) \leq \mu(C_n \setminus B) + \mu(B \setminus A_n) \\
\leq 2/n \to 0 \text{ as } n \to \infty.
\]

7.4 *Other characterizations of \( \mathcal{B} \)

This section may (probably should) be skipped on first reading. Its purpose is to serve as motivation for Carathéodory’s general construction theorem appearing in the optional Chapter 8.

Proposition 7.11 (*Other characterizations of \( \mathcal{B} \)). Let

\[ \mathcal{B} := \{ B \subset \Omega : \mu^*(E \cap B) + \mu^*(E \setminus B) = \mu^*(E) \forall E \in 2^\Omega \} \]

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;

\[ \mathcal{B}' := \{ B \subset \Omega : \mu^*(E \cap B) + \mu^*(E \setminus B) \leq \mu^*(E) \forall E \in 2^\Omega \}. \]

This is because the subadditivity of \( \mu^* \) automatically implies

\[ \mu^*(E) \leq \mu^*(E \cap B) + \mu^*(E \setminus B) \forall B, E \in 2^\Omega. \]

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 \( \mathcal{B}' \subseteq \mathcal{B}'' \). We will now complete the proof that \( \mathcal{B} = \mathcal{B}' = \mathcal{B}'' \) by showing \( \mathcal{B} \subseteq \mathcal{B}' \) and \( \mathcal{B}'' \subseteq \mathcal{B} \).

\( (\mathcal{B} \subset \mathcal{B}') \). If \( B \in \mathcal{B}, E \in 2^\Omega, \) and \( C \in \mathcal{A}_\sigma \) such that \( E \subset C \), then

\[
\mu^*(E \cap B) + \mu^*(E \setminus B) \leq \mu^*(C \cap B) + \mu^*(C \setminus B) \\
= \bar{\mu}(C \cap B) + \bar{\mu}(C \setminus B) = \bar{\mu}(C) = \mu_\sigma(C).
\]

Taking the infimum over all \( C \in \mathcal{A}_\sigma \) such that \( E \subset C \) shows

\[ \mu^*(E \cap B) + \mu^*(E \setminus B) \leq \mu^*(E) \]

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 \( B \subset C, B^c \subset D \),

\[ \bar{\mu}(C) = \mu_\sigma(C) \leq \mu^*(B) + \varepsilon, \text{ and } \bar{\mu}(D) = \mu_\sigma(D) \leq \mu^*(B^c) + \varepsilon. \]

Summing these inequalities while using \( B \in \mathcal{B}' \) implies,

\[ \bar{\mu}(C) + \bar{\mu}(D) \leq \mu(\Omega) + 2\varepsilon \implies \bar{\mu}(C) \leq \bar{\mu}(D^c) + 2\varepsilon. \]

As \( D^c \subset B \subset C \) and \( C \setminus D^c = C \cap D \in \mathcal{A}_\sigma \), it follows that
7 *Measure construction extras

\[ \mu^* (C \setminus D^c) = \mu_\sigma (C \setminus D^c) = \mu (C \setminus D^c) \leq 2\varepsilon. \]

Consequently we conclude that,

\[ d (B, C) = \mu^* (C \setminus B) \leq \mu^* (C \setminus D^c) \leq 2\varepsilon \quad \text{and} \]
\[ d (B, D^c) = \mu^* (B \setminus D^c) \leq \mu^* (C \setminus D^c) \leq 2\varepsilon. \]

Since \( \varepsilon > 0 \) is arbitrary these equations show \( B \in \mathcal{A}_d^d = \mathcal{B} \) and also that \( B \in \mathcal{A}_d^d \). Moreover, since \( \bar{\mu} = \mu_\sigma \) on \( \mathcal{A}_\sigma \) and \( \bar{\mu} \) is Lip–1 relative to \( d \), we also find,

\[ |\bar{\mu} (B) - \mu^* (B)| \leq |\bar{\mu} (B) - \bar{\mu} (C)| + |\mu_\sigma (C) - \mu^* (B)| \leq 2\varepsilon + \varepsilon = 3\varepsilon. \]

Again as \( \varepsilon > 0 \) is arbitrary we have also shown that \( \bar{\mu} = \mu^* \) on \( B \).

\textbf{Notation 7.12 (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 \textit{inner measure} (or \textit{inner content}) of \( B \).

\textbf{Remark 7.13.} 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 7.12.

\textbf{Corollary 7.14.} 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 7.12 and Proposition 6.6 and Proposition 6.14 respectively. Then (continuing to use the notation in Theorem 6.20)

\[ \mathcal{B} := \mathcal{A}_d^d = \left\{ B \subset \Omega : \mu_* (B) = \mu^* (B) \right\}. \]

\textbf{Proof.} From Proposition 7.11 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 \( A_n := C_n \cap A \) with \( A_n \subset B \), then

\[ \mu^* (B) = \mu (\Omega) - \mu^* (B^c) = \lim_{n \to \infty} [\mu (\Omega) - \bar{\mu} (C_n)] \]
\[ = \lim_{n \to \infty} \bar{\mu} (A_n) \leq \mu_* (B). \]

As \( \mu_* (B) \leq \mu^* (B) \) 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 \( d (B, C_n) = \mu^* (C_n \setminus B) \leq \mu^* (C_n \setminus A_n) \to 0 \) as \( n \to \infty \) which shows \( B \in \mathcal{A}_d^d = \mathcal{B} \).

\textbf{7.5 * Appendix: Alternate measure extension construction}

(The reader wanting for more motivation of the construction measure in this section may wish to read Section 7.3 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 \( \mu (A) = \mu (A) + \mu (A^c) \). From this observation we may extend \( \mu \) to a function on \( \mathcal{A}_\delta \cup \mathcal{A}_\sigma \) by defining

\[ \mu (A) := \mu (\Omega) - \mu (A^c) \quad \text{for all } A \in \mathcal{A}_\delta. \]

\textbf{Lemma 7.15.} 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 ?? and Eq. ?? 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 \( \mu (A) = \lim_{n \to \infty} \mu (A_n) \).
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 \( \mu (C \setminus A) = \mu (C) - \mu (A) \).

\textbf{Proof.} 

1. Since \( \mu^* (B) = \mu (\Omega) - \mu (B^c) \) and \( A \subset B \iff B^c \subset A^c \), it follows that

\[ \inf \{ \mu (B) : A \subset B \in \mathcal{A} \} = \inf \{ \mu (\Omega) - \mu (B^c) : A \supseteq B^c \subset A^c \} = \mu (\Omega) - \sup \{ \mu (B) : A \supseteq B \subset A^c \} = \mu (\Omega) - \mu (A^c) = \mu (A). \]

2. Similarly, since \( A_n^c \uparrow A^c \in \mathcal{A}_\sigma \), by the definition of \( \mu (A) \) and Proposition ?? it follows that

\[ \mu (A) = \mu (\Omega) - \mu (A^c) = \mu (\Omega) - \upto \lim_{n \to \infty} \mu (A_n^c) \]
\[ = \lim_{n \to \infty} [\mu (\Omega) - \mu (A_n^c)] = \lim_{n \to \infty} \mu (A_n). \]

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,
\[ \mu(A \cup B) + \mu(A \cap B) = \lim_{n \to \infty} \left[ \mu(A_n \cup B_n) + \mu(A_n \cap B_n) \right] = \lim_{n \to \infty} \left[ \mu(A_n) + \mu(B_n) \right] = \mu(A) + \mu(B). \]

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 \( \mathcal{A}_\delta \), it follows that
\[ \mu(\Omega) + \mu(A^c \cap B^c) = \mu(A^c) + \mu(B^c) \]
from which it follows that
\[ \mu(A \cup B) = \mu(\Omega) - \mu(A^c \cap B^c) = \mu(\Omega) - \left[ \mu(A^c) + \mu(B^c) - \mu(\Omega) \right] = \mu(A) + \mu(B). \]

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 \[7.12\] and \[7.13\] respectively) it is reasonable to define the measure of \( B \) as this common value. As we will see in Theorem \[7.18\] below, this extension becomes a \( \sigma \)-additive measure on a \( \sigma \)-algebra of subsets of \( \Omega \).

**Definition 7.16 (Measurable Sets).** Suppose \( \mu \) is a finite premeasure on an algebra \( \mathcal{A} \subset 2^\Omega \). We say that \( \mathcal{B} \subset \mathcal{A} \) is measurable if \( \mu^*(B) = \mu^*(B) \) where \( \mu^*(B) \) and \( \mu^*(B) \) as in Notations \[7.12\] and \[7.13\] 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
\[ \bar{\mu}(B) := \mu^*(B) = \mu^*(B) \] for all \( B \in \mathcal{B} \).

**Remark 7.17.** 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
\[ \mu(C \setminus A) = \mu(C) - \mu(A) > \varepsilon, \]
wherein we have used Lemma \[7.15\] 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
\[ \mu(A) \leq \mu^*(B) = \bar{\mu}(B) = \mu^*(B) \leq \mu(C). \]

**Theorem 7.18 (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 \[7.16\]. Then \( \mathcal{B} \) is a \( \sigma \)-algebra on \( \Omega \) which contains \( \mathcal{A} \) and \( \mu \) 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. \[7.6\] 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
\[ C \setminus A = (C_1 \setminus A_1) \cup (C_2 \setminus A_2) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2), \]
it follows from the sub-additivity of \( \mu \) that
\[ \mu(C \setminus A) \leq \mu(C_1 \setminus A_1) + \mu(C_2 \setminus A_2) < 2\varepsilon. \]
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 now know \( \mathcal{B} \) is an algebra, to show \( \mathcal{B} \) is a \( \sigma \)-algebra it suffices to show that \( B = \sum_{i=1}^\infty B_i \in \mathcal{B} \) 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 \( \mu(C_i \setminus A_i) < 2^{-i} \) for all \( i \). Let \( C := \bigcup_{i=1}^\infty C_i \in \mathcal{A}_\sigma \) and for \( n \in \mathbb{N} \) let \( A_n := \bigcup_{i=1}^n A_i \in \mathcal{A}_\delta \). Since the \( \{A_i\}_{i=1}^\infty \) are pairwise disjoint we may use Lemma \[7.15\] to show,
\[ \sum_{i=1}^n \mu(C_i) = \sum_{i=1}^n (\mu(A_i) + \mu(C_i \setminus A_i)) \]
\[ = \mu(A^n) + \sum_{i=1}^n \mu(C_i \setminus A_i) \leq \mu(\Omega) + \sum_{i=1}^\infty 2^{-i} \]
which on letting \( n \to \infty \) implies
\[ \sum_{i=1}^\infty \mu(C_i) \leq \mu(\Omega) + \varepsilon < \infty. \]
Using
\[ C \setminus A^n = \bigcup_{i=1}^\infty (C_i \setminus A^n) \subset \bigcup_{i=1}^n (C_i \setminus A_i) \cup \bigcup_{i=n+1}^\infty C_i \in \mathcal{A}_\sigma \]
and the sub-additivity of \( \mu \) on \( \mathcal{A}_\sigma \) it follows that
\[
\mu(C \setminus A^n) \leq \sum_{i=1}^{\infty} \mu(C_i \setminus A_i) + \sum_{i=n+1}^{\infty} \mu(C_i) \leq \varepsilon \sum_{i=1}^{\infty} 2^{-i} + \sum_{i=n+1}^{\infty} \mu(C_i)
\]
\[
\leq \varepsilon + \sum_{i=n+1}^{\infty} \mu(C_i) \to \varepsilon \text{ as } n \to \infty,
\]

wherein we have used Eq. (7.8) in computing the limit. In summary, \(B = \bigcup_{i=1}^{\infty} B_i\), \(A_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma\), \(C \setminus A^n \in \mathcal{A}_\sigma\) with \(\mu(C \setminus A^n) \leq 2\varepsilon\) 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
\[
\sum_{i=1}^{\infty} \mu(A_i) \nrightarrow \sum_{i=1}^{\infty} \mu(A_i) = \mu(A^n) \leq \bar{\mu}(B) \leq \mu(C) \leq \sum_{i=1}^{\infty} \mu(C_i). \tag{7.9}
\]
On the other hand, since \(A_i \subset B_i \subset C_i\), it follows (see Eq. 7.7) that \(\mu(A_i) \leq \bar{\mu}(B_i) \leq \mu(C_i)\) and therefore that
\[
\sum_{i=1}^{\infty} \mu(A_i) \leq \sum_{i=1}^{\infty} \bar{\mu}(B_i) \leq \sum_{i=1}^{\infty} \mu(C_i). \tag{7.10}
\]
Equations (7.9) and (7.10) 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
\[
\left| \bar{\mu}(B) - \sum_{i=1}^{\infty} \bar{\mu}(B_i) \right| \leq \sum_{i=1}^{\infty} \mu(C_i) - \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{\infty} \mu(C_i \setminus A_i) \leq \sum_{i=1}^{\infty} \varepsilon 2^{-i} = \varepsilon.
\]
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 6.2 below.

**Corollary 7.19.** 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 \(A_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma\) and \(\varepsilon > 0\) such that \(\mu(C \setminus A) < \varepsilon\) and
2. there exists \(A \in \mathcal{A}\) such that \(\mu(A \triangle B) < \varepsilon\).

**Exercise 7.2.** Prove corollary 7.19 by considering \(\bar{\nu}\) where \(\nu := \mu|\mathcal{A}\). **Hint:** you may find Exercise 5.6 useful here.
**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 results here in the remainder of this book.

### 8.1 General Extension and Construction Theorem

Proposition 7.11 motivates the following definition.

**Definition 8.1.** Let \( \mu \) be an outer measure, see Definition 6.5.

Define the \( \mu \)-measurable sets to be

\[
\mathcal{M}(\mu) := \{ B \subset \Omega : \mu^*(E) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \quad \forall \ E \subset \Omega \}. 
\]

Because of the sub-additivity of \( \mu^* \), we may equivalently define \( \mathcal{M}(\mu^*) \) by

\[
\mathcal{M}(\mu^*) = \{ B \subset \Omega : \mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c) \quad \forall \ E \subset \Omega \}. 
\] (8.1)

**Theorem 8.2 (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

\[
\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) = \mu^*(E \cap A \cap B) + \mu^*((E \cap A) \setminus B) + \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B). 
\] (8.2)

By the sub-additivity of \( \mu^* \) and the set identity,

\[
E \cap (A \cup B) = (E \cap A) \cup (E \cap B) = [(E \cap A) \setminus B] \cup (E \cap A \cap B) \cup [(E \cap B) \setminus A] \cup (E \cap A \cap B) = [E \cap A \cap B] \cup [(E \cap A) \setminus B] \cup [(E \cap A) \setminus B],
\]

we have

\[
\mu^*(E \cap A \cap B) + \mu^*((E \cap A) \setminus B) + \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B) \geq \mu^*(E \cap (A \cup B)).
\]

Using this inequality in Eq. (8.3) shows

\[
\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B))
\] (8.4)

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. (8.2) 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 \( B = \bigcup_{j=1}^\infty A_j \), then for \( E \subset \Omega \) we have

\[
\mu^*(E \cap B_n) = \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) = \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}),
\]

and so by induction,

\[
\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k).
\] (8.5)

Therefore we find that

\[
\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) = \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap B_n^c) \geq \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap B_n^c),
\]

where the last inequality is a consequence of the monotonicity of \( \mu^* \) and the fact that \( B_n^c \subset B_n^c \). Letting \( n \to \infty \) in this equation shows that
8. Carathéodory’s Construction of Measures

\[ \mu^*(E) \geq \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap B^c) \]
\[ \geq \mu^*(\bigcup_k (E \cap A_k)) + \mu^*(E \setminus B) \]
\[ = \mu^*(E \cap B) + \mu^*(E \setminus B) \geq \mu^*(E), \]

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. (8.5) to find

\[ \mu^*(E) \geq \sum_{k=1}^{\infty} \mu^*(E \cap A_k). \]

Letting \( E = B = \bigcup A_k \) 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 \( N \in \mathcal{M} \).

8.2 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 8.3.** 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. (8.1) 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 \( \nu \) is \( \mu \) 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 \( \mathcal{M} = \sigma(\mathcal{A})[\bar{\mu}|_{\sigma(\mathcal{A})}] \).

**Proof.** Recall from Proposition ?? 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 \( E \subset B := \bigcup_{j=1}^{\infty} B_j \) and

\[ \mu^*(E) + \varepsilon \geq \mu(B) = \sum_{j=1}^{\infty} \mu(B_j). \]

Since \( A \cap E \subset \bigcup_{j=1}^{\infty} (B_j \cap A^c) \) and \( E \cap A^c \subset \bigcup_{j=1}^{\infty} (B_j \cap A^c) \), using the sub-additivity of \( \mu^* \) and the additivity of \( \mu \) on \( \mathcal{A} \) we have,

\[ \mu^*(E) + \varepsilon \geq \sum_{j=1}^{\infty} \mu(B_j) = \sum_{j=1}^{\infty} [\mu(B_j \cap A) + \mu(B_j \cap A^c)] \]
\[ \geq \mu^*(E \cap A) + \mu^*(E \cap A^c). \]

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 [8.2]

3. If \( A := \bigcup_{j=1}^{\infty} A_j \) with \( \{A_j\} \subset \mathcal{A} \) 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

\[ \nu(B) \leq \inf \{\nu(A) : B \subset A \in \mathcal{A}_\sigma\} = \inf \{\mu(A) : B \subset A \in \mathcal{A}_\sigma\} = \mu(B). \]

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. (8.6), 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 \( \Omega_j \subset \mathcal{M} \) such that \( \Omega_j \uparrow \Omega \) and \( \mu^*(\Omega_j) < \infty \) then

\[ \bar{\mu}(B) = \lim_{j \to \infty} \mu(B \cap \Omega_j) = \lim_{j \to \infty} \nu(B \cap \Omega_j) = \nu(B). \]
Theorem 8.4 (Regularity Theorem). Suppose that \( \mu \) is a \( \sigma \)-finite pre-measure on an algebra \( A \), \( \tilde{\mu} \) is the extension described in Theorem 6.3 and \( B \in M := M(\mu^*) \). Then:

1. \( \tilde{\mu}(B) := \inf \{ \tilde{\mu}(C) : B \subset C \in A_\sigma \} \).
2. For any \( \varepsilon > 0 \) there exists \( A \subset B \subset C \) such that \( A \in A_\delta \), \( C \in A_\sigma \) and \( \tilde{\mu}(C \setminus A) < \varepsilon \).
3. There exists \( A \subset B \subset C \) such that \( A \in A_\delta \), \( C \in A_\sigma \) and \( \tilde{\mu}(C \setminus A) = 0 \).
4. The \( \sigma \)-algebra, \( \tilde{M} \), is the completion of \( \sigma(A) \) with respect to \( \tilde{\mu}|_{\sigma(A)} \).

Proof. The proofs of items 1.-3. are the same as the proofs of the corresponding results in Theorem 7.4, Corollary 7.9, and Theorem 6.28 and so will be omitted. Moreover, item 4. is a simple consequence of item 3. and Proposition 6.60.

The following proposition shows that measures may be “restricted” to non-measurable sets.

Proposition 8.5. Suppose that \( (\Omega, M, \mu) \) is a probability space and \( \Omega_0 \subset \Omega \) is any set. Let \( M_{\Omega_0} := \{ A \cap \Omega_0 : A \in M \} \) and set \( P(A \cap \Omega_0) := \mu^*(A \cap \Omega_0) \). Then \( P \) is a measure on the \( \sigma \)-algebra \( M_{\Omega_0} \). Moreover, if \( \mu^* \) is the outer measure generated by \( P \), then \( P^*(A) = \mu^*(A) \) for all \( A \in \Omega_0 \).

Proof. Let \( A, B \in M \) such that \( A \cap B = \emptyset \). Then since \( A \in M \subset M(\mu^*) \) it follows from Eq. 8.3 with \( E := (A \cup B) \cap \Omega_0 \) that

\[
\mu^*((A \cup B) \cap \Omega_0) = \mu^*((A \cup B) \cap \Omega_0 \cap A) + \mu^*((A \cup B) \cap \Omega_0 \cap A^c)
\]

\[
= \mu^*(\Omega_0 \cap A) + \mu^*(B \cap \Omega_0)
\]

which shows that \( P \) is finitely additive. Now suppose \( A = \bigcup_{j=1}^\infty A_j \) with \( A_j \in M \) and let \( B_n := \bigcup_{j=n+1}^\infty A_j \in M \). 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 subadditivity 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 \( P = \mu^*|_{M_{\Omega_0}} \) is a measure. Now let \( P^* \) be the outer measure generated by \( P \). For \( A \in \Omega_0 \), we have

\[
P^*(A) = \inf \{ P(B) : A \subset B \in M_{\Omega_0} \}
\]

\[
= \inf \{ P(B \cap \Omega_0) : A \subset B \in M \}
\]

\[
= \inf \{ \mu^*(B \cap \Omega_0) : A \subset B \in M \}
\]

and since \( \mu^*(B \cap \Omega_0) \leq \mu^*(B) \),

\[
P^*(A) \leq \inf \{ \mu^*(B) : A \subset B \in M \}
\]

\[
= \inf \{ \mu(B) : A \subset B \in M \} = \mu^*(A).
\]

On the other hand, for \( A \subset B \in M \), we have \( \mu^*(A) \leq \mu^*(B \cap \Omega_0) \) and therefore by Eq. (8.7)

\[
\mu^*(A) \leq \inf \{ \mu^*(B \cap \Omega_0) : A \subset B \in M \} = P^*(A).
\]

and we have shown

\[
\mu^*(A) \leq P^*(A) \leq \mu^*(A).
\]

8.3 More Motivation of Carathéodory’s Construction

Theorem 8.2

The next Proposition helps to motivate this definition and the Carathéodory’s construction Theorem 8.2.

Proposition 8.6. Suppose \( E = M \) is a \( \sigma \)-algebra, \( \rho = \mu : M \rightarrow [0, \infty] \) is a measure and \( \mu^* \) is defined as in Eq. 6.4. Then

1. For \( A \subset X \)

\[
\mu^*(A) = \inf \{ \mu(B) : B \in M \text{ and } A \subset B \}.
\]

In particular, \( \mu^* = \mu \) on \( M \).

2. Then \( M \subset M(\mu^*) \), i.e. if \( A \in M \) and \( E \subset X \) then

\[
\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c).
\]

3. Assume further that \( \mu \) is \( \sigma \)-finite on \( M \), then \( M(\mu^*) = \tilde{M} = M^\mu \) and \( \mu^*|_{M(\mu^*)} = \tilde{\mu} \) where \( (\tilde{M} = M^\mu, \tilde{\mu}) \) is the completion of \( (M, \mu) \).
Proof. Item 1. If $E_i \in \mathcal{M}$ such that $A \subset \bigcup E_i = B$ and $\tilde{E}_i = E_i \setminus (E_i \cup \cdots \cup E_{i-1})$ then
$$\sum \mu(E_i) \geq \sum \mu(\tilde{E}_i) = \mu(B)$$
so
$$\mu^*(A) \leq \sum \mu(\tilde{E}_i) = \mu(B) \leq \sum \mu(E_i).$$

Therefore, $\mu^*(A) = \inf \{\mu(B) : B \in \mathcal{M} \text{ and } A \subset B \}$.

Item 2. If $\mu^*(E) = \infty$ Eq. (8.8) 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
$$\mu^*(E) + \varepsilon \geq \mu(B) = \mu(B \cap A) + \mu(B \cap A^c)$$
$$\geq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

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,
$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \forall E \in 2^X. \quad (8.9)$$

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. (8.9) 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
$$A = D \cup [(B \setminus D) \cap A]$$
where $D \in \mathcal{M}$ and $(B \setminus D) \cap A \in \mathcal{N}$ showing that $A \in \mathcal{M}$ 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,
Random Variables

Notation 9.1 If \( f : X \to Y \) is a function and \( \mathcal{E} \subset 2^Y \) let
\[
\sigma(f^{-1}(\mathcal{E})) := f^{-1}(\sigma(\mathcal{E})).
\]
If \( \mathcal{G} \subset 2^X \), let
\[
f_*\mathcal{G} := \{ A \in 2^Y : f^{-1}(A) \in \mathcal{G} \}.
\]

Definition 9.2. Let \( \mathcal{E} \subset 2^X \) be a collection of sets, \( A \subset X \), \( i_A : A \to X \) be the inclusion map \((i_A(x) = x \text{ for all } x \in A)\) 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 9.3 (A key measurability lemma). If \( f : X \to Y \) is a function and \( \mathcal{E} \subset 2^Y \), then
\[
\sigma(f^{-1}(\mathcal{E})) = f^{-1}\sigma(\mathcal{E}). \tag{9.1}
\]
In particular, if \( A \subset Y \), then
\[
(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{9.2}
\]
(Similar assertion hold with \( \sigma(\cdot) \) being replaced by \( \mathcal{A}(\cdot, \cdot) \).

Proof. Since \( \mathcal{E} \subset \sigma(\mathcal{E}) \), it follows that \( f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E})) \). Moreover, by Exercise 9.1 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
\[
\mathcal{M} := \{ B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E})) \} = f_\ast \sigma(f^{-1}(\mathcal{E})).
\]
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 9.1) which contains \( \mathcal{E} \) and therefore \( \sigma(\mathcal{E}) \subset \mathcal{M} \).

Equation (9.2) is a special case of Eq. (9.1). 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 9.1. 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 \( f_\ast \mathcal{B} \) are \( \sigma \) - algebras (algebras).

Example 9.4. Let \( \mathcal{E} = \{ (a,b) : -\infty < a < b < \infty \} \) and \( \mathcal{B} = \sigma(\mathcal{E}) \) be the Borel \( \sigma \) - field on \( \mathbb{R} \). Then
\[
\mathcal{E}_{[0,1]} = \{ (a,b) : 0 \leq a < b \leq 1 \}
\]
and we have
\[
\mathcal{B}_{[0,1]} = \sigma(\mathcal{E}_{[0,1]}).
\]
In particular, if \( A \in \mathcal{B} \) such that \( A \subset (0,1] \), then \( A \in \sigma(\mathcal{E}_{[0,1]}) \).

Exercise 9.2. Given \( x \in \mathbb{R} \setminus \{0\} \) let
\[
x + B := \{ x + y : y \in B \} \quad \text{and} \quad x : B := \{ xy : y \in B \}. \tag{9.3}
\]
Show \( x + B \) and \( x : B \) are in \( \mathcal{B}_R \) for all \( B \in \mathcal{B}_R \). Hint: take \( \mathcal{E} = \{ (a,b) : -\infty < a < b < \infty \} \) and apply Lemma 9.3 with \( f(y) = y - x \) and \( f(y) = y/x \) respectively.

9.1 Measurable Functions

Definition 9.5. 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 \( \int f \, d\mu \) as a certain limit of sums of the form,
\[
\sum_{0 < a_1 < a_2 < a_3 < \ldots < a_i < a_{i+1}} \infty a_i \mu(f^{-1}(a_i, a_{i+1})].
\]
For this to make sense we will need to require \( f^{-1}(a,b) \in \mathcal{M} \) for all \( a < b \). Because of Corollary 9.11 below, this last condition is equivalent to the condition \( f^{-1}(\mathcal{B}_R) \subset \mathcal{M} \).
Definition 9.6. Let \((X, \mathcal{M})\) and \((Y, \mathcal{F})\) be measurable spaces. A function \(f : X \rightarrow Y\) is measurable if \(f^{-1} \mathcal{F} \subset \mathcal{M}\), i.e. if \(f^{-1}(A) \in \mathcal{M}\) for all \(A \in \mathcal{F}\).

Example 9.8 (Indicator Functions). Let \((X, \mathcal{M})\) be a measurable space and \(A \subset X\). Then \(1_A = (\mathcal{M}, \mathcal{B}_X)\) is 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 9.9. Suppose \(f : X \rightarrow Y\) 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 9.10. 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, \(f : X \rightarrow Y\) 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 \(\sigma(f^{-1}(\mathcal{E})) \subset \mathcal{M}\) and so making use of Lemma 9.3 \(f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma(f^{-1}(\mathcal{E})) \subset \mathcal{M}\).

Corollary 9.11. Suppose that \((X, \mathcal{M})\) is a measurable space. Then the following conditions on a function \(f : X \rightarrow \mathbb{R}\) are equivalent:

1. \(f\) is \((\mathcal{M}, \mathcal{B}_R)\) 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 9.3 (Look at but do not hand in). Prove Corollary 9.11 i.e. suppose that \((X, \mathcal{M})\) is a measurable space. Then the following conditions on a function \(f : X \rightarrow \mathbb{R}\) are equivalent:

1. \(f\) is \((\mathcal{M}, \mathcal{B}_R)\) 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 \(a \in \mathbb{R}\).

Hint: See Exercise 4.3

Exercise 9.4. 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 9.5. Let \((X, \mathcal{M})\) be a measure space and \(f_n : X \rightarrow \mathbb{R}\) be a sequence of measurable functions on \(X\). Show that \(\{x : \lim_{n \rightarrow \infty} f_n(x) \text{ exists in } \mathbb{R}\} \in \mathcal{M}\). Similarly show the same holds if \(\mathbb{R}\) is replaced by \(\mathbb{C}\).

Exercise 9.6. Show that every monotone function \(f : \mathbb{R} \rightarrow \mathbb{R}\) is \((\mathcal{B}_R, \mathcal{B}_R)\) measurable.

Definition 9.12. Given measurable spaces \((X, \mathcal{M})\) and \((Y, \mathcal{F})\) and a subset \(A \subset X\). We say a function \(f : A \rightarrow Y\) is measurable iff \(f\) is \(\mathcal{M}_A/\mathcal{F}\) measurable.

Proposition 9.13 (Localizing Measurability). Let \((X, \mathcal{M})\) and \((Y, \mathcal{F})\) be measurable spaces and \(f : X \rightarrow Y\) a function.

1. If \(f\) is measurable and \(A \subset X\) then \(f|_A : A \rightarrow Y\) is \(\mathcal{M}_A/\mathcal{F}\) measurable.
2. Suppose there exist \(A_n \in \mathcal{M}\) such that \(X = \bigcup_{n=1}^{\infty} A_n\) and \(f|_{A_n}\) is \(\mathcal{M}_{A_n}/\mathcal{F}\) measurable for all \(n\), then \(f\) is \(\mathcal{M}\) measurable.

Proof. 1. If \(f : X \rightarrow Y\) is measurable, \(f^{-1}(B) \in \mathcal{M}\) for all \(B \in \mathcal{F}\) and therefore \(f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A\) for all \(B \in \mathcal{F}\).

2. If \(B \in \mathcal{F}\), then \(f^{-1}(B) = \bigcup_{n=1}^{\infty} (f^{-1}(B) \cap A_n) = \bigcup_{n=1}^{\infty} f|_{A_n}^{-1}(B)\).

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 9.14 (Composing Measurable Functions). Suppose that \((X, \mathcal{M})\), \((Y, \mathcal{F})\) and \((Z, \mathcal{G})\) are measurable spaces. If \(f : (X, \mathcal{M}) \rightarrow (Y, \mathcal{F})\) and \(g : (Y, \mathcal{F}) \rightarrow (Z, \mathcal{G})\) are measurable functions then \(g \circ f : (X, \mathcal{M}) \rightarrow (Z, \mathcal{G})\) is measurable as well.

Proof. By assumption \(g^{-1}(\mathcal{G}) \subset \mathcal{F}\) and \(f^{-1}(\mathcal{F}) \subset \mathcal{M}\) so that \((g \circ f)^{-1}(\mathcal{G}) = f^{-1}(g^{-1}(\mathcal{G})) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}\)
Definition 9.15 (σ – 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.
\[
\sigma(f_\alpha : \alpha \in I) = \sigma(\{f_\alpha^{-1}(F) : F \in \mathcal{F}_\alpha\}).
\]

Example 9.16. Suppose that \(Y\) is a finite set, \(\mathcal{F} = 2^Y\), and \(X = Y^N\) for some \(N \in \mathbb{N}\). Let \(\pi \in Y^N \to Y\) be the projection maps, \(\pi_i(y_1, \ldots, y_N) = y_i\). Then, as the reader should check,
\[
\sigma(\pi_1, \ldots, \pi_n) = \{A \times A^{N-n} : A \subset A^n\}.
\]

Proposition 9.17. Assuming the notation in Definition 9.15 (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 \(f_\alpha \circ g : \mathcal{M} \to A\) is \((\mathcal{A}, \mathcal{F}_\alpha)\) – measurable for all \(\alpha \in I\).

Proof. (⇒) 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 9.14. (⇐) Since \(\sigma(f_\alpha : \alpha \in I) = \sigma(\mathcal{E})\) where \(\mathcal{E} := \bigcup_{\alpha \in I} f_\alpha^{-1}(\mathcal{F}_\alpha)\), according to Proposition 9.10 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 \(g^{-1}(A) = g^{-1}(f_\alpha^{-1}(B)) = (f_\alpha \circ g)^{-1}(B) \in \mathcal{M}\) because \(f_\alpha \circ g : Z \to Y_\alpha\) is assumed to be measurable.

Definition 9.18. If \(\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}\) is a collection of measurable spaces, then the product measure space, \((Y, \mathcal{F})\), is \(Y := \prod_{\alpha \in I} Y_\alpha\) 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 9.17.

Corollary 9.19. 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 \(A = \{1, 2, \ldots, n\}\), then \(Y = Y_1 \times \cdots \times Y_n\) and \(g = (g_1, \ldots, g_n) : Z \to Y\) is measurable if each component, \(g_i : Z \to Y_i\), is measurable. Here is another closely related result.

Proposition 9.20. 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 9.10 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 9.7. 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 \(Y_1 \in \mathcal{E}_1\) and \(\mathcal{F}_i = \sigma(\mathcal{E}_i)\) for \(i = 1, 2\). 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 \(\sigma(\mathcal{S}_1) \cup \sigma(\mathcal{S}_2) = \sigma(\mathcal{S}_1 \cup \mathcal{S}_2)\). (In fact, one has that \(\sigma(\bigcup_{\alpha \in I} \sigma(S_\alpha)) = \sigma(\bigcup_{\alpha \in I} S_\alpha)\) for any collection of non-empty subsets, \(\{S_\alpha\}_{\alpha \in I} \subset 2^Y\).)

2. After this you might start your proof as follows:

\[
\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\pi_1^{-1}(\mathcal{F}_1) \cup \pi_2^{-1}(\mathcal{F}_2)) = \sigma(\pi_1^{-1}(\mathcal{E}_1) \cup \pi_2^{-1}(\mathcal{E}_2)) = \ldots
\]

Remark 9.21. The reader should convince herself that Exercise 9.7 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
\[
\otimes_{i \in I} \mathcal{E}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j\text{ for all }j \in I\right\}\right)
\]
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 9.7.

Exercise 9.8 (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 9.7 with \(\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{F}_i\text{ for }i = 1, 2\}\).

Definition 9.22. A function \(f : X \to Y\) between two topological spaces is Borel measurable if \(f^{-1}(B_Y) \subset B_X\).
**Proposition 9.23.** 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 9.3 and $B_Y = \sigma(\tau_Y)$,

$$f^{-1}(B_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = B_X.$$

\[\square\]

**Example 9.24.** For $i = 1, 2, \ldots, n$, 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 $B_{\mathbb{R}^n}$ measurable.

**Lemma 9.25.** Let $E$ denote the collection of open rectangle in $\mathbb{R}^n$, then $B_{\mathbb{R}^n} = \sigma(E)$. We also have that $B_{\mathbb{R}^n} = \sigma(\pi_1, \ldots, \pi_n) = B_{\mathbb{R}} \otimes \cdots \otimes B_{\mathbb{R}}$ and in particular, $A_1 \times \cdots \times A_n \in B_{\mathbb{R}^n}$ whenever $A_i \in B_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$. Therefore $B_{\mathbb{R}^n}$ may be described as the $\sigma$ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in B_{\mathbb{R}}\}$. (Also see Remark 9.27.)

**Proof.** Assertion 1. Since $E \subset B_{\mathbb{R}^n}$, it follows that $\sigma(E) \subset B_{\mathbb{R}^n}$. Let

$$E_0 := \{(a, b) : a, b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a, b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1, 2, \ldots, n$ and let

$$(a, b) = (a_1, b_1) \times \cdots \times (a_n, b_n).$$

(9.4)

Since every open set, $V \subset \mathbb{R}^n$, may be written as a (necessarily) countable union of elements from $E_0$, we have

$$V \in \sigma(E_0) \subset \sigma(E),$$

i.e. $\sigma(E_0)$ and hence $\sigma(E)$ contains all open subsets of $\mathbb{R}^n$. Hence we may conclude that

$$B_{\mathbb{R}^n} = \sigma(\text{open sets}) \subset \sigma(E_0) \subset \sigma(E) \subset B_{\mathbb{R}^n}.$$

**Assertion 2.** Since each $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous, it is $B_{\mathbb{R}^n}$ measurable and therefore, $\sigma(\pi_1, \ldots, \pi_n) \subset B_{\mathbb{R}^n}$. Moreover, if $(a, b)$ is as in Eq. (9.4), then

$$(a, b) = \bigcap_{i=1}^n (a_i, b_i)^{-1} \in \sigma(\pi_1, \ldots, \pi_n).$$

Therefore, $E \subset \sigma(\pi_1, \ldots, \pi_n)$ and $B_{\mathbb{R}^n} = \sigma(E) \subset \sigma(\pi_1, \ldots, \pi_n)$.

**Assertion 3.** If $A_i \in B_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$, then

$$A_1 \times \cdots \times A_n = \bigcap_{i=1}^n (A_i)^{-1} \in \sigma(\pi_1, \ldots, \pi_n) = B_{\mathbb{R}^n}.$$

\[\square\]

**Corollary 9.26.** If $(X, M)$ is a measurable space, then

$$f = (f_1, f_2, \ldots, f_n) : X \to \mathbb{R}^n$$

is $(M, B_{\mathbb{R}^n})$ measurable if $f_i : X \to \mathbb{R}$ is $(M, B_{\mathbb{R}})$ measurable for each $i$. In particular, a function $f : X \to \mathbb{C}$ is $(M, B_{\mathbb{C}})$ measurable if $\text{Re} f$ and $\text{Im} f$ are $(M, B_{\mathbb{R}})$ measurable.

**Proof.** This is an application of Lemma 9.25 and Corollary 9.19 with $Y_i = \mathbb{R}$ for each $i$. 

**Corollary 9.27.** Let $(X, M)$ be a measurable space and $f, g : X \to \mathbb{C}$ be $(M, B_{\mathbb{C}})$ measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(M, B_{\mathbb{C}})$ measurable.

**Proof.** Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_{\pm} : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ 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 $(B_{\mathbb{C}^2}, B_{\mathbb{C}})$ measurable. Also $F = (M, B_{\mathbb{C}^2})$ measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(M, B_{\mathbb{C}})$ 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 9.28.** Let $\alpha \in \mathbb{C}$, $(X, M)$ be a measurable space and $f : X \to \mathbb{C}$ be a $(M, B_{\mathbb{C}})$ measurable function. Then

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if } f(x) \neq 0 \\ \alpha & \text{if } f(x) = 0 \end{cases}$$

is measurable.

**Proof.** Define $i : \mathbb{C} \to \mathbb{C}$ by

$$i(z) = \begin{cases} \frac{1}{z} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0. \end{cases}$$

For any open set $V \subset C$ 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 $B_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in B_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset B_{\mathbb{C}}$ and hence $i^{-1}(B_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) \subset B_{\mathbb{C}}$ which shows that $i$ is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, $F$ is also measurable. 

\[\square\]
Remark 9.29. For the real case of Lemma 9.28, define $i$ as above but now take $z$ to real. From the plot of $i$, Figure 9.29, the reader may easily verify that $i^{-1}((−∞, a])$ is an infinite half interval for all $a$ and therefore $i$ is measurable. See Example 9.34 for another proof of this fact.

2. if $A \subset \mathbb{R}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\mathbb{R}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm \infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\mathbb{R}}$ we may conclude that $A \in \mathcal{B}_{\mathbb{R}}$ as well.

This proves Eq. (9.6).

The proofs of the next two corollaries are left to the reader, see Exercises 9.9 and 9.10.

Corollary 9.31. Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{R}$ be a function. Then the following 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. $f^{-1}(\{−\infty, a\}) \in \mathcal{M}$ 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

$$f^0(x) := \begin{cases} f(x) & \text{if } f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm \infty\} \end{cases}$$

is measurable.

Corollary 9.32. Let $(X, \mathcal{M})$ be a measurable space, $f,g : X \to \mathbb{R}$ be functions and define $f \cdot g : X \to \mathbb{R}$ and $(f+g) : X \to \mathbb{R}$ 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.


Proposition 9.33 (Closure under sups, infs and limits). Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \mathbb{R}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ - 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 $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ - 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

$$\{x : g_+(x) \leq a\} \subset \{x : f_j(x) \leq a \ \forall j\}$$

is measurable.
so that \( g_+ \) is measurable. Similarly if \( g_-(x) = \inf_j f_j(x) \) then
\[
\{ x : g_-(x) \geq a \} = \cap_j \{ x : f_j(x) \geq a \} \in \mathcal{M}.
\]

Since
\[
\lim_{j \to \infty} \sup_n f_j = \inf_n \sup_j \{ f_j : j \geq n \}
\]
and
\[
\lim_{j \to \infty} \inf_n f_j = \sup_n \inf_j \{ f_j : j \geq n \}
\]
we are done by what we have already proved.

Example 9.34. As we saw in Remark 9.29, \( i : \mathbb{R} \to \mathbb{R} \) defined by
\[
i(z) = \begin{cases} 
\frac{1}{z} & \text{if } z \neq 0 \\
0 & \text{if } z = 0.
\end{cases}
\] (9.7)
is measurable by a simple direct argument. For an alternative argument, let
\[
i_n(z) := \frac{z}{z^2 + \frac{1}{n}} \quad \text{for all } n \in \mathbb{N}.
\]
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. (9.7) to be a function from \( \mathbb{C} \) to \( \mathbb{C} \). Again \( i(\cdot) \) is Borel measurable (see also Lemma 9.28) since \( i(z) = \lim_{n \to \infty} i_n(z) \) for all \( z \in \mathbb{C} \) where,
\[
i_n(z) := \frac{z}{|z|^2 + \frac{1}{n}} \quad \text{for all } n \in \mathbb{N}.
\]

Example 9.35. 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 \( f : \mathbb{R} \to \mathbb{R} \) 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 \( g_n(x) = \sqrt{|x - r_n|} \) 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 \( f : \mathbb{R} \to \mathbb{R} \) down then it is going to be measurable.

**Definition 9.36.** Given a function \( f : X \to \mathbb{R} \) let \( f_+(x) := \max \{ f(x), 0 \} \) and \( f_-(x) := \max \{-f(x), 0\} = -\min(\{ f(x)\} \). Notice that \( f = f_+ - f_- \).

**Corollary 9.37.** Suppose \( (X, \mathcal{M}) \) is a measurable space and \( f : X \to \overline{\mathbb{R}} \) is a function. Then \( f \) is measurable iff \( f_\pm \) are measurable.

**Proof.** If \( f \) is measurable, then Proposition 9.33 implies \( f_\pm \) are measurable. Conversely if \( f_\pm \) are measurable then so is \( f = f_+ - f_- \).

**Definition 9.38.** Let \( (X, \mathcal{M}) \) be a measurable space. A function \( \varphi : X \to \mathbb{F} \) (\( \mathbb{F} \) denotes either \( \mathbb{R} \), \( \mathbb{C} \) or \( [0,\infty] \subset \mathbb{R} \)) 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
\[
\varphi = \sum_{i=1}^{n} \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{9.8}
\]
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
\[
\varphi = \sum_{y \in \mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{9.9}
\]

The next theorem shows that simple functions are “pointwise dense” in the space of measurable functions.

**Theorem 9.39 (Approximation Theorem).** Let \( f : [0,\infty] \) be measurable and define, see Figure 9.1
\[
\varphi_n(x) := \sum_{k=0}^{2^n-1} \frac{k}{2^n} 1_{f^{-1}((\frac{k}{2^n}, \frac{k+1}{2^n}])}(x) + 2^n 1_{f^{-1}([2^n,\infty])}(x)
\]
\[
= \sum_{k=0}^{2^n-1} \frac{k}{2^n} 1_{\{ f < \frac{k+1}{2^n} \}}(x) + 2^n 1_{\{ f > 2^n \}}(x)
\]
then \( \varphi_n \leq f \) for all \( n \), \( \varphi_n(x) \uparrow f(x) \) for all \( x \in X \) and \( \varphi_n \uparrow f \) uniformly on the sets \( X_M := \{ x \in X : f(x) \leq M \} \) 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 \( n \to \infty \).

Fig. 9.1. Constructing the simple function, \( \varphi_2 \), approximating a function, \( f : X \to [0, \infty] \). The graph of \( \varphi_2 \) is in red.

**Proof.** Since \( f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) \) and \( f^{-1}\left(\left[2^n, \infty]\right) \) are in \( \mathcal{M} \) as \( f \) is measurable, \( \varphi_n \) is a measurable simple function for each \( n \). Because

\[
\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right],
\]

if \( x \in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right]\right) \) then \( \varphi_n(x) = \varphi_{n+1}(x) = \frac{2k+1}{2^n} \) and if \( x \in f^{-1}\left(\left[2^n, \infty]\right) \) then \( \varphi_n(x) = \frac{2k+1}{2^n} \). Similarly

\[
(2^n, \infty) = \left(\frac{2^n}{2^n}, 2^{n+1}\right] \cup (2^{n+1}, \infty],
\]

and so for \( x \in f^{-1}\left(\left(2^{n+1}, \infty]\right) \), \( \varphi_n(x) = 2^n < 2^{n+1} = \varphi_{n+1}(x) \) and for \( x \in f^{-1}\left(\left(2^{n}, 2^{n+1}\right]\right) \), \( \varphi_{n+1}(x) \geq 2^n = \varphi_n(x) \). Therefore \( \varphi_n \leq \varphi_{n+1} \) for all \( n \). It is clear by construction that \( 0 \leq \varphi_n(x) \leq f(x) \) for all \( x \) and that \( 0 \leq f(x) - \varphi_n(x) \leq 2^{-n} \) if \( x \in X_{2^n} = \{ f \leq 2^n \} \). 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 \( \varphi_n^\pm \leq |f_+ - f_-| = 0 \))

\[
|\varphi_n| = \varphi_n^+ + \varphi_n^- \leq \varphi_{n+1}^+ + \varphi_{n+1}^- = |\varphi_{n+1}|
\]

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 \( |u_n| \uparrow |\text{Re} f| \), \( |v_n| \uparrow |\text{Im} f| \), \( u_n \to \text{Re} f \) and \( v_n \to \text{Im} f \) as \( n \to \infty \). Let \( \varphi_n = u_n + iv_n \), then

\[
|\varphi_n|^2 = u_n^2 + v_n^2 \uparrow |\text{Re} f|^2 + |\text{Im} f|^2 = |f|^2
\]

and \( \varphi_n = u_n + iv_n \to \text{Re} f + i \text{Im} f = f \) as \( n \to \infty \).

### 9.2 Factoring Random Variables

**Lemma 9.40.** Suppose that \( (\mathcal{Y}, \mathcal{F}) \) is a measurable space and \( Y : \Omega \to \mathcal{Y} \) is a map. Then to every \((\sigma(Y), \mathcal{B}_{\mathbb{R}})\) measurable function, \( h : \Omega \to \mathbb{R} \), there is a \((\mathcal{F}, \mathcal{B}_{\mathbb{Y}})\) measurable function \( H : \mathcal{Y} \to \mathbb{R} \) such that \( h = H \circ Y \). More generally, \( \mathbb{R} \) may be replaced by any “standard Borel space,” i.e., a space, \((S, \mathcal{B}_S)\) which is measure theoretic isomorphic to a Borel subset of \( \mathbb{R} \).

\[
(\Omega, \sigma(Y)) \xymatrix{ \ar[r]^Y & (\mathcal{Y}, \mathcal{F}) } \\
(S, \mathcal{B}_S) \ar@{-->}[u]^H
\]

**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 \( \mathbb{R} \).

For a general \((\mathcal{F}, \mathcal{B}_{\mathbb{Y}})\) measurable function, \( h \), from \( \Omega \to \mathbb{R} \), choose simple functions \( h_n \) converging to \( h \). Let \( H_n : \mathcal{Y} \to \mathbb{R} \) be simple functions such that \( h_n = H_n \circ Y \). Then it follows that

\[
h = \lim_{n \to \infty} h_n = \limsup_{n \to \infty} h_n = \liminf_{n \to \infty} H_n \circ Y = H \circ Y
\]

1 Standard Borel spaces include almost any measurable space that we will consider in these notes. For example they include all complete separable metric spaces equipped with the Borel \( \sigma \) - algebra, see Section 12.11.
where $H := \limsup H_n$ – a measurable function from $\mathbb{Y}$ to $\mathbb{R}$.

For the last assertion we may assume that $S \in \mathcal{B}_R$ and $\mathcal{B}_S = (\mathcal{B}_R)_S = \{A \cap S : A \in \mathcal{B}_R\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \mathbb{R}$ which is $(\mathcal{F}, \mathcal{B}_S)$–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

$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on } H^{-1}(S) \\ \ast & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

where $\ast$ is some fixed arbitrary point in $S$. It follows from Proposition 9.13 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 9.41.** Suppose that $(\Omega, \mathcal{B})$ is a measurable space, $X_n : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_n$–measurable functions, and $\mathcal{B}_n := \sigma(X_1, \ldots, X_n) \subset \mathcal{B}$ 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 $\mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R}$–measurable such that $h = H(X_1, \ldots, X_n)$.

$$(\Omega, \mathcal{B}_n = \sigma(Y)) \xrightarrow{Y := (X_1, \ldots, X_n)} (\mathbb{R}^n, \mathcal{B}_\mathbb{R}^n)$$

$h$ $\mathcal{B}_\mathbb{R}$

**Proof.** By Lemma 9.20 and Corollary 9.19 the map, $Y := (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is $(\mathcal{B}, \mathcal{B}_\mathbb{R}^n = \mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R})$–measurable and by Proposition 9.20 $\mathcal{B}_n = \sigma(X_1, \ldots, X_n) = \sigma(Y)$. Thus we may apply Lemma 9.40 to see that there exists a $\mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R}$–measurable map, $H : \mathbb{R}^n \to \mathbb{R}$, such that $h = H \circ Y = H(X_1, \ldots, X_n)$.

### 9.3 Summary of Measurability Statements

It may be worthwhile to gather the statements of the main measurability results of Sections 9.1 and 9.2 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 9.8).
2. If $I \subset M$ generates $M$ (i.e. $M = \sigma(I)$), then a map, $g : \Omega \to X$ is $(\mathcal{M}, \mathcal{M})$–measurable iff $g^{-1}(I) \subset \mathcal{B}$ (Lemma 9.3 and Proposition 9.10).
3. The notion of measurability may be localized (Proposition 9.13).
4. Composition of measurable functions are measurable (Lemma 9.14).
5. Continuous functions between two topological spaces are also Borel measurable (Proposition 9.23).
6. $\sigma(f) = \sigma(f_\alpha : \alpha \in I)$ (Proposition 9.20).
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 9.17).
8. A map, $h : X \to Y$ is $(\mathcal{M}, \mathcal{F})$–measurable if $\pi_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$–measurable for all $\alpha \in I$ (Corollary 9.19).
9. If $I = \{1, 2, \ldots, n\}$, 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 9.21.
10. $\mathcal{B}_n = \mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R}$ ($n$–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 9.25).
11. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\mathbb{R}^n, \mathcal{B}_\mathbb{R}^n)$ is closed under the usual pointwise algebraic operations (Corollary 9.32). They are also closed under the countable supremums, infimums, and limits (Proposition 9.33).
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 9.27 and Proposition 9.33). 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 $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ may be well approximated by measurable simple functions (Theorem 9.39).
14. If \( X_i : \Omega \to \mathbb{R} \) are \( \mathcal{B}/\mathcal{B}_\mathbb{R} \) – measurable maps and \( \mathcal{B}_n := \sigma(X_1, \ldots, X_n) \), then \( h : \Omega \to \mathbb{R} \) is \( \mathcal{B}_n \) – measurable iff \( h = H(X_1, \ldots, X_n) \) for some \( \mathcal{B}_\mathbb{R}^n/\mathcal{B}_\mathbb{R} \) – measurable map, \( H : \mathbb{R}^n \to \mathbb{R} \) (Corollary 9.41).

15. We also have the more general factorization Lemma 9.40.

For the most part most of our future measurability issues can be resolved by one or more of the items on this list.

### 9.4 Distributions / Laws of Random Vectors

The proof of the following proposition is routine and will be left to the reader.

**Proposition 9.42.** 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 \(\text{Law}_\mu(f)\) and call \(f_* \mu\) the push-forward of \(\mu\) by \(f\) or the law of \(f\) under \(\mu\).

**Definition 9.43.** Suppose that \(\{X_i\}_{i=1}^n\) is a sequence of random variables on a probability space, \((\Omega, \mathcal{B}, P)\). The probability measure, \(\mu = (X_1, \ldots, X_n)_* P = P \circ (X_1, \ldots, X_n)^{-1}\) on \(\mathcal{B}_\mathbb{R}\) (see Proposition 9.42) is called the **joint distribution (or law)** of \((X_1, \ldots, X_n)\). To be more explicit, 

\[
\mu(B) := P((X_1, \ldots, X_n) \in B) := P(\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega)) \in B\})
\]

for all \(B \in \mathcal{B}_\mathbb{R}^n\).

**Corollary 9.44.** The joint distribution, \(\mu\) is uniquely determined from the knowledge of

\[
P((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n)
\]

for all \(A_i \in \mathcal{B}_\mathbb{R}\)

or from the knowledge of

\[
P(X_1 \leq x_1, \ldots, X_n \leq x_n)
\]

for all \(x = (x_1, \ldots, x_n) \in \mathbb{R}^n\).

**Proof.** Apply Proposition 6.55 with \(\mathcal{P}\) being the \(\pi\) – systems defined by

\[
\mathcal{P} := \{A_1 \times \cdots \times A_n \in \mathcal{B}_\mathbb{R}^n : A_i \in \mathcal{B}_\mathbb{R}\}\]

for the first case and

\[
\mathcal{P} := \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \in \mathcal{B}_\mathbb{R}^n : x_i \in \mathbb{R}\}
\]

for the second case.

**Definition 9.45.** 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 \((X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n)\) if \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_n)\) have the **same distribution / law**, i.e. if
Proposition 9.46. Let us continue using the notation in Definition 9.45. Further let
\[ X = (X_1, X_2, \ldots, X_n) : \Omega \rightarrow \mathbb{R}^n \quad \text{and} \quad Y = (Y_1, Y_2, \ldots, Y_n) : \Omega' \rightarrow \mathbb{R}^n \]
and let \( \mathcal{F} := \otimes_{n \in \mathbb{N}} \mathcal{B}_\mathbb{R} \) be the product \( \sigma \)-algebra on \( \mathbb{R}^n \). Then \( \{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty \iff X_iP = Y_iP' \) as measures on \( (\mathbb{R}^n, \mathcal{F}) \).

Proof. Let
\[ \mathcal{P} := \cup_{n=1}^\infty \{ A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^N : A_i \in \mathcal{B}_\mathbb{R} \text{ for } 1 \leq i \leq n \} . \]
Notice that \( \mathcal{P} \) is a \( \sigma \)-system and it is easy to show \( \sigma(\mathcal{P}) = \mathcal{F} \) (see Exercise 9.7). Therefore by Proposition 6.55, \( X_iP = Y_iP' \) iff \( X_iP = Y_iP' \) on \( \mathcal{P} \). Now for \( A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^N \in \mathcal{P} \) we have,
\[ X_iP( A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^N ) = P( (X_1, \ldots, X_n) \in A_1 \times A_2 \times \cdots \times A_n ) \]
and hence the condition becomes,
\[ P( (X_1, \ldots, X_n) \in A_1 \times A_2 \times \cdots \times A_n ) = P'( (Y_1, \ldots, Y_n) \in A_1 \times A_2 \times \cdots \times A_n ) \]
for all \( n \in \mathbb{N} \) and \( A_i \in \mathcal{B}_\mathbb{R} \). Another application of Proposition 6.55 or using Corollary 9.44 allows us to conclude that shows that \( X_iP = Y_iP' \) iff \( (X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n) \) for all \( n \in \mathbb{N} \).

Corollary 9.47. Continue the notation above and assume that \( \{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty \). Further let
\[ X_\pm = \begin{cases} \limsup_{n \rightarrow \infty} X_n & \text{if } + \\ \liminf_{n \rightarrow \infty} X_n & \text{if } - \end{cases} \]
and define \( Y_\pm \) similarly. Then \( (X_-, X_+) \overset{d}{=} (Y_-, Y_+) \) as random variables into \( (\mathbb{R}^2, \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathbb{R}) \). In particular,
\[ P \left( \lim_{n \rightarrow \infty} X_n \text{ exists in } \mathbb{R} \right) = P' \left( \lim_{n \rightarrow \infty} Y_n \text{ exists in } \mathbb{R} \right) . \] (9.10)

Proof. First suppose that \( (\Omega', \mathcal{B}', P') = (\mathbb{R}^N, \mathcal{F}, P' := X_iP) \) where \( Y_i(a_1, a_2, \ldots) := a_i \in \pi_i(a_1, a_2, \ldots) \). Then for \( C \in \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathbb{R} \) we have,
\[ X^{-1}( \{ (Y_-, Y_+) \in C \} ) = \{ (Y_- \circ X, Y_+ \circ X) \in C \} = \{ (X_-, X_+) \in C \} \]
since, for example,
\[ Y_- \circ X = \liminf_{n \rightarrow \infty} Y_n \circ X = \liminf_{n \rightarrow \infty} X_n = X_- . \]
Therefore it follows that
\[ P( (X_-, X_+) \in C ) = P \circ X^{-1}( \{ (Y_-, Y_+) \in C \} ) = P'( \{ (Y_-, Y_+) \in C \} ) . \] (9.11)

The general result now follows by two applications of this special case. For the last assertion, take
\[ C = \{ (x, x) : x \in \mathbb{R} \} \in \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathbb{R} \subset \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathbb{R} . \]
Then \( (X_-, X_+) \in C \) iff \( x_- = x_+ \in \mathbb{R} \) which happens iff \( \lim_{n \rightarrow \infty} X_n \) exists in \( \mathbb{R} \). Similarly, \( (Y_-, Y_+) \in C \) iff \( \lim_{n \rightarrow \infty} Y_n \) exists in \( \mathbb{R} \) and therefore Eq. (9.10) holds as a consequence of Eq. (9.11).

Exercise 9.11. 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, \( S_n := X_1 + \cdots + X_n \) and \( T_n := Y_1 + \cdots + Y_n \). Prove the following assertions.
1. Suppose that \( f : \mathbb{R}^n \rightarrow \mathbb{R}^k \) is a \( \mathcal{B}_\mathbb{R}^n / \mathcal{B}_\mathbb{R}^k \) measurable function, then \( f( (X_1, \ldots, X_n) ) \overset{d}{=} f( (Y_1, \ldots, Y_n) ) \).
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 \( f : \mathbb{R}^n \rightarrow \mathbb{R}^n \).

9.5 Generating All Distributions from the Uniform Distribution

If \( U \in (0, 1) \) is a random variable with the uniform distribution, \( G : (0, 1) \rightarrow \mathbb{R} \) is a non-decreasing function, and \( F(x) := P(G(U) \leq x) \) is the cumulative distribution function of \( G(U) \), then
\[ F(x) = m \{ y \in (0, 1) : G(y) \leq x \} = \sup \{ y \in (0, 1) : G(y) \leq x \} . \] (9.12)
Now suppose that \( F : \mathbb{R} \rightarrow [0, 1] \) 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 \( \lim_{x \rightarrow \infty} F(x) = 1 \), and \( \lim_{x \rightarrow -\infty} F(x) = 0 \). We would like to find a function \( G \) as above such that Eq. (9.12) holds. If \( F \) happened to be continuous and
strictly increasing we should take \( G(y) = F^{-1}(y) \). For general distributions functions \( (F) \) we will show that
\[
G(y) := \inf \{ x \in \mathbb{R} : y \leq F(x) \} \quad \text{for all } y \in (0, 1),
\]
(see see Figure 9.2) is valid.

\[\text{Theorem 9.48. If } F : \mathbb{R} \to [0, 1] \text{ is a cumulative distribution function of a probability measure } \mu = \mu_F \text{ on } (\mathbb{R}, \mathcal{B}_\mathbb{R}) \text{ and } G \text{ is defined as in Eq. (9.13), then } G : (0, 1) \to \mathbb{R} \text{ is a non-decreasing (hence Borel measurable) function such that } \text{Law}_F (G(U)) = G_* \mu = \mu_F.\]

\[\text{Proof. Since } y < 1, G(y) < \infty \text{ and since } y > 0, G(y) = -\infty \text{ wherein we have used } \lim_{x \to -\infty} F(x) = 1, \text{ and } \lim_{x \to -\infty} F(x) = 0 \text{ respectively. Since } F \text{ is non-decreasing it is easily seen that } G \text{ is non-decreasing. To finish the proof it suffices to prove } G(y) \leq x \text{ if } y \leq F(x). \text{ For once this is done it easily follows that Eq. (9.12) is valid.} \]

By definition of \( G(y) \), \( G(y) \leq x \) iff there exists a non-increasing sequence \( \{x_n\} \subset \mathbb{R} \) such that \( y \leq F(x_n) \) for all \( n \) and \( \lim_{n \to \infty} x_n = x \leq x \). Since \( \lim_{x \to -\infty} F(x) = 0 \) and \( y > 0 \), it follows that \( \lim_{n \to \infty} x_n = x \in \mathbb{R} \) for some \( x \). Because \( F \) is right continuous we may conclude that \( F(x_0) \geq y \). Thus we have shown \( G(y) \leq x \) iff there exists \( x \leq x \) such that \( y \leq F(x) \). As \( F \) is non-decreasing this last equivalence is equivalent to \( G(y) \leq x \) iff \( y \leq F(x) \). \( \square \)

\[\text{Theorem 9.49 (Durrett’s Version). Given a distribution function, } F : \mathbb{R} \to [0, 1] \text{ let } Y : (0, 1) \to \mathbb{R} \text{ be defined (see Figure 9.3) by,} \]
\[Y(x) := \sup \{ y : F(y) < x \}. \]

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 \(-\infty < a < b < \infty \).

\[\text{Proof. Since } Y : (0, 1) \to \mathbb{R} \text{ is a non-decreasing function, } Y \text{ is measurable. Also observe, if } y < Y(x), \text{ then } F(y) < x \text{ and hence,} \]
\[F(Y(x)^-) = \lim_{y \uparrow Y(x)} F(y) \leq x. \]

For \( y > Y(x) \), we have \( F(y) \geq x \) and therefore,
\[F(Y(x)^+) = \lim_{y \downarrow Y(x)} F(y) \geq x \]
and so we have shown
\[F(Y(x)^-) \leq x \leq F(Y(x)^+). \]

We will now show
\[\{ x \in (0, 1) : Y(x) \leq y_0 \} = (0, F(y_0)] \cap (0, 1). \quad (9.14)\]

For the inclusion “\( \subset \)” if \( x \in (0, 1) \) and \( Y(x) \leq y_0 \), then \( x \leq F(Y(x)) \leq F(y_0) \), i.e. \( x \in (0, F(y_0)] \cap (0, 1) \). Conversely if \( x \in (0, 1) \) and \( x \leq F(y_0) \) then (by definition of \( Y(x) \)) \( y_0 \geq Y(x) \).

From the identity in Eq. (9.14), it follows that \( Y \) is measurable and
\[\{ Y_* m ((-\infty, y_0)) = m (Y^{-1}((-\infty, y_0))) = m ((0, F(y_0)] \cap (0, 1)) = F(y_0). \]

Therefore, \( \text{Law}(Y) = \mu_F \) as desired. \( \square \)
Integration Theory

In this chapter, we will greatly extend the “simple” integral or expectation which was developed in Section 5.5 above. Recall there that if \((\Omega,B,\mu)\) was measurable space and \(\varphi: \Omega \to [0,\infty)\) was a measurable simple function, then we let

\[
\mathbb{E}_\mu \varphi := \sum_{\lambda \in [0,\infty)} \lambda \mu(\varphi = \lambda).
\]

The conventions being used here is that \(0 \cdot \mu(\varphi = 0) = 0\) even when \(\mu(\varphi = 0) = \infty\). This convention is necessary in order to make the integral linear – at a minimum we will want \(\mathbb{E}_\mu \{0\} = 0\). Please be careful not blindly apply the \(0 \cdot \infty = 0\) convention in other circumstances.

10.1 Integrals of positive functions

**Definition 10.1.** Let \(L^+ = L^+(B) = \{f: \Omega \to [0,\infty]: f \text{ is measurable}\}\). Define

\[
\int_{\Omega} f(\omega) d\mu(\omega) = \int_{\Omega} f d\mu := \sup \{\mathbb{E}_\mu \varphi: \varphi \text{ is simple and } \varphi \leq f\}.
\]

We say the \(f \in L^+\) is integrable if \(\int_{\Omega} f d\mu < \infty\). If \(A \in B\), let

\[
\int_{A} f(\omega) d\mu(\omega) = \int_{A} f d\mu := \int_{A} 1_A f d\mu.
\]

We also use the notation,

\[
\mathbb{E} f = \int_{\Omega} f d\mu \text{ and } \mathbb{E} \{f: A\} := \int_{A} f d\mu.
\]

**Remark 10.2.** 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 10.3.** Let \(f, g \in L^+(B)\). Then:

1. if \(\lambda \geq 0\), 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 \(0 \leq f \leq g\), then

\[
\int_{\Omega} f d\mu \leq \int_{\Omega} g d\mu. \tag{10.1}
\]

3. For all \(\varepsilon > 0\) and \(p > 0\),

\[
\mu(f \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int_{\Omega} f^p 1_{\{f \geq \varepsilon\}} d\mu \leq \frac{1}{\varepsilon^p} \int_{\Omega} f^p d\mu. \tag{10.2}
\]

The inequality in Eq. (10.2) 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,

\[
\int_{\Omega} \lambda f d\mu = \sup \{\mathbb{E}_\mu \varphi: \varphi \text{ is simple and } \varphi \leq \lambda f\}
\]

\[
= \sup \{\mathbb{E}_\mu \varphi: \varphi \text{ is simple and } \lambda^{-1} \varphi \leq f\}
\]

\[
= \sup \{\mathbb{E}_\mu [\lambda \varphi]: \varphi \text{ is simple and } \varphi \leq f\}
\]

\[
= \lambda \int_{\Omega} f d\mu.
\]

2. Since

\[
\{\varphi \text{ is simple and } \varphi \leq f\} \subset \{\varphi \text{ is simple and } \varphi \leq g\},
\]

Eq. (10.1) follows from the definition of the integral.

3. Since \(1_{\{f \geq \varepsilon\}} \leq \frac{1}{\varepsilon} f \leq \frac{1}{\varepsilon} f\) we have

\[
1_{\{f \geq \varepsilon\}} \leq \frac{1}{\varepsilon^p} \int_{\Omega} f^p 1_{\{f \geq \varepsilon\}} d\mu \leq \frac{1}{\varepsilon^p} \int_{\Omega} f^p d\mu
\]

and by monotonicity and the multiplicative property of the integral,

\[
\mu(f \geq \varepsilon) = \int_{\Omega} 1_{\{f \geq \varepsilon\}} d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_{\Omega} 1_{\{f \geq \varepsilon\}} f^p d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_{\Omega} f^p d\mu.
\]
4. If \( \mu(f = \infty) > 0 \), then \( \varphi_n := n1_{(f = \infty)} \) is a simple function such that
\[ \varphi_n \leq f \]
for all \( n \) and hence
\[ n\mu(f = \infty) = E\mu(\varphi_n) \leq \int f \, d\mu \]
for all \( n \). Letting \( n \to \infty \) shows \( \int f \, d\mu = \infty \). Thus if \( \int f \, d\mu < \infty \) then
\[ \mu(f = \infty) = 0. \]
Moreover,
\[ \{ f > 0 \} = \bigcup_{n=1}^{\infty} \{ f > 1/n \} \]
with \( \mu(f > 1/n) \leq n \int f \, d\mu \leq \infty \) for each \( n \).

**Theorem 10.4 (Monotone Convergence Theorem).** Suppose \( f_n \in L^+ \) is a sequence of functions such that \( f_n \uparrow f \) \((f \text{ is necessarily in } L^+)\) then
\[ \int f_n \uparrow \int f \text{ as } n \to \infty. \]

**Proof.** Since \( f_n \leq f_m \leq f \), for all \( n \leq m < \infty \),
\[ \int f_n \leq \int f_m \leq \int f \]
from which if follows \( \int f_n \) is increasing in \( n \) and
\[ \lim_{n \to \infty} \int f_n \leq \int f. \] (10.3)

For the opposite inequality, let \( \varphi : \Omega \to [0, \infty) \) be a simple function such that \( 0 \leq \varphi \leq f \), \( \alpha \in (0,1) \) and \( \Omega_n := \{ f_n \geq \alpha \varphi \} \)
\[ \text{Notice that } \Omega_n \uparrow \Omega \text{ and } f_n \geq \alpha \Omega_n \cdot \varphi \text{ and so by definition of } \int f_n, \]
\[ \int f_n \geq E\mu(\alpha 1_{\Omega_n} \varphi) = \alpha E\mu(1_{\Omega_n} \varphi). \] (10.4)

Then using the identity
\[ 1_{\Omega_n} \varphi = 1_{\Omega_n} \sum_{y>0} y1_{\{ \varphi=y \}} = \sum_{y>0} y1_{\{ \varphi=y \}\cap \Omega_n}, \]
and the linearity of \( E\mu \) we have,
\[ \lim_{n \to \infty} E\mu(1_{\Omega_n} \varphi) = \lim_{n \to \infty} \sum_{y>0} y \cdot \mu(\Omega_n \cap \{ \varphi = y \}) \]
\[ = \sum_{y>0} \lim_{n \to \infty} \mu(\Omega_n \cap \{ \varphi = y \}) \text{ (finite sum)} \]
\[ = \sum_{y>0} \mu(\{ \varphi = y \}) = E\mu[\varphi], \]

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. (10.4) to conclude
\[ \lim_{n \to \infty} \int f_n \geq \alpha E\mu[\varphi] \]
and since \( \alpha \in (0,1) \) was arbitrary we may further conclude, \( E\mu[\varphi] \leq \lim_{n \to \infty} \int f_n \). The latter inequality being true for all simple functions \( \varphi \) with \( \varphi \leq f \) then implies that
\[ \int f = \sup_{0 \leq \varphi \leq f} E\mu[\varphi] \leq \lim_{n \to \infty} \int f_n, \]
which combined with Eq. (10.3) proves the theorem.

**Remark 10.5.** 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 \( \Omega_n := \{ f_n \geq \alpha \varphi \} \) to be measurable.

**Remark 10.6 ("Explicit" Integral Formula).** Given \( f : \Omega \to [0, \infty] \) measurable, we know from the approximation Theorem 9.39 \( \varphi_n \uparrow f \) where
\[ \varphi_n := \sum_{k=0}^{2n-1} \frac{k}{2^n} 1_{\{ \frac{k}{2^n} < f \leq \frac{k+1}{2^n} \}} + 2^n 1_{\{ f > 2^n \}}. \]

Therefore by the monotone convergence theorem,
\[ \int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu \]
\[ = \lim_{n \to \infty} \sum_{k=0}^{2n-1} \frac{k}{2^n} \mu(\{ \frac{k}{2^n} < f \leq \frac{k+1}{2^n} \}) + 2^n \mu(\{ f > 2^n \}). \]

**Corollary 10.7.** 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.
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,
\[
\int (f_1 + f_2) = \lim_{n \to \infty} \int (\varphi_n + \psi_n) = \lim_{n \to \infty} \left( \int \varphi_n + \int \psi_n \right) = \lim_{n \to \infty} \int \varphi_n + \lim_{n \to \infty} \int \psi_n = \int f_1 + \int f_2.
\]

Now to the general case. Let \( g_N := \sum_{n=1}^{\infty} f_n \) and \( g = \sum_{n=1}^{\infty} f_n \), then \( g_N \uparrow g \) and so again by monotone convergence theorem and the additivity just proved,
\[
\sum_{n=1}^{\infty} \int f_n := \lim_{N \to \infty} \sum_{n=1}^{N} \int f_n = \lim_{N \to \infty} \int \sum_{n=1}^{N} f_n = \lim_{N \to \infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.
\]

Example 10.8 (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 [10.7] that
\[
\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} \int_{\{n\}} f (n) 1_{\{n\}} d\mu = \sum_{n=1}^{\infty} f (n) \mu (\{n\}) = \sum_{n=1}^{\infty} f (n).
\]
Thus the integral relative to counting measure is simply the infinite sum.

Lemma 10.9 (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
\[
\sum_{\omega \in \Omega} \varphi (\omega) \sum_{z \in \Omega (\omega)} z 1_{\{\varphi (z) \}} (\omega) = \sum_{z \in [0, \infty)} z \sum_{\omega \in \Omega} \rho (\omega) 1_{\{\varphi (z) \}} (\omega)
\]
\[
= \sum_{z \in [0, \infty)} z \mu (\{\varphi (z) \}) = \int_\Omega \varphi d\mu.
\]
So if \( \varphi : \Omega \to [0, \infty] \) is a simple function such that \( \varphi \leq f \), then
\[
\int_\Omega \varphi d\mu = \sum_{\omega \in \Omega} \varphi (\omega) \leq \sum_{\omega \in \Omega} f (\omega).
\]
Taking the sup over \( \varphi \) in this last equation then shows that
\[
\int_{\Omega} f d\mu \leq \sum_{\omega \in \Omega} f (\omega).
\]
For the reverse inequality, let \( A \subset \Omega \) be a finite set and \( N \in (0, \infty) \). Set \( f^N (\omega) = \min \{ N, f (\omega) \} \) and let \( \varphi_{N,A} \) be the simple function given by \( \varphi_{N,A} (\omega) := 1_{A (\omega)} f^N (\omega) \). Because \( \varphi_{N,A} (\omega) \leq f (\omega) \),
\[
\sum_{\omega \in \Omega} f^N (\omega) = \sum_{\omega \in A} \varphi_{N,A} (\omega) = \int_A f^N d\mu \leq \int_A f d\mu.
\]
Since \( f^N \uparrow f \) as \( N \to \infty \), we may let \( N \to \infty \) in this last equation to conclude
\[
\sum_{\omega \in A} f (\omega) \leq \int_A f d\mu.
\]
Since \( A \) is arbitrary, this implies
\[
\sum_{\omega \in \Omega} f (\omega) \leq \int_{\Omega} f d\mu.
\]

Exercise 10.1. 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 10.10. Suppose that \( f \geq 0 \) is a measurable function. Then \( \int f d\mu = 0 \) if \( f = 0 \) a.e. Also if \( f, g \geq 0 \) are measurable functions such that \( f \leq g \) a.e. then \( \int f d\mu \leq \int g d\mu \). In particular if \( f = g \) a.e. then \( \int f d\mu = \int g d\mu \).
**Proof.** If \( f = 0 \) a.e. and \( \varphi \leq f \) 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 10.3),
\[
\mu(f \geq 1/n) \leq n \int f \, d\mu = 0 \text{ for all } n.
\]
Therefore, \( \mu(f > 0) \leq \sum_{n=1}^\infty \mu(f \geq 1/n) = 0 \), 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 \( 1_{E^c} f \leq 1_{E^c} g \) 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 \( 1_{E^c} f \leq 1_{E^c} g \) everywhere,
\[
\int f \, d\mu = \int 1_{E^c} f \, d\mu \leq \int 1_{E^c} g \, d\mu = \int g \, d\mu.
\]

**Corollary 10.11.** 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 10.10,
\[
\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.
\]

**Lemma 10.12 (Fatou’s Lemma).** If \( f_n : \Omega \to [0, \infty) \) is a sequence of measurable functions then
\[
\int \liminf_{n \to \infty} f_n \leq \liminf_{n \to \infty} \int f_n
\]

**Proof.** Define \( g_k := \inf_{n \geq k} f_n \) so that \( g_k \uparrow \liminf_{n \to \infty} f_n \) as \( k \to \infty \). Since \( g_k \leq f_n \) for all \( k \leq n \),
\[
\int g_k \leq \int f_n \text{ for all } n \geq k
\]
and therefore
\[
\int g_k \leq \liminf_{n \to \infty} \int f_n \text{ for all } k.
\]
We may now use the monotone convergence theorem to let \( k \to \infty \) to find
\[
\int \liminf_{n \to \infty} f_n = \int \liminf_{k \to \infty} g_k = \lim_{k \to \infty} \int g_k \leq \liminf_{n \to \infty} \int f_n.
\]

The following Corollary and the next lemma are simple applications of Corollary 10.7.

**Corollary 10.13.** 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 \( i \neq j \), then
\[
\mu(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mu(A_n).
\]

**Proof.** Since
\[
\mu(\bigcup_{n=1}^\infty A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} \, d\mu \text{ and }
\sum_{n=1}^\infty \mu(A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} \, d\mu
\]
it suffices to show
\[
\sum_{n=1}^\infty 1_{A_n} = 1_{\bigcup_{n=1}^\infty A_n} \text{ } \mu \text{- a.e. (10.5)}
\]
Now \( \sum_{n=1}^\infty 1_{A_n} \geq 1_{\bigcup_{n=1}^\infty A_n} \) and \( \sum_{n=1}^\infty 1_{A_n}(\omega) \neq 1_{\bigcup_{n=1}^\infty A_n}(\omega) \) iff \( \omega \in A_i \cap A_j \) for some \( i \neq j \), that is
\[
\left\{ \omega : \sum_{n=1}^\infty 1_{A_n}(\omega) \neq 1_{\bigcup_{n=1}^\infty A_n}(\omega) \right\} = \bigcup_{i<j} A_i \cap A_j
\]
and the latter set has measure zero. This proves Eq. (10.5) and hence the corollary.

**Lemma 10.14 (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 \}'s \}
\[
= \bigcap_{N=1}^\infty \bigcup_{N \geq n} A_n.
\]

If \( \sum_{n=1}^\infty \mu(A_n) < \infty \) then \( \mu(\{ A_n \text{ i.o.} \}) = 0 \).
To shorten notation in this chapter we may simply write this fact:

\[
\sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty.
\]

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:

\[
\mu(A_n \text{ i.o.}) = \lim_{N \to \infty} \mu\left( \bigcup_{n \geq N} A_n \right)
\]

\[
\leq \lim_{N \to \infty} \sum_{n \geq N} \mu(A_n)
\]

and the last limit is zero since \(\sum_{n=1}^{\infty} \mu(A_n) < \infty\).

**Example 10.15.** 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, \(P(\lim_{n \to \infty} X_n = 0) = 1\).

### 10.2 Integrals of Complex Valued Functions

**Definition 10.16.** A measurable function \(f : \Omega \to \mathbb{R}\) is integrable if \(f_+ := f1_{\{f \geq 0\}}\) and \(f_- := -f1_{\{f \leq 0\}}\) are integrable. We write \(L^1(\mu; \mathbb{R})\) for the space of real valued integrable functions. For \(f \in L^1(\mu; \mathbb{R})\), 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 \(f, g : \Omega \to \mathbb{R}\) are two measurable functions, let \(f + g\) denote the collection of measurable functions \(h : \Omega \to \mathbb{R}\) 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 \(f, g \in L^1(\mu; \mathbb{R})\) and \(h_1, h_2 \in f + g\), then \(h_1 = h_2\) a.e. because \(|f| < \infty\) and \(|g| < \infty\) a.e.

**Notation 10.17 (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 10.18.** Since

\[
f_\pm \leq |f| \leq f_+ + f_-,
\]

a measurable function \(f\) is integrable iff \(\int |f| ~d\mu < \infty\). Hence

\[
L^1(\mu; \mathbb{R}) := \left\{ f : \Omega \to \mathbb{R} : f \text{ is measurable and } \int_{\Omega} |f| ~d\mu < \infty \right\}.
\]

If \(f, g \in L^1(\mu; \mathbb{R})\) and \(f = g\) a.e. then \(f_\pm = g_\pm\) a.e. and so it follows from Proposition 10.10 that \(\int f d\mu = \int g d\mu\). In particular if \(f, g \in L^1(\mu; \mathbb{R})\) we may define

\[
\int_{\Omega} (f + g) d\mu = \int_{\Omega} h d\mu
\]

where \(h\) is any element of \(f + g\).

**Proposition 10.19.** The map

\[
f \in L^1(\mu; \mathbb{R}) \to \Omega \to f d\mu \in \mathbb{R}
\]

is linear and has the monotonicity property: \(\int f d\mu \leq \int g d\mu\) for all \(f, g \in L^1(\mu; \mathbb{R})\) such that \(f \leq g\) a.e.

**Proof.** Let \(f, g \in L^1(\mu; \mathbb{R})\) 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 \(af + bg \in L^1(\mu; \mathbb{R})\) because

\[
|af + bg| \leq |a||f| + |b||g| \in L^1(\mu; \mathbb{R}).
\]

If \(a < 0\), then

\[
(af)_+ = -af_- \quad \text{and} \quad (af)_- = -af_+
\]

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...
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 \( f_+ - f_- = f \leq g = g_+ - g_- \) then \( f_+ + g_- \leq g_+ + f_- \) which implies that
\[ \int f_+ + \int g_- \leq \int g_+ + \int f_- \]
or equivalently that
\[ \int f = \int f_+ - \int f_- \leq \int g_+ - \int g_- = \int g. \]

The monotonicity property is also a consequence of the linearity of the integral, the fact that \( \|f\|_{\infty} \leq \sqrt{2} \max(\|\text{Re} f\|, \|\text{Im} f\|) \), \( \int |f| \ d\mu < \infty \) iff
\[ \int |\text{Re} f| \ d\mu + \int |\text{Im} f| \ d\mu < \infty. \]

For \( f \in L^1(\mu; \mathbb{C}) \) define
\[ \int f \ d\mu = \int \text{Re} f \ d\mu + i \int \text{Im} f \ d\mu. \]

Proposition 10.21. Suppose that \( f \in L^1(\mu; \mathbb{C}) \), then
\[ \left| \int f d\mu \right| \leq \int |f| d\mu. \quad (10.6) \]

**Proof.** Start by writing \( \int f d\mu = \text{Re} e^i\theta \) with \( R \geq 0 \). We may assume that \( R = |\int f d\mu| > 0 \) since otherwise there is nothing to prove. Since
\[ R = e^{-i\theta} \int f d\mu = \int e^{-i\theta} f d\mu = \int \text{Re} e^{-i\theta} f d\mu + i \int \text{Im} e^{-i\theta} f d\mu, \]
it must be that \( \int \text{Im} e^{-i\theta} f d\mu = 0 \). Using the monotonicity in Proposition 10.10,
\[ \left| \int f d\mu \right| = \int \text{Re} e^{-i\theta} f d\mu \leq \int \left| \text{Re} e^{-i\theta} f \right| d\mu \leq \int |f| d\mu. \]

Proposition 10.22. Let \( f, g \in L^1(\mu; \mathbb{C}) \), then
1. The set \( \{ f \neq 0 \} \) is \( \sigma \)-finite, in fact \( \{ |f| \geq \frac{1}{n} \} \) \( \uparrow \{ f \neq 0 \} \) and \( \mu(\{ |f| \geq \frac{1}{n} \}) < \infty \) for all \( n \).  
2. The following are equivalent 
   a) \( \int_E f = \int_E g \) for all \( E \in \mathcal{B} \)  
   b) \( \int |f - g| = 0 \)  
   c) \( f = g \) a.e.

**Proof.** 1. By Chebyshev’s inequality, Lemma 10.3
\[ \mu(\{ |f| \geq \frac{1}{n} \}) \leq n \int_{\Omega} |f| d\mu < \infty \]
for all \( n \).
2. (a) \( \implies \) (c) Notice that
\[ \int_E f = \int_E g \iff \int_E (f - g) = 0 \]
for all \( E \in \mathcal{B} \). Taking \( E = \{ \text{Re}(f - g) > 0 \} \) and using \( 1_E \text{Re}(f - g) \geq 0 \), we learn that
\[ 0 = \text{Re} \int_E (f - g) d\mu = \int 1_E \text{Re}(f - g) \implies 1_E \text{Re}(f - g) = 0 \text{ a.e.} \]
This implies that \( 1_E = 0 \text{ a.e.} \) which happens iff...
\[ \mu(\{\text{Re}(f-g) > 0\}) = \mu(\{\text{Re}(f-g) < 0\}) = \mu(E) = 0. \]

Similar \( \mu(\text{Re}(f-g) < 0) = 0 \) so that \( \text{Re}(f-g) = 0 \) a.e. Similarly, \( \text{Im}(f-g) = 0 \) 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

\[ \left| \int_E f - \int_E g \right| \leq \int |f - g| = 0. \]

\[ \Box \]

Lemma 10.23 (Integral Comparison I). Suppose that \( h \in L^1(\mu) \) satisfies

\[ \int_A h d\mu \geq 0 \text{ for all } A \in \mathcal{B}, \tag{10.7} \]

then \( h \geq 0 \) a.e.

\[ \textbf{Proof.} \] Since by assumption,

\[ 0 = \text{Im} \int_A h d\mu = \int_A \text{Im} h d\mu \text{ for all } A \in \mathcal{B}, \]

we may apply Proposition 10.22 to conclude that \( \text{Im} h = 0 \) a.e. Thus we may now assume that \( h \) is real valued. Taking \( A = \{h < 0\} \) in Eq. (10.7) implies

\[ \int_{\Omega} 1_A |h| d\mu = \int_{\Omega} -1_A h d\mu = -\int_A h d\mu \leq 0. \]

However \( 1_A |h| \geq 0 \) and therefore it follows that \( \int_{\Omega} 1_A |h| d\mu = 0 \) and so Proposition 10.22 implies \( 1_A |h| = 0 \) a.e. which then implies \( 0 = \mu(A) = \mu(h < 0) = 0. \]

\[ \Box \]

Lemma 10.24 (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 \( f \geq g \) a.e. iff

\[ \int_{\Omega} f d\mu \geq \int_{\Omega} g d\mu \text{ for all } A \in \mathcal{B}. \tag{10.8} \]

In particular \( f = g \) a.e. iff equality holds in Eq. (10.8).

\[ \textbf{Proof.} \] It was already shown in Proposition 10.10 that \( f \geq g \) a.e. implies Eq. (10.8). For the converse assertion, let \( B_n := \{f \leq n1_{\Omega_n}\} \). Then from Eq. (10.8),

\[ \infty > n\mu(\Omega_n) \geq \int f1_{B_n} d\mu \geq \int g1_{B_n} d\mu \]

from which it follows that both \( f1_{B_n} \) and \( g1_{B_n} \) are in \( L^1(\mu) \) and hence \( h := f1_{B_n} - g1_{B_n} \in L^1(\mu) \). Using Eq. (10.8) again we know that

\[ \int_A h = \int f1_{B_n \cap A} - \int g1_{B_n \cap A} \geq 0 \text{ for all } A \in \mathcal{B}. \]

An application of Lemma 10.23 implies \( h \geq 0 \) a.e., i.e. \( f1_{B_n} \geq g1_{B_n} \) a.e. Since \( B_n \uparrow \{f < \infty\} \), we may conclude that

\[ f1_{\{f < \infty\}} = \lim_{n \to \infty} f1_{B_n} \geq \lim_{n \to \infty} g1_{B_n} = g1_{\{f < \infty\}} \text{ a.e.} \]

Since \( f \geq g \) whenever \( f = \infty \), we have shown \( f \geq g \) a.e.

If equality holds in Eq. (10.8), then we know that \( g \leq f \) and \( f \leq g \) a.e., i.e. \( f = g \) a.e.

\[ \boxed{\text{Notice that we can not drop the } \sigma \text{-finiteness assumption in Lemma 10.24.}} \]

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. (10.8) holds even though \( f < g \) everywhere.

\[ \textbf{Definition 10.25.} \] Let \( (\Omega, \mathcal{B}, \mu) \) be a measure space and \( L^1(\mu) = L^1(\Omega, \mathcal{B}, \mu) \) denote the set of \( L^1(\mu) \) 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 \( \rho_1(f, g) = \|f - g\|_{L^1} \).

\[ \textbf{Warning:} \] in the future we will often not make much of a distinction between \( L^1(\mu) \) and \( L^1(\mu) \). On occasion this can be dangerous and this danger will be pointed out when necessary.

\[ \textbf{Remark 10.26.} \] 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.

\[ \text{We will see in later that} \]

\[ \|f\|_{L^p} = \left( \int |f|^p \ d\mu \right)^{1/p} \text{ for } f \in L^p(\mu) \]

is a norm and \( (L^p(\mu), \|\cdot\|_{L^p}) \) is a Banach space in this norm and in particular,

\[ \|f + g\|_p \leq \|f\|_p + \|g\|_p \text{ for all } f, g \in L^p(\mu). \]
Theorem 10.27 (Dominated Convergence Theorem). Suppose \( f_n, g_n, g \in L^1(\mu) \), \( f_n \to f \) a.e., \( |f_n| \leq g_n \in L^1(\mu) \), \( g_n \to g \) a.e. and \( \int \Omega g_n d\mu \to \int \Omega g d\mu \). Then \( f \in L^1(\mu) \) and
\[
\int \Omega f d\mu = \lim_{n \to \infty} \int \Omega f_n d\mu.
\]
(In most typical applications of this theorem \( g_n = g \in L^1(\mu) \) for all \( n \).)

Proof. Notice that \( |f| = \lim_{n \to \infty} |f_n| \leq \lim_{n \to \infty} |g_n| \leq g \) a.e. so that \( f \in L^1(\mu) \). 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,
\[
\int \Omega (g + f) d\mu = \int \Omega \liminf_{n \to \infty} (g_n + f_n) d\mu \leq \liminf_{n \to \infty} \int \Omega (g_n + f_n) d\mu
\]
\[
= \lim_{n \to \infty} \int \Omega g_n d\mu + \liminf_{n \to \infty} \left( \pm \int \Omega f_n d\mu \right)
\]
\[
= \int \Omega g d\mu + \liminf_{n \to \infty} \left( \pm \int \Omega f_n d\mu \right)
\]
Since \( \liminf_{n \to \infty} (-a_n) = -\limsup_{n \to \infty} a_n \), we have shown,
\[
\int \Omega g d\mu \pm \int \Omega f d\mu \leq \int \Omega g d\mu + \left\{ \liminf_{n \to \infty} \int \Omega f_n d\mu - \limsup_{n \to \infty} \int \Omega f_n d\mu \right\}
\]
and therefore
\[
\limsup_{n \to \infty} \int \Omega f_n d\mu \leq \int \Omega f d\mu \leq \liminf_{n \to \infty} \int \Omega f_n d\mu.
\]
This shows that \( \lim_{n \to \infty} \int \Omega f_n d\mu \) exists and is equal to \( \int \Omega f d\mu \). \( \square \)

Exercise 10.2. Give another proof of Proposition [10.21] by first proving Eq. [10.6] with \( f \) being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 9.39 along with the dominated convergence Theorem 10.27 to handle the general case.

Corollary 10.28. Let \( \{f_n\}_{n=1}^{\infty} \subset L^1(\mu) \) be a sequence such that \( \sum_{n=1}^{\infty} \|f_n\|_{L^1(\mu)} < \infty \), 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 \( \sum_{n=1}^{\infty} \|f_n\|_{L^1(\mu)} < \infty \) is equivalent to \( \sum_{n=1}^{\infty} |f_n| \in L^1(\mu) \). Hence \( \sum_{n=1}^{\infty} f_n \) is almost everywhere convergent and if \( S_N := \sum_{n=1}^{N} f_n \), then
\[
|S_N| \leq \sum_{n=1}^{N} |f_n| \leq \sum_{n=1}^{\infty} |f_n| \in L^1(\mu).
\]
So by the dominated convergence theorem,
\[
\int \Omega \left( \sum_{n=1}^{\infty} f_n \right) d\mu = \int \Omega \lim_{N \to \infty} S_N d\mu = \lim_{N \to \infty} \int \Omega S_N d\mu
\]
\[
= \lim_{N \to \infty} \sum_{n=1}^{N} \int \Omega f_n d\mu = \sum_{n=1}^{\infty} \int \Omega f_n d\mu.
\]

Example 10.29 (Sums as integrals). Suppose, \( \Omega = \mathbb{N} \), \( \mathcal{B} := 2^\mathbb{N} \), \( \mu \) is counting measure on \( \mathcal{B} \) (see Example 10.8), and \( f : \mathbb{N} \to \mathbb{C} \) is a function. From Example 10.8 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 10.28 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,
\[
\int \mathbb{N} f d\mu = \infty.
\]
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 10.30 (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 \( g \in L^1(\mu) \) such that \( |\frac{\partial f}{\partial t}(t, \cdot)| \leq g \) for each \( t \in J \).

Then \( f(t, \cdot) \in L^1(\mu) \) 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,

\[
|f(t, \omega) - f(t_0, \omega)| \leq g(\omega) |t - t_0| \quad \text{for all } t \in J \quad (10.9)
\]

and hence

\[
|f(t, \omega)| \leq |f(t, \omega) - f(t_0, \omega)| + |f(t_0, \omega)| \leq g(\omega) |t - t_0| + |f(t_0, \omega)|.
\]

This shows \( f(t, \cdot) \in L^1(\mu) \) 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,

\[
\lim_{t \to t_0} \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} = \frac{\partial f}{\partial t}(t, \omega) \quad \text{for all } \omega \in \Omega
\]

and by Eq. (10.9),

\[
\left| \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} \right| \leq g(\omega) \quad \text{for all } t \in J \text{ and } \omega \in \Omega.
\]

Therefore, we may apply the dominated convergence theorem to conclude

\[
\lim_{n \to \infty} \frac{G(t_n) - G(t_0)}{t_n - t_0} = \lim_{n \to \infty} \int_{\Omega} \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} \, d\mu(\omega)
\]

\[
= \int_{\Omega} \lim_{n \to \infty} \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} \, d\mu(\omega)
\]

\[
= \int_{\Omega} \frac{\partial f}{\partial t}(t_0, \omega) \, d\mu(\omega)
\]

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 10.31.** 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

\[
f'(z) = \sum_{n=1}^{\infty} n a_n (z - z_0)^{n-1} = \sum_{n=1}^{\infty} n a_n (z - z_0)^{n-1}.
\]

By induction it follows that \( f^{(k)} \) exists for all \( k \) and that

\[
f^{(k)}(z) = \sum_{n=0}^{\infty} n(n-1) \ldots (n-k+1) a_n (z - z_0)^{n-1}.
\]

**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 10.30 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

\[
|g'(z, n)| = |n a_n (z - z_0)^{n-1}| \leq n |a_n| \rho^{n-1}
\]

\[
\leq \frac{1}{r} (\frac{\rho}{r})^{n-1} |a_n| r^n \leq \frac{1}{r} (\frac{\rho}{r})^{n-1} M
\]
and the function \( G(n) := \frac{d}{dz} n (z^n - 1) \) is summable (by the Ratio test for example), we may use \( G \) as our dominating function. It then follows from Corollary 10.30

\[
f(z) = \int_X g(z, n) d\mu(n) = \sum_{n=0}^{\infty} a_n(z - z_0)^n
\]
is complex differentiable with the differential given as in Eq. (10.10).

**Definition 10.32 (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 

\[
M_X(t) := \mathbb{E}[e^{tX}].
\]

**Proposition 10.33.** Suppose there exists \(\varepsilon > 0\) such that \(\mathbb{E}[e^{\varepsilon X}] < \infty\), then 

\(M_X(t)\) is a smooth function of \(t \in (0, \varepsilon)\)

\[
M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n \text{ if } |t| \leq \varepsilon.
\]  

(10.11) 

In particular,

\[
\mathbb{E}X^n = \left( \frac{d}{dt} \right)^n |t=0| M_X(t) \text{ for all } n \in \mathbb{N}_0.
\]

(10.12) 

**Proof.** If \(|t| \leq \varepsilon\), then

\[
\mathbb{E} \left[ \sum_{n=0}^{\infty} \frac{|t^n|}{n!} |X|^n \right] \leq \mathbb{E} \left[ \sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!} |X|^n \right] = \mathbb{E}[e^{\varepsilon |X|}] < \infty.
\]

It follows that, for \(|t| \leq \varepsilon\)

\[
M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E} \left[ \sum_{n=0}^{\infty} \frac{t^n}{n!} X^n \right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n.
\]

Equation (10.12) now is a consequence of Corollary 10.31. 

**Exercise 10.3.** 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 \(x^\omega := x_1^\omega \ldots x_d^\omega\). Further suppose that \(f: \Omega \to \mathbb{C}\) is function and \(r_i > 0\) for \(1 \leq i \leq d\) such that 

\[
\sum_{\omega \in \Omega} |f(\omega)||r^\omega| = \int_{\Omega} |f(\omega)||r^\omega| d\mu(\omega) < \infty,
\]

where \(r := (r_1, \ldots, r_d)\). Show;

1. There is a constant, \(C < \infty\) such that \(|f(\omega)| \leq \frac{C}{r^\omega}\) for all \(\omega \in \Omega\).
2. Let 

\[
U := \{x \in \mathbb{R}^d : |x_i| < r_i \forall i\} \text{ and } \bar{U} = \{x \in \mathbb{R}^d : |x_i| \leq r_i \forall i\}
\]

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

\[
F(x) = \sum_{\omega \in \Omega} f(\omega)x^\omega
\]

is continuous on \(\bar{U}\).

3. Show, for all \(x \in U\) and \(1 \leq i \leq d\), that 

\[
\partial_i F(x) = \sum_{\omega \in \Omega} \omega_i f(\omega)x^{\omega-e_i}
\]

where \(e_i = (0, \ldots, 0, 1, 0, \ldots, 0)\) is the \(i\)th standard basis vector on \(\mathbb{R}^d\).

4. For any \(\alpha \in \Omega\), let \(\partial^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \ldots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}\) and \(\alpha! := \prod_{i=1}^{d} \alpha_i!\). Explain why we may now conclude that

\[
\partial^\alpha F(x) = \sum_{\omega \in \Omega} \alpha! f(\omega)x^{\omega-\alpha}
\]

for all \(x \in U\).

5. Conclude that \(f(\alpha) = \frac{\partial F(0)}{\partial \alpha}\) 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 \(\omega \in \Omega\).

**10.2.1 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

\[
0 \leq |X - Y|^2 = |X|^2 + |Y|^2 - 2|X||Y|,
\]

it follows that

\[
|XY| \leq \frac{1}{2}|X|^2 + \frac{1}{2}|Y|^2 \in L^1(P).
\]

In particular by taking \(Y = 1\), we learn that \(|X| \leq \frac{1}{2}(1 + |X|^2)\) which shows that every square integrable random variable is also integrable.

**Definition 10.34.** The covariance, \(\text{Cov}(X, Y)\), of two square integrable random variables, \(X\) and \(Y\), is defined by

\[
\text{Cov}(X, Y) = \mathbb{E}[(X - a_X)(Y - a_Y)] = \mathbb{E}[XY] - a_Xa_Y
\]
where \( a_X := \mathbb{E}X \) and \( a_Y := \mathbb{E}Y \). The variance of \( X \),
\[
\text{Var} (X) := \text{Cov} (X, X) = \mathbb{E} [X^2] - (\mathbb{E}X)^2
\]  
(10.14)

We say that \( X \) and \( Y \) are uncorrelated if \( \text{Cov} (X,Y) = 0 \), 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 \( \text{Cov} (X_i, X_j) = 0 \) for all \( i \neq j \).

It follows from Eq. (10.14) that
\[
\text{Var} (X) \leq \mathbb{E} [X^2] \quad \text{for all} \quad X \in L^2 (P).
\]  
(10.15)

Lemma 10.35. The covariance function, \( \text{Cov} (X,Y) \) is bilinear in \( X \) and \( Y \) and \( \text{Cov} (X, Y) = 0 \) if either \( X \) or \( Y \) is constant. For any constant \( k \), \( \text{Var} (X + k) = \text{Var} (X) \) and \( \text{Var} (kX) = k^2 \text{Var} (X) \). If \( \{X_k\}_{k=1}^n \) are uncorrelated \( L^2 (P) \) random variables, then
\[
\text{Var} (S_n) = \sum_{k=1}^n \text{Var} (X_k).
\]

Proof. We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove \( \text{Var} (X + k) = \text{Var} (X) \);
\[
\text{Var} (X + k) = \text{Cov} (X + k, X + k) = \text{Cov} (X + k, X) + \text{Cov} (X + k, k)
\]
\[
= \text{Cov} (X + k, X) = \text{Cov} (X, X) + \text{Cov} (k, k)
\]
\[
= \text{Cov} (X, X) = \text{Var} (X),
\]
wherein we have used the bilinearity of \( \text{Cov} (\cdot, \cdot) \) and the property that \( \text{Cov} (Y, k) = 0 \) whenever \( k \) is a constant.

Exercise 10.4 (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 \( \sigma^2 := \text{Var} (X_k) \) (these are independent of \( k \)). Show:
\[
\mathbb{E} \left( \frac{S_n}{n} \right) = \mu,
\]
\[
\mathbb{E} \left( \frac{S_n}{n} - \mu \right)^2 = \text{Var} \left( \frac{S_n}{n} \right) = \frac{\sigma^2}{n}, \quad \text{and}
\]
\[
P \left( \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \leq \frac{\sigma^2}{n \varepsilon^2}
\]
for all \( \varepsilon > 0 \) and \( n \in \mathbb{N} \). (Compare this with Exercise 5.18.)

10.2.2 Some Discrete Distributions

Definition 10.36 (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
\[
G_N (z) := \mathbb{E} [z^N] = \sum_{n=0}^\infty P (N = n) z^n \quad \text{for} \quad |z| \leq 1.
\]  
(10.16)

By Corollary 10.31 it follows that \( P (N = n) = \frac{1}{\Pi} G_N^{(n)} (0) \) so that \( G_N \) can be used to completely recover the distribution of \( N \).

Proposition 10.37 (Generating Functions). The generating function satisfies,
\[
G_N^{(k)} (z) = \mathbb{E} [N (N - 1) \ldots (N - k + 1) z^{N-k}] \quad \text{for} \quad |z| < 1
\]
and
\[
G^{(k)} (1) = \lim_{z \uparrow 1} G^{(k)} (z) = \mathbb{E} [N (N - 1) \ldots (N - k + 1)],
\]
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 \( \mathbb{E}N^2 < \infty \),
\[
\text{Var} (N) = G'' (1) + G' (1) - [G' (1)]^2.
\]  
(10.17)

Proof. By Corollary 10.31 for \( |z| < 1 \),
\[
G_N^{(k)} (z) = \sum_{n=0}^\infty P (N = n) \cdot n (n-1) \ldots (n-k+1) z^{n-k}
\]
\[
= \mathbb{E} [N (N - 1) \ldots (N - k + 1) z^{N-k}].
\]  
(10.18)

Since, for \( z \in (0,1) \),
\[
0 \leq N (N - 1) \ldots (N - k + 1) z^{N-k} \wedge N (N - 1) \ldots (N - k + 1) \quad \text{as} \quad z \uparrow 1,
\]
we may apply the MCT to pass to the limit as \( z \uparrow 1 \) in Eq. (10.18) to find,
\[
G^{(k)} (1) = \lim_{z \uparrow 1} G^{(k)} (z) = \mathbb{E} [N (N - 1) \ldots (N - k + 1)].
\]

Exercise 10.5 (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 \( \text{Var} (N) \).
1. Bernoulli($p$) : $P(N = 1) = p$ and $P(N = 0) = 1 - p$. You should find $\text{EN} = p$ and $\text{Var}(N) = p - p^2$.
2. Binomial($n=p, \lambda$) : $P(N = k) = \binom{n}{k}p^k(1-p)^{n-k}$ for $k = 0, 1, \ldots, n$. ($P(N = k)$ is the probability of $k$ successes in a sequence of $n$ independent yes/no experiments with probability of success being $p$.) You should find $\text{EN} = np$ and $\text{Var}(N) = n(p - p^2)$.
3. Geometric($p$) : $P(N = k) = p(1 - p)^{k-1}$ for $k \in \mathbb{N}$. ($P(N = k)$ is the probability that the $k^{th}$ trial is the first time of success out of a sequence of independent trials with probability of success being $p$.) You should find $\text{EN} = 1/p$ and $\text{Var}(N) = \frac{1}{p^2}$.
4. Poisson($\lambda$) : $P(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for all $k \in \mathbb{N}_0$. You should find $\text{EN} = \lambda = \text{Var}(N)$.

**Exercise 10.6.** 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 $n$,

$$P(\text{Binomial} (n,p) = k) \approx P(\text{Poisson} (pn) = k) = \frac{(pn)^k}{k!} e^{-pn}.$$ (We will come back to the Poisson distribution and the related Poisson process later on.)

### 10.3 Integration on $\mathbb{R}$

**Notation 10.38** If $m$ is Lebesgue measure on $\mathbb{R}$, $f$ is a non-negative Borel measurable function and $a < b$ with $a, b \in \mathbb{R}$, 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 10.39.** Suppose $-\infty < a < b < \infty$, $f \in C([a,b], \mathbb{R})$ and $m$ be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi = \{a = a_0 < a_1 < \cdots < a_n = b\},$$

let

$$\text{mesh}(\pi) := \max\{|a_j - a_{j-1}| : j = 1, \ldots, n\}$$

and

$$f_\pi(x) := \sum_{i=0}^{n-1} f(a_i) 1_{(a_i, a_{i+1}]}(x).$$

Then

$$\int_a^b f_\pi \, dm = \sum_{i=0}^{n-1} f(a_i) m((a_i, a_{i+1})) = \sum_{i=0}^{n-1} f(a_i) (a_{i+1} - a_i)$$

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k \to \infty} \text{mesh}(\pi_k) = 0$, we know that

$$\lim_{k \to \infty} \int_a^b f_{\pi_k} \, dm = \int_a^b f(x) \, dx \quad (10.19)$$

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 $|f_{\pi_k}(x)| \leq g(x) := M 1_{(a,b)}(x)$ for all $x \in (a, b)$ where $M := \max_{x \in [a, b]} |f(x)| < \infty$. Since $\int_a^b g \, dm = M(b-a) < \infty$, 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. (10.19) 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 10.82 below for a more general statement along these lines.

**Theorem 10.40 (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 $F(x) := \int_a^x f(y) \, dm(y), \lim_{x \to a^+} 1_{(a,x)}(y) f(y) \, dm(y), \lim_{x \to b^-} 1_{(a,x]}(y) f(y) \, dm(y)$ is an $L^1$ – function, it follows from the dominated convergence Theorem 10.27 that $F$ is continuous on $[a,b]$. Simple manipulations show,
\[
\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| = \frac{1}{|h|} \left\{ \begin{array}{ll}
\int_{x}^{x+h} |f(y) - f(x)| \, dy \text{ if } h > 0 \\
\int_{x}^{x+h} |f(y) - f(x)| \, dy \text{ if } h < 0 
\end{array} \right.
\]

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

\[
\int_{a}^{b} f(x) \, dx = F(b) - F(a) = (G(b) + C) - (G(a) + C) = G(b) - G(a).
\]

We can use the above results to integrate some non-Riemann integrable functions:

**Example 10.41.** For all \( \lambda > 0 \),

\[
\int_{0}^{\infty} e^{-\lambda x} \, dx = \lambda^{-1}
\]

and

\[
\int_{\mathbb{R}}^{\infty} \frac{1}{1 + x^2} \, dx = \pi.
\]

The proof of these identities are similar. By the monotone convergence theorem, Example 10.39 and the fundamental theorem of calculus for Riemann integrals (or Theorem 10.40 below),

\[
\int_{0}^{\infty} e^{-\lambda x} \, dx = \lim_{N \to \infty} \int_{0}^{N} e^{-\lambda x} \, dx = \lim_{N \to \infty} \int_{0}^{N} e^{-\lambda x} \, dx = -\lim_{N \to \infty} \lambda e^{-\lambda x} \bigg|_{0}^{N} = \lambda^{-1}
\]

and

\[
\int_{\mathbb{R}}^{\infty} \frac{1}{1 + x^2} \, dx = \lim_{N \to \infty} \int_{-N}^{N} \frac{1}{1 + x^2} \, dx = \lim_{N \to \infty} \int_{-N}^{N} \frac{1}{1 + x^2} \, dx = \lim_{N \to \infty} \left[ \tan^{-1}(N) - \tan^{-1}(-N) \right] = \pi.
\]

Let us also consider the functions \( x^{-p} \). Using the MCT and the fundamental theorem of calculus,

\[
\int_{(0, 1]}^{\infty} \frac{1}{x^p} \, dx = \lim_{n \to \infty} \int_{0}^{1} \frac{1}{(\frac{1}{n})^p} \, dx \frac{1}{x^p} \, dx = \lim_{n \to \infty} \int_{0}^{1} \frac{1}{x^p} \, dx = \lim_{n \to \infty} \frac{x^{-p+1}}{-1} \bigg|_{0}^{1/n} = \left\{ \begin{array}{ll}
\frac{1}{p} & \text{if } p < 1 \\
\infty & \text{if } p > 1
\end{array} \right.
\]

If \( p = 1 \) we find

\[
\int_{(0, 1]} \frac{1}{x} \, dx = \lim_{n \to \infty} \int_{0}^{1} \frac{1}{x} \, dx = \lim_{n \to \infty} \ln(x) \bigg|_{1/n}^{1} = \infty.
\]

**Exercise 10.7.** Show

\[
\int_{1}^{\infty} \frac{1}{x^p} \, dx = \left\{ \begin{array}{ll}
\infty & \text{if } p \leq 1 \\
\frac{1}{p-1} & \text{if } p > 1.
\end{array} \right.
\]

**Example 10.42 (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

\[
\int_{\alpha}^{\beta} \left( \sum_{n=0}^{\infty} a_n x^n \right) \, dx = \sum_{n=0}^{\infty} a_n \int_{\alpha}^{\beta} x^n \, dx = \sum_{n=0}^{\infty} a_n \frac{\beta^{n+1} - \alpha^{n+1}}{n+1}
\]

for all \( -R < \alpha < \beta < R \). Indeed this follows from Corollary 10.28 since

\[
\sum_{n=0}^{\infty} \left| a_n \right| \left| x \right|^n \, dx \leq \sum_{n=0}^{\infty} \left( \int_{0}^{\beta} \left| a_n \right| \left| x \right|^n \, dx + \int_{0}^{\alpha} \left| a_n \right| \left| x \right|^n \, dx \right) \leq 2r \sum_{n=0}^{\infty} \left| a_n \right| \left| r \right|^n < \infty
\]

where \( r = \max(|\beta|, |\alpha|) \).

**Example 10.43.** Let \( \{r_n\}_{n=1}^{\infty} \) be an enumeration of the points in \( Q \cap [0, 1] \) and define

\[
f(x) = \sum_{n=1}^{\infty} \frac{2^{-n}}{\sqrt{|x - r_n|}}
\]

with the convention that

\[
\frac{1}{\sqrt{|x - r_n|}} = 5 \text{ if } x = r_n.
\]

Since, By Theorem 10.40
\[ \int_0^1 \frac{1}{\sqrt{|x-r_n|}} \, dx = \int_0^{r_n} \frac{1}{\sqrt{r_n-x}} \, dx + \int_{r_n}^{\infty} \frac{1}{\sqrt{r_n-x}} \, dx \]
\[ = 2\sqrt{r_n-x} \bigg|_{r_n}^{1} - 2\sqrt{r_n-x} \bigg|_{0}^{r_n} = 2 \left( \sqrt{1-r_n} - \sqrt{r_n} \right) \leq 4, \]

we find

\[ \int_{[0,1]} f(x) \, dm(x) = \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x-r_n|}} \, dx \leq \sum_{n=1}^{\infty} 2^{-n} 4 = 4 < \infty. \]

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 10.44.** The following limit holds,

\[ \lim_{n \to \infty} \int_0^1 \left( 1 - \frac{x}{n} \right)^n \, dm(x) = 1. \quad (10.20) \]

**DCT Proof.** To verify this, let \( f_n(x) := \left( 1 - \frac{x}{n} \right)^n \mathbb{1}_{[0,n]}(x) \). Then \( \lim_{n \to \infty} f_n(x) = e^{-x} \) for all \( x \geq 0 \). Moreover by simple calculus

\[ 1 - x \leq e^{-x} \text{ for all } x \in \mathbb{R}. \]

Therefore, for \( x < n \), we have

\[ 0 \leq 1 - \frac{x}{n} \leq e^{-x/n} \implies \left( 1 - \frac{x}{n} \right)^n \leq \left[ e^{-x/n} \right]^n = e^{-x}, \]

from which it follows that

\[ 0 \leq f_n(x) \leq e^{-x} \text{ for all } x \geq 0. \]

From Example 10.43\footnote{Since \( y = 1 - x \) is the tangent line to \( y = e^{-x} \) at \( x = 0 \) and \( e^{-x} \) is convex up, it follows that \( 1 - x \leq e^{-x} \) for all \( x \in \mathbb{R} \).}, we know

\[ \int_{\mathbb{R}} e^{-x} \, dm(x) = 1 < \infty, \]

so that \( e^{-x} \) is an integrable function on \([0, \infty)\). Hence by the dominated convergence theorem,

\[ \lim_{n \to \infty} \int_0^n \left( 1 - \frac{x}{n} \right)^n \, dm(x) = \lim_{n \to \infty} \int_0^\infty f_n(x) \, dm(x) \]
\[ = \int_0^\infty \lim_{n \to \infty} f_n(x) \, dm(x) = \int_0^\infty e^{-x} \, dm(x) = 1. \]

**MCT Proof.** The limit in Eq. (10.20) 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 \( n > x \),

\[ \frac{d}{dn} \ln f_n(x) = \frac{d}{dn} \left[ n \ln \left( 1 - \frac{x}{n} \right) \right] = \ln \left( 1 - \frac{x}{n} \right) + \frac{n}{1 - x/n} \]
\[ = \ln \left( 1 - \frac{x}{n} \right) + \frac{\frac{x}{n}}{1 - \frac{x}{n}} = h(x/n) \]

where, for \( 0 \leq y < 1 \),

\[ h(y) := \ln(1-y) + \frac{y}{1-y}. \]

Since \( h(0) = 0 \) and

\[ h'(y) = -\frac{1}{1-y} + \frac{1}{1-y} + \frac{y}{(1-y)^2} > 0 \]

it follows that \( h \geq 0 \). Thus we have shown, \( f_n(x) \uparrow e^{-x} \) as \( n \to \infty \) as claimed.

**Example 10.45.** Suppose that \( f_n(x) := n \mathbb{1}_{[0,\frac{1}{n}]}(x) \) for \( n \in \mathbb{N} \). Then \( \lim_{n \to \infty} f_n(x) = 0 \) for all \( x \in \mathbb{R} \) while

\[ \lim_{n \to \infty} \int_{\mathbb{R}} f_n(x) \, dx = \lim_{n \to \infty} 1 = 1 \neq 0 = \int_{\mathbb{R}} \lim_{n \to \infty} f_n(x) \, dx. \]

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

\[ \int_{\mathbb{R}} g(x) \, dx = \sum_{n=1}^{\infty} n \cdot \left( \frac{1}{n} - \frac{1}{n+1} \right) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty. \]
Example 10.46 (Jordan’s Lemma). In this example, let us consider the limit:

$$\lim_{n \to \infty} \int_0^\pi \cos \left( \frac{\theta}{n} \right) e^{-n \sin(\theta)} d\theta.$$ 

Let

$$f_n(\theta) := 1_{(0,\pi)}(\theta) \cos \left( \frac{\theta}{n} \right) e^{-n \sin(\theta)}.$$ 

Then

$$|f_n| \leq 1_{(0,\pi)}(\theta) \in L^1(m)$$

and

$$\lim_{n \to \infty} f_n(\theta) = 1_{(0,\pi)}(\theta) 1_{\{\pi\}}(\theta).$$

Therefore by the D.C.T.,

$$\lim_{n \to \infty} \int_0^\pi \cos \left( \frac{\theta}{n} \right) e^{-n \sin(\theta)} d\theta = \int_\mathbb{R} 1_{\{\pi\}}(\theta) dm(\theta) = m(\{\pi\}) = 0.$$ 

Example 10.47. Recall from Example 10.41 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x} dm(x) \text{ for all } \lambda > 0.$$ 

Let $\varepsilon > 0$. For $\lambda \geq 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

$$0 \leq \left( -\frac{d}{d\lambda} \right)^n e^{-\lambda x} = x^n e^{-\lambda x} \leq C_n(\varepsilon) e^{-\varepsilon x}.$$

Using this fact, Corollary 10.30 and induction gives

$$n! \lambda^{-n-1} = \left( -\frac{d}{d\lambda} \right)^n (n-1) \lambda^{-1} = \int_{[0,\infty)} \left( -\frac{d}{d\lambda} \right)^n e^{-\lambda x} dm(x)$$

$$= \int_{[0,\infty)} x^n e^{-\lambda x} dm(x).$$

That is

$$n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x} dm(x). \quad (10.21)$$

Remark 10.48. Corollary 10.30 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 10.30 to $g(t) := \int_0^\infty 1_{x \leq t} dm(x)$,

$$\dot{g}(t) = \frac{d}{dt} \int_0^\infty 1_{x \leq t} dm(x) = \int_0^\infty \frac{\partial}{\partial t} 1_{x \leq t} dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t} 1_{x \leq t} = 0$ 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 10.30 has been violated in this example.)

Exercise 10.8 (Folland 2.28 on p. 60.). Compute the following limits and justify your calculations:

1. $\lim_{n \to \infty} \int_0^\infty \frac{\sin(\frac{t}{n})}{(1+\frac{t}{n})^n} \, dx$.
2. $\lim_{n \to \infty} \int_1^n \frac{1}{(1+nx^2)^n} \, dx$.
3. $\lim_{n \to \infty} \int_0^n \frac{n \sin(nx/n)}{x(1+x^2)} \, dx$.
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 10.9 (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:

$$\int_\mathbb{R} f'(x) \cdot g(x) \, dx = -\int_\mathbb{R} f(x) \cdot g'(x) \, dx. \quad (10.22)$$

Similarly, show that if $f, g : [0, \infty) \to [0, \infty)$ are two continuously differentiable functions such that $f'g, fg'$, and $fg$ are all Lebesgue integrable functions on $[0, \infty)$, then

$$\int_0^\infty f'(x) \cdot g(x) \, dx = -\int_0^\infty f(x) \cdot g'(x) \, dx. \quad (10.23)$$

Outline: 1. First notice that Eq. (10.22) holds if $f(x) = 0$ for $|x| \geq N$ 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 $|x| \leq 1$ and $\psi(x) = 0$ if $|x| \geq 2$. For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$.

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. (10.23).

Definition 10.49 (Gamma Function). The Gamma function, $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \quad (10.24)$$

(The reader should check that $\Gamma(x) < \infty$ for all $x > 0$.)
Here are some of the more basic properties of this function.

**Example 10.50 (Γ - 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
\[
\Gamma(n + 1) = n! \text{ for all } n \in \mathbb{N}. \tag{10.25}
\]

(Equation 10.25 was also proved in Eq. (10.21).)

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 12.29) that
\[
\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{10.26}
\]

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 \((2n-1)!! = (2n-1)(2n-3)\ldots 3 \cdot 1\) for \( n \in \mathbb{N}\). Letting the mesh of \( \Pi \) tend to zero using the uniform continuity of \( f \)
then shows \( \lambda(f) = \mu(f) \).

### 10.4 Densities and Change of Variables Theorems

**Exercise 10.10 (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
   \[
   \int_X f d\nu = \int_X f \rho d\mu. \tag{10.27}
   \]

**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 \( |f| \in L^1(\mu) \)
   and if \( f \in L^1(\nu) \) then Eq. (10.27) still holds.

**Notation 10.51** It is customary to informally describe \( \nu \) defined in Exercise 10.10 by writing \( d\nu = \rho d\mu \).

**Exercise 10.11 (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
   \[
   \int_Y g d\nu = \int_X (g \circ f) d\mu \tag{10.28}
   \]
   for all measurable functions \( g : Y \to [0, \infty] \). **Hint:** see the hint from Exercise 10.10.

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. (10.28) holds for all \( g \in L^1(\nu) \).

**Example 10.52.** Suppose \((\Omega, \mathcal{B}, P)\) is a probability space and \(\{X_i\}_{i=1}^n\) are random variables on \(\Omega\) with \(\nu := \text{Law} P (X_1, \ldots, X_n)\), then
\[
\mathbb{E}[g(X_1, \ldots, X_n)] = \int_{\mathbb{R}^n} g d\nu
\]
for all \( g : \mathbb{R}^n \to \mathbb{R} \) which are Borel measurable and either bounded or non-negative. This follows directly from Exercise 10.11 with \( f := (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n \) and \( \mu = P \).

**Remark 10.53.** As a special case of Example 10.52 suppose that \(X\) is a random variable on a probability space, \((\Omega, \mathcal{B}, P), F(x) := P(X \leq x)\). Then
\[ \mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) dF(x) \]  
(10.29)

where \( dF(x) \) is shorthand for \( d\mu_F(x) \) and \( \mu_F \) is the unique probability measure on \((\mathbb{R}, B_{\mathbb{R}})\) 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

\[ d\mu_F(x) = F'(x) \, dm(x) \]  
(10.30)

and Eq. (10.29) may be written as

\[ \mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x) \, dm(x). \]  
(10.31)

To verify Eq. (10.30) it suffices to observe, by the fundamental theorem of calculus, that

\[ \mu_F((a,b]) = F(b) - F(a) = \int_a^b F'(x) \, dx = \int_{[a,b]} F' \, dm. \]

From this equation we may deduce that \( \mu_F(A) = \int_A F' \, dm \) for all \( A \in B_{\mathbb{R}} \). Equation (10.31) now follows from Exercise 10.11.

**Exercise 10.12.** 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 \( B_{\mathbb{R}} \) and

\[ \nu(A) = m(F(A)) = m((F^{-1})^{-1}(A)) = (F^{-1}m)(A) \]

for all \( A \in B_{\mathbb{R}} \). Show \( d\nu = F' \, dm \). Use this result to prove the change of variable formula,

\[ \int_{\mathbb{R}} h \circ F \cdot F' \, dm = \int_{\mathbb{R}} h \, dm \]  
(10.32)

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 7.7 to conclude \( d\nu = F' \, dm \) on all of \( B_{\mathbb{R}} \). To prove Eq. (10.32) apply Exercise 10.11 with \( g = h \circ F \) and \( f = F^{-1} \).

### 10.5 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 \( \|f\|_\infty \coloneqq \max_{x \in X} |f(x)| \).

**Definition 10.54.** A positive linear functional, \( I \), on \( C_c(X) \) is a linear map, \( I: C_c(X) \to \mathbb{R} \) such that \( I(f) \geq 0 \) whenever \( f \geq 0 \).

**Example 10.55.** If \( \mu \) is any measure on \((X, B_X)\) with \( \mu(K) < \infty \) for all compact subset \( K \subset X \), then \( I(f) := \mu(f) := \int_X f \, d\mu \) 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 10.56.** 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 \( |I(f)| \leq C_K \|f\|_\infty \) for all \( f \in C_c(X) \) with \( \text{supp}(f) \subset K \).

**Proof.** By Urysohn’s Lemma for metric spaces (see Lemma 2.22 or also ??), there exists \( \varphi \in C_c([0,1]) \) such that \( \varphi = 1 \) on \( K \). Then for all \( f \in C_c(X, \mathbb{R}) \) such that \( \text{supp}(f) \subset K \), \( |I(f)| \leq \|f\|_\infty \|\varphi\|_{C_c([0,1])} \). Hence \( \|\varphi\|_{C_c([0,1])} = \int \varphi(x) \, dx \leq \sup_{x \in X} \varphi(x) \leq \sup_{x \in K} \varphi(x) \). Letting \( C_K := \|\varphi\|_{C_c([0,1])} \), we have shown that \( |I(f)| \leq C_K \|f\|_\infty \) for all \( f \in C_c(X, \mathbb{R}) \) with \( \text{supp}(f) \subset K \). For general \( f \in C_c(X, \mathbb{C}) \) with \( \text{supp}(f) \subset K \), choose \( \alpha = 1 \) such that \( \alpha I(f) \geq 0 \). Then

\[ |I(f)| = |\alpha I(f)| = |\alpha I(f)| = |I(\alpha f)| \leq C_K \|\alpha f\|_\infty \leq C_K \|f\|_\infty. \]

**Notation 10.57** Let \( C(X) = \{f_X: X \to [0, \infty) | \exists f_n \in C(X) \ni f_n \downarrow f_X\} \)

where \( f_n \downarrow f_X \) means \( f_n(x) \geq f_{n+1}(x) \) for all \( n \in \mathbb{N} \) and \( x \in X \) and \( f_X(x) = \lim_{n \to \infty} f_n(x) \in [-\infty, \infty] \) for all \( x \in X \).

**Example 10.58.** If \( F \) is a closed subset of \( X \), then \( 1_F \in C(X) \). To prove this first observe that \( \rho_F(x) := \inf \{\rho(x, y) : y \in F\} \geq 0 \) 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 \( h_n \downarrow 1_{\{0\}} \). It then follows that \( f_n = h_n \circ \rho_F \in C(X) \) with \( f_n \downarrow 1_{\{0\}} \).

**Theorem 10.59.** 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 \( f_n \downarrow f_X \in C(X) \) with \( f_X \leq f \), then \( \lim_{n \to \infty} \lambda(f_n) \leq \lambda(f) \).
2. If \( f_n, g_n \in C(X) \) such that \( f_n \downarrow f_X \) and \( g_n \downarrow g_X \) \( \text{pointwise} \) with \( f_X \leq g_X \), then \( \lim_{n \to \infty} \lambda(f_n) \leq \lim_{n \to \infty} \lambda(g_n) \). In particular if \( f_n \downarrow f_X \) and \( g_n \downarrow g_X \), 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 \( C(X) \) by setting 
\[ \lambda(f) := \lim_{n \to \infty} \lambda(f_n) \]
whenever \( f_n \in C(X) \) with \( f_1 \leq f_2 \). The extension satisfies:

a) \( \lambda(f) \leq \lambda(g) \) for all \( f, g \in C(X) \) with \( f \leq g \)

b) for all \( f, g \in C(X) \) and \( a \geq 0 \), \( f + ag \in C(X) \) and
\[ \lambda(f + ag) = \lambda(f) + a\lambda(g) \]  
(10.33)

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

1. If \( f_n, f \in C(X) \) and \( f_\infty \in C(X) \) are as in item 1., then \( f_n \leq f \) uniformly on \( X \). Because \( \lambda \) is a bounded linear functional (see Propositions 10.56) it follows that
\[ \lim_{n \to \infty} \lambda(f_n) \leq \lim_{n \to \infty} \lambda(f) \]  

2. If \( f_n, g_n \in C(X) \) and \( f_\infty, g_\infty \in C(X) \) are as in item 2., then for each \( m \in \mathbb{N} \),
\[ f_n \leq f_n \vee g_n \leq f \vee g_m = g_m \]  

Hence by item 1., \( \lim_{n \to \infty} \lambda(f_n) \leq \lambda(g_m) \) and the result follows by letting \( m \to \infty \) in the last inequality.

3. The monotonic assertion of item 3a. follows directly from item 1. The proof 3b. is straightforward and will be left to the reader.

**Remark 10.60.** Because of Theorem 10.59 and Example 10.58 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 \( \lambda(1_{F_1}) \leq \lambda(1_{F_2}) \). It is fact true it this level of generality that there exists a Borel measure, \( \mu \), on \((X, B_X)\) 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 10.61.** To each \( b \in \mathbb{R} \) and \( \varepsilon > 0 \) let \( \varphi_{b, \varepsilon} \in C(\mathbb{R}, [0, 1]) \) be defined by
\[ \varphi_{b, \varepsilon}(x) = \begin{cases} 
1 & \text{if } x \leq b \\
1 - \frac{x - b}{\varepsilon} & \text{if } b \leq x \leq b + \varepsilon \\
0 & \text{if } x \geq b + \varepsilon. 
\end{cases} \]  
(10.34)

The key property of these functions are that \( \varphi_{b, \varepsilon} \downarrow 1_{(-\infty, b]} \) as \( \varepsilon \downarrow 0 \).

**Theorem 10.62 (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 \( B_X \) such that
\[ \lambda(f) = \int_0^1 f \, d\mu \]  
for all \( f \in C(X) \).

**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 \( F(b) := \lambda([0, b]) \) which is a well defined non-decreasing function on \([0, 1]\) by Remark 10.60. It is also true that \( F \) is right continuous. To prove this choose \( \delta_n \) so that
\[ 0 < \delta_n < \frac{1}{n} - \frac{1}{n + 1} = \frac{1}{n(n + 1)}. \]  
(10.36)

Then
\[ F(b) = \lim_{n \to \infty} F \left( b + \frac{1}{n} \right) \leq \lim_{n \to \infty} \lambda(f_n) = F(b) \leq F(b+) \]  
and hence
\[ F(b+) = \lim_{n \to \infty} F(b + \frac{1}{n}) \leq \lim_{n \to \infty} \lambda(f_n) = F(b) \leq F(b+) \]  

2. By item 1. and Corollary 10.35, there exists a unique measure \( \mu \) on \( B_{[0, 1]} \) such that \( F(b) = \mu([0, b]) \) for all \( b \in [0, 1] \).

3. If \( a, b \in [0, 1] \) with \( a < b \), then \( \lambda([a, b]) \leq \mu([a, b]) \). Since
\[ \mu([0, b]) = F(b) = \lambda([0, b]) \]  
we may assume that \( 0 < a \leq b \leq 1 \). For any \( 0 < \delta < a \) we have,
\[ 1_{[0, \delta]} + 1_{[a, b]} \leq 1_{[0, b]} \]  
and therefore
\[ F(\delta) + \lambda([a, b]) = \lambda(1_{[0, \delta]} + 1_{[a, b]}) \leq \lambda(1_{[0, b]}) = F(b) \]  
From this equation it follows that
\[ \lambda([a, b]) \leq F(b) = F(\delta) + \mu([a, b]) \]

4. Since \( \sum_{x \in [0, 1]} \mu(\{x\}] \leq \mu([1]) < \infty \), it follows that \( U := \{x \in [0, 1] \mid \mu(\{x\}) > 0\} \) is at most a countable set.

5. (Verification of Eq. 10.35.) First assume \( f \in C(X) \) is non-negative, i.e. \( f \geq 0 \). 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 \( 1 \leq j < n \). Let
\[ c_i^n := \max \{|f(x)| : a_i \leq x \leq a_{i+1}\} \]  
for \( 0 \leq i < n \),
and define $f_\pi$ and $\hat{f}_\pi$ by

$$f_\pi = c_0 1_{[0,a_1]} + \sum_{i=1}^{n-1} c_i 1_{[a_i,a_{i+1}]}$$

and

$$\hat{f}_\pi = c_0 1_{[0,a_1]} + \sum_{i=1}^{n-1} c_i 1_{[a_i,a_{i+1}]} \in C([0,1])_+.$$  

It is easy to see that $f \leq f_\pi \leq \hat{f}_\pi$ and that $f_\pi \to f$ uniformly on $X$. Hence it follows that

$$\lambda(f) \leq \lambda(\hat{f}_\pi) = c_0 \lambda(1_{[0,a_1]}) + \sum_{i=1}^{n-1} c_i \lambda(1_{[a_i,a_{i+1}]})$$

$$\leq c_0 \mu([0,a_1]) + \sum_{i=1}^{n-1} c_i \mu([a_i,a_{i+1}])$$

$$= c_0 \mu([0,a_1]) + \sum_{i=1}^{n-1} c_i \mu((a_i,a_{i+1}])$$

$$= \mu(f_\pi) \to \mu(f) \text{ as mesh } (\pi) \to 0$$

wherein we have used $\mu(\{a_i\}) = 0$ for all $1 \leq i < n$ in the second to last inequality and the dominated convergence theorem to take the limit. Thus we have shown that $\lambda(f) \leq \mu(f)$ for all $f \geq 0$. Since $\lambda(1) = \mu(1)$, if $f \in C(X)$ and $M = \|f\|_\infty$ we find that

$$M\lambda(1) + \lambda(f) = \lambda(f + M) \leq \mu(f + M) = \mu(f) + M\mu(1)$$

and thus $\lambda(f) \leq \mu(f)$ for all $f \in C(X)$. Replacing $f$ by $-f$ implies that $\mu(f) \leq \lambda(f)$ and therefore that $\lambda(f) = \mu(f)$.  

The next theorem is the multi-dimensional extension of Theorem 10.62. The proof will follow the same pattern as its one dimensional cousin with a few complications due to the cumbersome nature of multi-dimensional cumulative distribution functions.

**Theorem 10.63 (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 $B_X$ such that

$$\lambda(f) = \mu(f) := \int_X f \, d\mu \text{ for all } f \in C(X).$$  

### Proof.

For $b \in [0,1]^d$ and $\varepsilon > 0$ let $\varphi_{b,\varepsilon}(x) := \prod_{i=1}^d \varphi_{b_i,\varepsilon}(x_i)$ so that $\varphi_{b,\varepsilon} | X \subseteq C(X)$ and $\varphi_{b,\varepsilon} \downarrow 1_{[0,b]}$ as $\varepsilon \downarrow 0$ which again shows $1_{[0,b]} \subseteq C(X)$. We will carry the proof in a number of steps.

1. We start by defining $F : [0,1]^d \to [0,\infty)$ by $F(b) := \lambda(1_{[0,b]})$ for all $b \in [0,1]^d$ which exists by Remark 10.60. Let $\mu_F$ be the unique finitely additive measure on $A([0,1]^d)$ such that

$$\mu_F((a,b]) = \sum_{\gamma \in \{1,\ldots,d\}} \varepsilon_\gamma (-1)^{\gamma} F(a_\gamma \times b_\gamma).$$  

2. Next we will show $F$ is right continuous and $\mu_F \geq 0$ and hence there exists a unique extension, $\mu$, of $\mu_F$ to $B_X$. We now prove these claims.

a) **Right continuity of $F$.** We must show if $b_n \in [0,1]^d$ with $a \leq b_n$ and $\lim_{n \to \infty} b_n = a$, then $\lim_{n \to \infty} F(b_n) = F(a)$. If we let $c_m := a + \frac{1}{m} 1$ for $m \in \mathbb{N}$, we will have $[0,b_n] \subseteq [0,c_m]$ for all sufficiently large $n$ and hence

$$\limsup_{n \to \infty} F(b_n) = \limsup_{n \to \infty} \lambda(1_{[0,b_n]}) \leq \lambda(1_{[0,c_m]}) = F(c_m).$$

Therefore

$$F(a) \leq \liminf_{n \to \infty} F(b_n) \leq \limsup_{n \to \infty} F(b_n) \leq \lim_{m \to \infty} F(c_m)$$

and so to finish the proof of right continuity it suffices to show

$$\lim_{n \to \infty} F(c_n) \leq F(a).$$

If we let $\delta_n > 0$ be as in Eq. (10.36), then

$$1_{[0,c_n]} \leq \varphi_{c_n,\delta_n} \downarrow 1_{[0,a]}$$

and therefore

$$F(c_n) \leq \lambda(\varphi_{c_n,\delta_n}) \downarrow \lambda(1_{[0,a]}) = F(a)$$

which shows $\lim_{n \to \infty} F(c_n) \leq F(a)$.

b) **$\mu_F$ is positive.** Let $a, b \in [0,1]^d$. From Eq. (5.43), we have

$$\sum_{|\gamma| \text{ even}} w_\gamma 1_{[0,a_\gamma \times b_\gamma]} \geq \sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0,a_\gamma \times b_\gamma]}$$

wherein we have used $\mu(\{a_i\}) = 0$ for all $1 \leq i < n$ in the second to last inequality and the dominated convergence theorem to take the limit. Thus we have shown that $\lambda(f) \leq \mu(f)$ for all $f \geq 0$. Since $\lambda(1) = \mu(1)$, if $f \in C(X)$ and $M = \|f\|_\infty$ we find that

$$M\lambda(1) + \lambda(f) = \lambda(f + M) \leq \mu(f + M) = \mu(f) + M\mu(1)$$

and thus $\lambda(f) \leq \mu(f)$ for all $f \in C(X)$. Replacing $f$ by $-f$ implies that $\mu(f) \leq \lambda(f)$ and therefore that $\lambda(f) = \mu(f)$.  

The next theorem is the multi-dimensional extension of Theorem 10.62. The proof will follow the same pattern as its one dimensional cousin with a few complications due to the cumbersome nature of multi-dimensional cumulative distribution functions.
and therefore using Theorem 10.59
\[
\sum_{|\gamma| \text{ even}} w_\gamma F(a_\gamma \times b_\gamma) = \lambda \left( \sum_{|\gamma| \text{ even}} w_\gamma \left[ 1_{[0,a_\gamma \times b_\gamma]} \right] \right) \\
\geq \lambda \left( \sum_{|\gamma| \text{ odd}} w_\gamma 1_{[0,a_\gamma \times b_\gamma]} \right) = \sum_{|\gamma| \text{ odd}} w_\gamma F(a_\gamma \times b_\gamma).
\]

The inequality combined with Eq. 10.38 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 \( \hat{a}_i = 0 \) and and \( \{a_i, b_i\} = [0, b_i] \) if \( a_i = 0 \) and \( \hat{a}_i \in (0, a_i) \) and \( \{a_i, b_i\} = (a_i, b_i) \) if \( a_i > 0 \). If we further let \( \xi_{\gamma} := \prod_{i=\gamma_1 \neq 0}^{\gamma_n} 1_{\hat{a}_i \times b_\gamma} \), then we find,
\[
1_{[a,b]}(x) = \prod_{i=1}^{d} 1_{[a_i, b_i]}(x_i) = \prod_{i=1}^{d} \left[ 1_{[0,b_i]}(x_i) - 1_{[0,a_i]}(x_i) \right] \\
\leq \prod_{i=1}^{d} \left[ 1_{[0,b_i]}(x_i) - 1_{[\hat{a}_i,b_i]}(x_i) \right] \\
= \sum_{\gamma \subset \{1,2,\ldots,d\}} \xi_{\gamma} (-1)^{|\gamma|} 1_{\hat{a}_\gamma \times b_\gamma}(x).
\]

This inequality may be rewritten as
\[
1_{[a,b]} + \sum_{|\gamma| \text{ odd}} \xi_{\gamma} 1_{\hat{a}_\gamma \times b_\gamma} \leq \sum_{|\gamma| \text{ even}} \xi_{\gamma} 1_{\hat{a}_\gamma \times b_\gamma}.
\]

Applying \( \lambda \) to the last inequality and solving the result for \( \lambda \left( 1_{[a,b]} \right) \) shows
\[
\lambda \left( 1_{[a,b]} \right) \leq \sum_{\gamma \subset \{1,2,\ldots,d\}} \xi_{\gamma} (-1)^{|\gamma|} \lambda \left( 1_{\hat{a}_\gamma \times b_\gamma} \right) \\
= \sum_{\gamma \subset \{1,2,\ldots,d\}} \xi_{\gamma} (-1)^{|\gamma|} F(\hat{a}_\gamma \times b_\gamma) = \mu_F(\{\hat{a}, b\})
\]

Letting \( \hat{a} \uparrow a \) in this inequality then implies \( \lambda \left( 1_{[a,b]} \right) \leq \mu_F([a,b]) \).

4. (Verification of Eq. 10.37.) Let \( \pi = \{0 = s_0 < s_1 < \cdots < s_n = 1\} \) be partition of \([0,1]\), for \( s = s_k \in \pi \backslash \{1\} \) let \( s_+ := s_{k+1} \) and let \( H \) be the partition of \([0,1]^n\) consisting of rectangles \( Q \) of the form \( \{a, a_+\} \)

where \( a = (a_1, \ldots, a_d) \in [\pi \backslash \{1\}[^n] \), \( a_+ = ((a_1)_+, \ldots, (a_d)_+) \), and \( \{a, a_+\} = \prod_{i=1}^{d} \{a_i, (a_i)_+\} \) with \( \{a_i, (a_i)_+\} = \{a_i, (a_i)_+\} \) if \( a_i > 0 \) and \( \{a_i, (a_i)_+\} = \{a_i, (a_i)_+\} \) if \( a_i = 0 \). Let us first assume that \( f \in C(X) \) is a non-negative function and for \( Q \in \Pi \) let \( c_Q := \max_{x \in Q} f(x) \) and define \( f^\pi \) and \( \hat{f}^\pi \) by
\[
f^\pi := \sum_{Q \in \Pi} c_Q 1_Q \quad \text{and} \quad \hat{f}^\pi := \sum_{Q \in \Pi} c_Q 1_{\hat{Q}}.
\]

Since \( 0 \leq f \leq f^\pi \leq \hat{f}^\pi \in C(X)_+ \), we conclude that
\[
0 \leq \lambda(f) \leq \lambda \left( \hat{f}^\pi \right) = \sum_{Q \in \Pi} c_Q \lambda(1_Q) \leq \sum_{Q \in \Pi} c_Q \mu_F(Q) \tag{10.39}
\]

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 \( \pi \backslash \{0,1\} \subset U^c \) in which case \( \mu_F(Q) = \mu_F(Q) \) for all \( Q \in \Pi \). Thus under previous restriction on \( \pi \) we the we inequality in Eq. (10.39) becomes
\[
0 \leq \lambda(f) \leq \sum_{Q \in \Pi} c_Q \mu_F(Q) = \mu(F^\pi).
\]

Since \( f^\pi \rightarrow f \) boundedly as \( |\pi| \rightarrow 0 \) with \( \pi \backslash \{0,1\} \subset U^c \) 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 \( \mu(1) = F(1) = \lambda(1) \), 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 \( \lambda(f) = \mu(f) \).

\[\blacksquare\]

**Theorem 10.64.** 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 \((0,1)^d, B_{(0,1)^d}\) such that \( \lambda = \mu \) for all \( f \in C_c \left( (0,1)^d \right) \).
Proof. Given $\varphi \in C_c\left(\left(0,1\right)^d,\left[0,1\right]\right)$, let $\lambda_\varphi(f) := \lambda(\varphi \cdot f)$ for all $f \in C_c\left(\left(0,1\right)^d\right)$. As $\lambda_\varphi$ is a positive linear functional on $C\left(\left(0,1\right)^d\right)$ we may apply Theorem 10.63 to find a unique measure $\mu_\varphi$ on $\left(\left[0,1\right]^d,B_{\left[0,1\right]^d}\right)$ such that $\lambda_\varphi(f) = \mu_\varphi(f)$ for all $f \in C\left(\left(0,1\right)^d\right)$. We now complete the proof in a number of steps.

1. If $\psi \in C_c\left(\left(0,1\right)^d,\left[0,1\right]\right)$ with $\varphi \leq \psi$ then $\mu_\varphi \leq \mu_\psi$. To see this is the case first observe that if $A$ is a closed subset of $\left[0,1\right]^d$ and $f_n \in C\left(\left(0,1\right)^d\right)$ are chosen so that $f_n \downarrow 1_F$, then

$$\mu_\varphi(F) = \lim_{n \to \infty} \lambda_\varphi(f_n) = \lim_{n \to \infty} \lambda(\varphi f_n) = \lambda(\varphi \cdot 1_F).$$

As $\varphi \cdot 1_F \leq \psi \cdot 1_F$ it follows

$$\mu_\varphi(F) = \lambda(\varphi \cdot 1_F) \leq \lambda(\psi \cdot 1_F) = \mu_\psi(F).$$

Then using the regularity properties of $\mu_\varphi$ and $\mu_\psi$ (see Exercise 6.7 and/or ??) if $A \in B_{\left[0,1\right]^d}$ we have,

$$\mu_\varphi(A) = \sup \{ \mu_\varphi(F) : F \subset A \& F \text{ closed} \} \leq \sup \{ \mu_\psi(F) : F \subset A \& F \text{ closed} \} = \mu_\psi(A).$$

2. If $A \in B_{\left[0,1\right]^d}$ 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. (10.40) it follows that $\mu_\varphi(F) = \mu_\psi(F)$ and in this case the inequality in Eq. (10.41) becomes an equality.

3. Let $Q_n := \left[\frac{1}{n},1 - \frac{1}{n}\right]^d$ for $n \geq 3$ and let $\varphi_n \in C_c\left(\left(0,1\right)^d,\left[0,1\right]\right)$ be chosen so that $1_{Q_n} \leq \varphi_n$ and supp$\left(\varphi_n\right) \subset Q_{n+1}$ and hence $\varphi_n \uparrow 1_{\left(0,1\right)^d}$. We then define the measure $\mu$ on $B_{\left(0,1\right)^d}$ by

$$\mu(A) = \uparrow \lim_{n \to \infty} \mu_{\varphi_n}(A) \forall A \in B_{\left(0,1\right)^d}.$$ 

4. An easy consequence of item 2. is that if $f$ is a bounded $B_{\left(0,1\right)^d}$ measurable function with compact support in $\left(0,1\right)^d$, then $\mu(f) = \mu_{\varphi_n}(f)$ for all $n$ such that supp$\left(f\right) \subset Q_n$. In particular if $f \in C_c\left(\left(0,1\right)^d\right)$ with supp$\left(f\right) \subset Q_n$, then

$$\mu(f) = \mu_{\varphi_n}(f) = \lambda_{\varphi_n}(f) = \lambda(\varphi_n f) = \lambda(f)$$

and the proof is complete.

\begin{corollary}
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 $\left(\mathbb{R}^d,B_{\mathbb{R}^d}\right)$ such that $\lambda(f) = \mu(f)$ for all $f \in C_c\left(\mathbb{R}^d\right)$.

\begin{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

$$\psi(t_1,\ldots,t_d) := (\psi(t_1),\ldots,\psi(t_d)).$$

Using this homeomorphic identification of $(0,1)^d$ with $\mathbb{R}^d$ we may easily translate the statement in Theorem 10.64 to the statement in the corollary.
\end{proof}
\end{corollary}
Thus it follows that

\[ M_X(t) = \frac{1}{b} \int_0^b e^{tx} \, dx = \frac{1}{bt} \left( e^{tb} - 1 \right) \]

\[ = \sum_{n=1}^{\infty} \frac{1}{n!} (bt)^{n-1} = \sum_{n=0}^{\infty} \frac{b^n}{(n+1)!} t^n. \]

On the other hand (see Proposition [10.33]),

\[ M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E X^n. \]

Thus it follows that

\[ E X^n = \frac{b^n}{n+1}. \]

Of course this may be calculated directly just as easily,

\[ E X^n = \frac{1}{b} \int_0^b x^n \, dx = \frac{1}{b(n+1)} \frac{x^{n+1}}{n+1} |_0^b = \frac{b^n}{n+1}. \]

**Definition 10.67.** A random variable \( T \geq 0 \) is said to be **exponential with parameter** \( \lambda \in [0, \infty) \) provided, \( P(T > t) = e^{-\lambda t} \) for all \( t \geq 0 \). 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 \( P(T \in (t, t+dt)) = \lambda_1 \geq 0 e^{-\lambda t} \, dt \). Applying Corollary [10.30] repeatedly implies,

\[ ET = \int_0^{\infty} \tau \lambda e^{-\lambda \tau} \, d\tau = \lambda \left( -\frac{d}{d\lambda} \right) \int_0^{\infty} e^{-\lambda \tau} \, d\tau = \lambda \left( -\frac{d}{d\lambda} \right) \lambda^{-1} = \lambda^{-1} \]

and more generally that

\[ ET^k = \int_0^{\infty} \tau^k \lambda e^{-\lambda \tau} \, d\tau = \lambda \left( -\frac{d}{d\lambda} \right)^k \int_0^{\infty} e^{-\lambda \tau} \, d\tau = \lambda \left( -\frac{d}{d\lambda} \right)^k \lambda^{-1} = k! \lambda^{-k}. \]

(10.42)

In particular we see that

\[ \text{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \]

(10.43)

Alternatively we may compute the moment generating function for \( T \),

\[ M_T(a) := E [e^{aT}] = \int_0^{\infty} e^{at} \lambda e^{-\lambda t} \, dt \]

\[ = \int_0^{\infty} e^{at} \lambda e^{-\lambda t} \, dt = \frac{\lambda}{\lambda - a} = \frac{1}{1 - a\lambda^{-1}} \]  

(10.44)

which is valid for \( a < \lambda \). On the other hand (see Proposition [10.33]), we know that

\[ E [e^{aT}] = \sum_{n=0}^{\infty} \frac{a^n}{n!} E [T^n] \text{ for } |a| < \lambda. \]

(10.45)

Comparing this with Eq. (10.44) again shows that Eq. (10.42) is valid.

Here is yet another way to understand and generalize Eq. (10.44). We simply make the change of variables, \( u = \lambda \tau \) in the integral in Eq. (10.42) to learn,

\[ E T^k = \lambda^{-k} \int_0^{\infty} u^k e^{-u} \, du = \lambda^{-k} \Gamma(k+1). \]

This last equation is valid for all \( k \in (-1, \infty) \) – in particular \( k \) need not be an integer.

**Theorem 10.68 (Memoryless property).** A random variable, \( T \in (0, \infty] \) has an exponential distribution if it satisfies the memoryless property:

\[ P(T > s + t | T > s) = P(T > t) \text{ for all } s, t \geq 0, \]

where as usual, \( P(A|B) := P(A \cap B) / P(B) \) when \( P(B) > 0 \). (Note that \( T \overset{d}{=} E(\lambda) \) means that \( P(T > t) = e^{-\lambda t} = 1 \) 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, \( e^{-\lambda} = g(1) = g(1/q) \). 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 \( T \sim \mathcal{E}(\lambda) \). \( \square \)

**Exercise 10.13 (Gamma Distributions).** Let \( X \) be a positive random variable. For \( k, \theta > 0 \), we say that \( X \sim \text{Gamma}(k, \theta) \) if

$$f(x; k, \theta) := x^{k-1} e^{-x/\theta} \theta^k / \Gamma(k), \text{ for } x > 0, \text{ and } k, \theta > 0.$$

Find the moment generating function (see Definition 10.32), \( M_X(t) = \mathbb{E}[e^{tx}] \) for \( t < \theta^{-1} \). Differentiate your result in \( t \) to show

$$\mathbb{E}[X^m] = k(k+1) \ldots (k+m-1) \theta^m \text{ for all } m \in \mathbb{N}_0.$$

In particular, \( \mathbb{E}[X] = k\theta \) and \( \text{Var}(X) = k\theta^2 \). (Notice that when \( k = 1 \) and \( \theta = \lambda^{-1} \), \( X \sim \mathcal{E}(\lambda) \).

**10.6.1 Normal (Gaussian) Random Variables**

**Definition 10.69 (Normal / Gaussian Random Variables).** A random variable, \( Y \), is normal with mean \( \mu \) standard deviation \( \sigma^2 \) iff

$$P(Y \in B) = \frac{1}{\sqrt{2\pi}\sigma^2} \int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \, dy \text{ for all } B \in \mathcal{B}(\mathbb{R}).$$

(10.46)

We will abbreviate this by writing \( Y \sim N(\mu, \sigma^2) \). When \( \mu = 0 \) and \( \sigma^2 = 1 \) we will simply write \( N \) for \( N(0,1) \) and if \( Y \sim N \), we will say \( Y \) is a standard normal random variable.

Observe that Eq. (10.46) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi}\sigma^2} \int \sigma \, e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \, dy$$

for all bounded measurable functions, \( f : \mathbb{R} \rightarrow \mathbb{R} \). Also observe that \( Y \sim N(\mu, \sigma^2) \) is equivalent to \( Y \sim \sigma N + \mu \). Indeed, by making the change of variable, \( y = \sigma x + \mu \), we find

$$\mathbb{E}[f(\sigma N + \mu)] = \frac{1}{\sqrt{2\pi}\sigma^2} \int f(\sigma x + \mu) e^{-\frac{1}{2\sigma^2}x^2} \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \, dy = \frac{1}{\sqrt{2\pi}\sigma^2} \int f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \, dy.$$

Lastly the constant, \((2\pi\sigma^2)^{-1/2}\) is chosen so that

$$\frac{1}{\sqrt{2\pi}\sigma^2} \int e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \, dy = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}y^2} \, dy = 1,$$

see Example 10.50 and Lemma 12.29.

**Exercise 10.14.** Suppose that \( X \sim N(0,1) \) and \( f : \mathbb{R} \rightarrow \mathbb{R} \) is a \( C^1 \) function such that \( Xf(X), f'(X) \) and \( f(X) \) are all integrable random variables. Show

$$\mathbb{E}[Xf(X)] = -\frac{1}{\sqrt{2\pi}} \int f(x) \frac{d}{dx} e^{-\frac{1}{2}x^2} \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int f'(x) e^{-\frac{1}{2}x^2} \, dx = \mathbb{E}[f'(X)].$$

**Example 10.70.** Suppose that \( X \sim N(0,1) \) and define \( \alpha_k := \mathbb{E}[X^{2k}] \) for all \( k \in \mathbb{N}_0 \). By Exercise 10.14.
Exercise 10.16. Suppose that $f \equiv 0$ to see that $f$ for all continuous functions,

and by a simple induction argument,

$$E[X^{2k}] = \alpha_k = (2k - 1)!!,$$ (10.47)

where $(-1)!! := 0$. Actually we can use the $\Gamma$ - function to say more. Namely for any $\beta > -1$,

$$E[|X|^\beta] = \frac{1}{\sqrt{\pi}} \int_0^\infty (2y)^{\beta/2} e^{-y} y^{-1/2} dy$$

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,

$$E[|X|^\beta] = \frac{1}{\sqrt{\pi}} \int_0^\infty (2y)^{\beta/2} e^{-y} y^{-1/2} dy$$

$$= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\beta+1/2)}{2^{\beta/2}}$$ (10.48)

Exercise 10.15. Suppose that $X \equiv N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := E[e^{i\lambda X}] = \exp\left(-\lambda^2/2\right).$$ (10.49)

Hint: Use Corollary 10.30 to show, $f'(\lambda) = iE[Xe^{i\lambda X}]$ and then use Exercise 10.14 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Exercise 10.16. Suppose that $X \equiv N(0,1)$ and $t \in \mathbb{R}$. Show $E[e^{tX}] = \exp(t^2/2)$ . (You could follow the hint in Exercise 10.15 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 10.17. Use Exercise 10.16 and Proposition 10.33 to give another proof that $E[X^{2k}] = (2k - 1)!!$ when $X \equiv N(0,1)$.

Exercise 10.18. Let $X \equiv N(0,1)$ and $\alpha \in \mathbb{R}$, find $\rho : \mathbb{R}_+ \to \mathbb{R}_+ := (0,\infty)$ such that

$$E[f(|X|^\alpha)] = \int_{\mathbb{R}_+} f(x) \rho(x) dx$$

for all continuous functions, $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support in $\mathbb{R}_+$.

Lemma 10.71 (Gaussian tail estimates). Suppose that $X$ is a standard normal random variable, i.e.

$$P(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx$$ for all $A \in \mathcal{B}_{\mathbb{R}},$

then for all $x \geq 0$,

$$P(X \geq x) \leq \min \left( \frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) \leq \frac{1}{2} e^{-x^2/2}. \quad (10.50)$$

Moreover (see [35, Lemma 2.5]),

$$P(X \geq x) \geq \max \left( \frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) \quad (10.51)$$

which combined with Eq. (10.50) proves Mill’s ratio (see [18]):

$$\lim_{x \to \infty} \frac{P(X \geq x)}{x e^{-x^2/2}} = 1. \quad (10.52)$$

Proof. See Figure 10.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min \left( \frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right),$$

the red is the plot of $1/2 e^{-x^2/2}$, and the blue is the plot of

$$\max \left( \frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) .$$

The formal proof of these estimates for the reader who is not convinced by Figure 10.1 is given below.

We begin by observing that

$$P(X \geq x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2} dy$$

$$\leq -\frac{1}{\sqrt{2\pi}} \frac{1}{x^2} e^{-y^2/2} \big|_x^\infty = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \quad (10.53)$$

If we only want to prove Mill’s ratio (10.52), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,
This equation along with Eq. (10.53) gives the first equality in Eq. (10.50). To for $x$ from which it follows, $P(x) = 1$ we must show, $P(x) ≥ 1$ for all $0 ≤ x < ∞$. So to finish the proof of Eq. (10.51) we must show,

$$f(x) := \frac{1}{\sqrt{2\pi}} xe^{-x^2/2} - (1 + x^2) P(X ≥ x)$$

$$= \frac{1}{\sqrt{2\pi}} \left[ xe^{-x^2/2} - (1 + x^2) \int_x^∞ e^{-y^2/2}dy \right] ≥ 0 \text{ for all } 0 ≤ x < ∞.$$

This follows by observing that $f(0) = -1/2 < 0$, $\lim_{x \to ∞} f(x) = 0$ and

$$f'(x) = \frac{1}{\sqrt{2\pi}} \left[ e^{-x^2/2} (1 - x^2) - 2x P(X ≥ x) + (1 + x^2) e^{-x^2/2} \right]$$

$$= 2 \left( \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - x P(X ≥ y) \right) ≥ 0,$$

where the last inequality is a consequence Eq. (10.50).

10.7 Stirling’s Formula

On occasion one is faced with estimating an integral of the form, $\int_J e^{-G(t)}dt$, where $J = (a,b) ⊂ \mathbb{R}$ and $G(t)$ is a $C^1$ – function with a unique (for simplicity) global minimum at some point $t_0 ∈ J$. The idea is that the majority contribution of the integral will often come from some neighborhood, $(t_0 - α, t_0 + α)$, of $t_0$. Moreover, it may happen that $G(t)$ can be well approximated on this neighborhood by its Taylor expansion to order 2;

$$G(t) ∼ G(t_0) + \frac{1}{2} \dot{G}(t_0) (t - t_0)^2.$$  

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 $G(t_0) ≠ 0$ and hence $\ddot{G}(t_0) > 0$. Under these hypothesis we will have,
Making the change of variables, \( s = \sqrt{\tilde{G}(t_0)} (t - t_0) \), in the above integral then gives,
\[
\int e^{-G(t)} dt \cong e^{-G(t_0)} \int_{|s|<\sqrt{\tilde{G}(t_0)} \cdot \alpha} e^{-\frac{1}{2} s^2} ds
\]
\[
= \frac{1}{\sqrt{\tilde{G}(t_0)}} e^{-G(t_0)} \left[ \sqrt{2\pi} - \int_{\sqrt{\tilde{G}(t_0)} \cdot \alpha}^{\infty} e^{-\frac{1}{2} s^2} ds \right]
\]
\[
= \frac{1}{\sqrt{\tilde{G}(t_0)}} e^{-G(t_0)} \left[ \sqrt{2\pi} - O \left( \frac{1}{\sqrt{\tilde{G}(t_0)} \cdot \alpha} \right) \right].
\]

If \( \alpha \) is sufficiently large, for example if \( \sqrt{\tilde{G}(t_0)} \cdot \alpha = 3 \), then the error term is about 0.0037 and we should be able to conclude that
\[
\int e^{-G(t)} dt \cong \frac{2\pi}{\sqrt{\tilde{G}(t_0)}} e^{-G(t_0)}. \quad (10.55)
\]

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 10.72 (Stirling’s formula).** The Gamma function (see Definition 10.49), satisfies Stirling’s formula,
\[
\lim_{x \to \infty} \frac{\Gamma(x+1)}{\sqrt{2\pi} e^{-x} x^{x+1/2}} = 1. \quad (10.56)
\]

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 10.77 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 \( \Gamma(x+1) \);
\[
\Gamma(x+1) = \int_0^\infty e^{-t} t^x dt = \int_0^\infty e^{-G_x(t)} dt,
\]
where
\[
G_x(t) := t - x \ln t.
\]

Then \( \dot{G}_x(t) = 1 - 1/t \), \( \ddot{G}_x(t) = x/t^2 \), \( \dot{G}_x \) has a global minimum (since \( \ddot{G}_x > 0 \)) at \( t_0 = x \) where \( G_x(x) = x - x \ln x \) and \( \dot{G}_x(x) = 1/x \).

So if Eq. (10.55) is valid in this case we should expect,
\[
\Gamma(x+1) \cong \sqrt{2\pi} e^{-x} x^{x+1/2}
\]
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 \( s = \sqrt{\tilde{G}(t_0)} (t - t_0) = \frac{1}{\sqrt{x}} (t - x) \) as suggested above. Then
\[
G_x(t) - G_x(x) = (t - x) - x \ln (t/x) = \sqrt{x} s - x \ln \left( \frac{x + \sqrt{x} s}{x} \right)
\]
\[
= x \left[ \frac{s}{\sqrt{x}} - \ln \left( 1 + \frac{s}{\sqrt{x}} \right) \right] = s^2 q \left( \frac{s}{\sqrt{x}} \right)
\]
where
\[
q(u) := \frac{1}{u^2} [u - \ln (1 + u)] \quad \text{for } u > -1 \quad \text{with } q(0) := \frac{1}{2}
\]
Setting \( q(0) = 1/2 \) makes \( q \) a continuous and in fact smooth function on \((-1, \infty) \), see Figure 10.2. Using the power series expansion for \( \ln (1 + u) \) we find,
\[
q(u) = \frac{1}{2} + \sum_{k=3}^{\infty} \frac{(-u)^{k-2}}{k} \quad \text{for } |u| < 1.
\]

Making the change of variables, \( t = x + \sqrt{x} s \) in the second integral in Eq. (10.57) 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
\[
I(x) = \int_{-\sqrt{x}}^{\infty} e^{-q \left( \frac{s}{\sqrt{x}} \right)} s^2 ds = \int_{-\infty}^{\infty} 1_{s \geq -\sqrt{x}} \cdot e^{-q \left( \frac{s}{\sqrt{x}} \right)} s^2 ds.
\]

From Eq. (10.58) it follows that \( \lim_{u \to 0} q(u) = 1/2 \) and therefore,
\[
\int_{-\infty}^{\infty} \lim_{x \to \infty} 1_{s \geq -\sqrt{x}} \cdot e^{-q \left( \frac{s}{\sqrt{x}} \right)} s^2 ds = \int_{-\infty}^{\infty} e^{-\frac{1}{2} s^2} ds = \sqrt{2\pi}.
\]
We will sometimes use the following variant of Eq. (10.56):

\[
\lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi x} \left(\frac{x}{e}\right)^x} = 1
\]  

(10.61)

To prove this let \( x \to x - 1 \) in Eq. (10.56) in order to find,

\[
1 = \lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi x} e^{-x} \cdot (x-1)^{x-1/2}} = \lim_{x \to \infty} \frac{\Gamma(x)}{\sqrt{2\pi x} \left(\frac{x}{e}\right)^x} \cdot e \cdot (1 - \frac{1}{x})^x
\]

which gives Eq. (10.61) since

\[
\lim_{x \to \infty} \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^x = 1.
\]

**Remark 10.73 (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, 

\[
\hat{h}(1) - \hat{h}(0) = \int_0^1 \dot{h}(t) \, dt = - \int_0^1 \hat{h}(t) \, d(1-t)
\]

\[
= -\dot{h}(t) (1-t) \big|_0^1 + \int_0^1 \ddot{h}(t) \, dt
\]

\[
= \hat{h}(0) + \frac{1}{2} \int_0^1 \ddot{h}(t) \, dv(t)
\]

(10.62)

where \( dv(t) := 2(1-t) \, dt \) 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,

\[
F(b) - F(a) = (b - a) \hat{F}(a) + \frac{1}{2} (b - a)^2 \int_0^1 \ddot{F}(a + t(b - a)) \, dv(t). 
\]

(10.63)

(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,

\[
u - \ln(1 + u) = \frac{1}{2} u^2 \int_0^1 \frac{1}{(1 + tu)^2} \, dv(t),
\]

i.e.

\[
q(u) = \frac{1}{2} \int_0^1 \frac{1}{(1 + tu)^2} \, dv(t).
\]
From this expression for \( q(u) \) it now easily follows that
\[
q(u) \geq \frac{1}{2} \int_0^1 \frac{1}{(1+0)^2} d\nu(t) = \frac{1}{2} \quad \text{if} \quad -1 < u \leq 0
\]
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 \( \varepsilon(M) = (1 + M)^{-2}/2 \).

### 10.7.1 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.
\[
P_x(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k} \quad \text{for} \quad 0 \leq k \leq n. \tag{10.64}
\]
Recall that \( ES_n = nx \) and \( \text{Var}(S_n) = n\sigma^2 \) where \( \sigma^2 := x(1-x) \). The weak law of large numbers states (Exercise 5.18) that
\[
\text{law of large numbers states (Exercise 5.18)}.
\]
(The next central limit theorem describes the fluctuations of \( S_n \).

**Theorem 10.74 (De Moivre-Laplace Central Limit Theorem).** For all \( -\infty < a < b < \infty \),
\[
\lim_{n \to \infty} P\left( a \leq \frac{S_n - nx}{\sigma \sqrt{n}} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy = P(a \leq N \leq b)
\]
where \( N \overset{d}{=} N(0,1) \). Informally, \( \frac{S_n - nx}{\sigma \sqrt{n}} \overset{d}{=} N \) or equivalently, \( S_n \overset{d}{=} nx + \sigma \sqrt{n} \cdot N \) 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,
\[
p_n := P\left( a \leq \frac{S_n - nx}{\sigma \sqrt{n}} \leq b \right) = P\left( S_n \in nx + \sigma \sqrt{n} [a,b] \right)
\]
\[
= \sum_{k \in nx + \sigma \sqrt{n} [a,b]} P(S_n = k)
\]
\[
= \sum_{k \in nx + \sigma \sqrt{n} [a,b]} \binom{n}{k} x^k (1-x)^{n-k}.
\]
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
\[
p_n = \sum_{y_k \in [a,b]} \sigma \sqrt{n} \binom{n}{k} x^k (1-x)^{n-k} \Delta y_k. \tag{10.65}
\]
So to finish the proof we need to show, for \( k = O(\sqrt{n}) \) \( y_k = O(1) \), that
\[
\sigma \sqrt{n} \binom{n}{k} x^k (1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \text{ as } n \to \infty \tag{10.66}
\]
in which case the sum in Eq. (10.65) 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,
\[
\sigma \sqrt{n} \binom{n}{k} = \sigma \sqrt{n} \frac{1}{k!} \frac{n!}{(n-k)!} \sim \frac{\sigma \sqrt{n}}{\sqrt{2\pi}} \frac{n^{k+1/2}}{k^{k+1/2} (n-k)^{n-k+1/2}}
\]=
\[
= \frac{\sigma}{\sqrt{2\pi}} \frac{1}{k^{k+1/2} (n-k)^{n-k+1/2}}.
\]
\[
= \frac{\sigma}{\sqrt{2\pi}} \frac{1}{\sqrt{x(1-x)} (x + \sigma \sqrt{n} y_k)^{k+1/2} (1-x - \sigma \sqrt{n} y_k)^{n-k+1/2}}
\]=
\[
= \frac{\sigma}{\sqrt{2\pi}} \frac{1}{\sqrt{x(1-x)} (x + \sigma \sqrt{n} y_k)^{k} (1-x - \sigma \sqrt{n} y_k)^{n-k}}
\]
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,
\[
\sqrt{2\pi} \sigma \sqrt{n} \binom{n}{k} x^k (1 - x)^{n-k} \sim \frac{x^k (1 - x)^{n-k}}{(x + z_k)^2 (1 - x - z_k)^{n-k}}
\]

\[
= \frac{1}{(1 + \frac{1}{x} z_k)^n (1 - \frac{1}{1-x} z_k)^{n-k}}
\]

\[
= \frac{1}{(1 + \frac{1}{x} z_k)^{n(x+z_k)} (1 - \frac{1}{1-x} z_k)^{n(1-x-z_k)}} =: q(n,k).
\]

Taking logarithms and using Taylor’s theorem we learn

\[
n (x + z_k) \ln \left(1 + \frac{1}{x} z_k\right)
\]

\[
= n (x + z_k) \left(\frac{1}{x} z_k - \frac{1}{2x^2} z_k^2 + O \left(n^{-3/2}\right)\right)
\]

\[
= nz_k + \frac{n}{2x} z_k^2 + O \left(n^{-3/2}\right)
\]

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. (10.67) 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. (10.66).

The previous central limit theorem has shown that

\[
\frac{S_n}{n} \overset{d}{=} x + \frac{\sigma}{\sqrt{n}} N
\]

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 10.75 (Binomial Large Deviation Bounds).** Let us continue to use the notation in Theorem 10.74. Then for all \(y \in (0, x)\),

\[
\lim_{n \to \infty} \frac{1}{n} \ln P_x \left(\frac{S_n}{n} \leq y\right) = y \ln \frac{y}{x} + (1 - y) \ln \frac{1 - x}{1 - y},
\]

Roughly speaking,

\[
P_x \left(\frac{S_n}{n} \leq y\right) \approx e^{-n I_x(y)}
\]

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 10.3 for the graph of \(I_{1/2}\).

![Fig. 10.3. A plot of the rate function, \(I_{1/2}\).](image)

**Proof.** By definition of the binomial distribution,

\[
P_x \left(\frac{S_n}{n} \leq y\right) = P_x \left(S_n \leq ny\right) = \sum_{k \leq ny} \binom{n}{k} x^k (1 - x)^{n-k}.
\]
If \( a_k \geq 0 \), then we have the following crude estimates on \( \sum_{k=0}^{m-1} a_k \),

\[
\max_{k \leq m} a_k \leq \sum_{k=0}^{m-1} a_k \leq m \cdot \max_{k \leq m} a_k.
\tag{10.68}
\]

In order to apply this with \( a_k = \binom{n}{k} x^k (1-x)^{n-k} \) and \( m = \lfloor ny \rfloor \), 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,

\[
\frac{a_{k+1}}{a_k} = \frac{\binom{n}{k+1} x^{k+1} (1-x)^{n-k-1}}{\binom{n}{k} x^k (1-x)^{n-k}} = \frac{k!(n-k)! \cdot x}{(k+1)! \cdot (n-k)! \cdot (1-x)} = \frac{(n-k) \cdot x}{(k+1) \cdot (1-x)}.
\]

Therefore, where the latter expression is greater than or equal to 1 iff

\[
\frac{a_{k+1}}{a_k} \geq 1 \iff (n-k) \cdot x \geq (k+1) \cdot (1-x) \iff nx \geq k+1-x \iff k < (n-1)x-1.
\]

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. (10.68) implies,

\[
\left( \sum_{[ny]} x^{[ny]} (1-x)^{n-[ny]} \right) \leq P_x \left[ \frac{S_n}{n} \leq y \right] \leq [ny] \left( \sum_{[ny]} x^{[ny]} (1-x)^{n-[ny]} \right)
\]

or equivalently,

\[
\frac{1}{n} \ln \left[ \left( \sum_{[ny]} x^{[ny]} (1-x)^{n-[ny]} \right) \right] \leq \frac{1}{n} \ln P_x \left[ \frac{S_n}{n} \leq y \right] \leq \frac{1}{n} \ln \left( [ny] \sum_{[ny]} x^{[ny]} (1-x)^{n-[ny]} \right).
\]

By Stirling’s formula, for \( k \) such that \( k \) and \( n-k \) is large we have,

\[
\binom{n}{k} \sim \frac{1}{\sqrt{2\pi}} k^{k+1/2} (n-k)^{n-k+1/2} = \frac{1}{\sqrt{2\pi}} \frac{(k^{1/2})^{k+1/2}}{(n-k)^{1/2} \cdot (1-k)^{n-k+1/2}}
\]

and therefore,

\[
\frac{1}{n} \ln \left( \binom{n}{k} \right) \sim -\frac{1}{n} \ln \left( \frac{k}{n} \right) - \left( 1 - \frac{k}{n} \right) \ln \left( 1 - \frac{k}{n} \right).
\]

So taking \( k = \lfloor ny \rfloor \), we learn that

\[
\lim_{n \to \infty} \frac{1}{n} \ln \left( \binom{n}{\lfloor ny \rfloor} \right) = -y \ln y - (1-y) \ln (1-y)
\]

and therefore,

\[
\lim_{n \to \infty} \frac{1}{n} \ln \left( S_n / n \right) = -y \ln y - (1-y) \ln (1-y) + y \ln x + (1-y) \ln (1-x)
\]

As a consistency check it is worth noting, by Jensen’s inequality described below, that

\[
-I_x (y) = y \ln \frac{x}{y} + (1-y) \ln \left( \frac{1-x}{1-y} \right) \leq \ln \left( \frac{x}{y} + (1-y) \frac{1-x}{1-y} \right) = \ln (1) = 0.
\]

This must be the case since

\[
-I_x (y) = \lim_{n \to \infty} \frac{1}{n} \ln P_x \left( \frac{S_n}{n} \leq y \right) \leq \lim_{n \to \infty} \frac{1}{n} \ln 1 = 0.
\]

### 10.7.2 A primitive Stirling type approximation

**Theorem 10.76.** 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

\[
s_n - \frac{1}{2} \left( f(n) + f(1) \right) \leq \int_{1}^{n} f(x) \, dx \leq s_n - \frac{1}{2} [f(n+1) + 2f(1)] + \frac{1}{2} f(2)
\]

\[
\leq s_n - \frac{1}{2} [f(n) + 2f(1)] + \frac{1}{2} f(2).
\]
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 it follows that
\[
\int_{1}^{n} f(x) \, dx \leq \sum_{k=2}^{n} f(k) - \frac{1}{2} (f(n+1) - f(2)) = s_n - \frac{1}{2} [f(n+1) + 2f(1)] + \frac{1}{2} f(2) \leq s_n - \frac{1}{2} [f(n) + 2f(1)] + \frac{1}{2} f(2).
\]

Example 10.77 (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 conclude 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,
\[
\frac{e}{\sqrt{2}} \cdot e^{-n \frac{n+1}{2}} \leq n! \leq e \cdot e^{-n \frac{n+1}{2}}
\]
which compares well with Strirling’s formula (Theorem 10.72) which states,
\[
n! \sim \sqrt{2\pi} e^{-n \frac{n+1}{2}}.
\]
Observe that
\[
\frac{e}{\sqrt{2}} \approx 1.9221 \leq \sqrt{2\pi} \approx 2.506 \leq e \approx 2.7183.
\]

10.8 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
of \([a, b]\) let

\[
\text{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1, \ldots, n\},
\]

\[
M_j = \sup \{f(x) : t_j \leq x \leq t_{j-1}\}, \quad m_j = \inf \{f(x) : t_j \leq x \leq t_{j-1}\}
\]

\[
G_\pi = f(a)1_{\{a\}} + \sum_{j=1}^{n} M_j 1_{(t_{j-1}, t_j]}, \quad g_\pi = f(a)1_{\{a\}} + \sum_{j=1}^{n} m_j 1_{(t_{j-1}, t_j]}
\]

and which case the Riemann integral

\[
\int_a^b f(x) \, dx = \inf \pi S_\pi f \quad \text{and} \quad \int_a^b f(x) \, dx = \sup \pi s_\pi f.
\]

**Definition 10.78.** The function \(f\) is **Riemann integrable** if \(\int_a^b f = \int_a^b f \in \mathbb{R}\) and which case the Riemann integral \(\int_a^b f\) is defined to be the common value:

\[
\int_a^b f(x) \, dx = \int_a^b f(x) \, dx = \int_a^b f(x) \, dx.
\]

The proof of the following Lemma is left to the reader as Exercise 10.29.

**Lemma 10.79.** If \(\pi'\) and \(\pi\) are two partitions of \([a, b]\) and \(\pi \subset \pi'\) then

\[
G_\pi \geq G_{\pi'} \geq f \geq g_{\pi'} \geq g_\pi
\]

and

\[
S_\pi f \geq S_{\pi'} f \geq s_{\pi'} f \geq s_\pi f.
\]

There exists an increasing sequence of partitions \(\{\pi_k\}_{k=1}^\infty\) such that \(\text{mesh}(\pi_k) \downarrow 0\) and

\[
S_{\pi_k} f \downarrow \int_a^b f \quad \text{and} \quad s_{\pi_k} f \uparrow \int_a^b f \quad \text{as} \quad k \to \infty.
\]

If we let

\[
G := \lim_{k \to \infty} G_{\pi_k} \quad \text{and} \quad g := \lim_{k \to \infty} g_{\pi_k}
\]

then by the dominated convergence theorem,

\[
\int_{[a,b]} g dm = \lim_{k \to \infty} \int_{[a,b]} g_{\pi_k} = \lim_{k \to \infty} s_{\pi_k} f = \int_a^b f(x) \, dx
\]

and

\[
\int_{[a,b]} G dm = \lim_{k \to \infty} \int_{[a,b]} G_{\pi_k} = \lim_{k \to \infty} S_{\pi_k} f = \int_a^b f(x) \, dx.
\]

**Notation 10.80.** For \(x \in [a, b]\), let

\[
H(x) = \limsup_{y \to x} \{f(y) : |y - x| \leq \varepsilon, \ y \in [a, b]\}
\]

and

\[
h(x) = \liminf_{y \to x} \{f(y) : |y - x| \leq \varepsilon, \ y \in [a, b]\}.
\]

**Lemma 10.81.** The functions \(H, h : [a, b] \to \mathbb{R}\) satisfy:

1. \(h(x) \leq f(x) \leq H(x)\) 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 an increasing sequence of partitions such that \(\text{mesh}(\pi_k) \downarrow 0\) and \(G\) and \(g\) are defined as in Eq. (10.70), then

\[
G(x) = H(x) \geq f(x) \geq h(x) = g(x) \quad \forall \ x \notin \pi := \cup_{k=1}^\infty \pi_k.
\]

(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 \(h(x) \leq f(x) \leq H(x)\) 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 \(x \notin \pi\),

\[
G_k(x) \geq H(x) \geq f(x) \geq g_k(x) \quad \forall \ k
\]

and letting \(k \to \infty\) in this equation implies

\[
G(x) \geq H(x) \geq f(x) \geq h(x) = g(x) \quad \forall \ x \notin \pi.
\]

Moreover, given \(\varepsilon > 0\) and \(x \notin \pi\),

\[
\sup \{f(y) : |y - x| \leq \varepsilon, \ y \in [a, b]\} \geq G_k(x)
\]

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

\[
\sup \{f(y) : |y - x| \leq \varepsilon\} \geq H(x)
\]

for \(x \notin \pi\).
for all $x \notin \pi$. Combining this equation with Eq. (10.74) 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. (10.73) 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 10.82.** Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then

$$
\int_a^b f = \int_{[a,b]} \mathcal{H}dm\quad\text{and}\quad\int_a^b f = \int_{[a,b]} hdm
$$

(10.75)

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 $\overline{m}$-null set.
3. $f$ is Riemann integrable.

If $f$ is Riemann integrable then $f$ is Lebesgue measurable\(^4\) 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 $\overline{m}$ denote the completion of $m$, then

$$
\int_{[a,b]} \mathcal{H}dm = \int_a^b f(x)dx = \int_{[a,b]} f \overline{m} = \int_{[a,b]} hdm.
$$

(10.76)

**Proof.** Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 10.79 and let $G$ and $g$ be defined as in Lemma 10.83. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (10.75) is a consequence of Eqs. (10.71) and (10.72). From Eq. (10.75), $f$ is Riemann integrable iff

$$
\int_{[a,b]} \mathcal{H}dm = \int_{[a,b]} hdm
$$

and because $h \leq f \leq H$ this happens iff $h(x) = H(x)$ for $m$-a.e. $x$. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to $E$ being a $m$-null set. In light of these results and Eq. (10.73), the remaining assertions including Eq. (10.76) are now consequences of Lemma 10.85.

**Notation 10.83** In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_{[a,b]} f \overline{m}$.

\(^4\) Recall this means that if $N \subset X$ is a set such that $N \subset A \subset \mathcal{M} \text{ and } \mu(A) = 0$, then $N \notin \mathcal{M}$ as well.

10.9 Measurability on Complete Measure Spaces\(^*\)

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

**Proposition 10.84.** Suppose that $(X, \mathcal{B}, \mu)$ is a complete measure space\(^*\) 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 $E = \{x : f(x) \neq g(x)\}$ 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,

$$
(1_{E}g)^{-1}(A) = \begin{cases}
E^c \cup (1_{E}g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\
(1_{E}g)^{-1}(A) & \text{if } 0 \notin A
\end{cases}
$$

(10.77)

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. (10.77) shows that $1_{E}g$ is measurable. 2. Let $E = \{x : \lim_{n \to \infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_{E}f = \lim_{n \to \infty} 1_{E}f_n$, $g$ 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 $X = \{0, 1, 2\}$, $\mathcal{B} = \{\{0\}, \{1, 2\}, X, \varnothing\}$ 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 10.85.** Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\overline{\mathcal{M}}$ is the completion of $\mathcal{M}$ relative to $\mu$ and $\mu$ is the extension of $\mu$ to $\overline{\mathcal{M}}$. Then a function $f : X \to \mathbb{R}$ is $(\mathcal{M}, \mathcal{B} = \mathcal{B}_\mathbb{R})$ - measurable iff there exists a function $g : X \to \mathbb{R}$ that is $(\mathcal{M}, \mathcal{B})$ - measurable such $E = \{x : f(x) \neq g(x)\} \in \overline{\mathcal{M}}$ and $\mu(E) = 0$, i.e. $f(x) = g(x)$ for $\mu$ - a.e. $x$. Moreover for such a pair $f$ and $g$, $f, g \in L^1(\mu)$ iff $g \in L^1(\mu)$ and in which case

$$
\int_X f d\mu = \int_X g d\mu.
$$

**Proof.** Suppose first that such a function $g$ exists so that $\mu(E) = 0$. Since $g$ is also $(\overline{\mathcal{M}}, \mathcal{B})$ - measurable, we see from Proposition 10.84 that $f$ is $(\overline{\mathcal{M}}, \mathcal{B})$ - measurable. Conversely if $f$ is $(\overline{\mathcal{M}}, \mathcal{B})$ - measurable, by considering $f_\pm$, we may
assume that \( f \geq 0 \). Choose \((\mathcal{M}, \mathcal{B})\) – 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 \( A_k \in \mathcal{M} \), we may choose \( B_k \in \mathcal{M} \) such that \( B_k \subset A_k \) and \( \check{\mu}(A_k \setminus B_k) = 0 \). Letting
\[
\check{\varphi}_n := \sum a_k 1_{B_k}
\]
we have produced a \((\mathcal{M}, \mathcal{B})\) – measurable simple function \( \check{\varphi}_n \geq 0 \) such that \( E_n := \{ \varphi_n \neq \check{\varphi}_n \} \) has zero \( \check{\mu} \) – measure. Since \( \check{\mu}(\cap_n E_n) \leq \sum \check{\mu}(E_n) \), there exists \( F \in \mathcal{M} \) such that \( \cap_n E_n \subset F \) and \( \mu(F) = 0 \). It now follows that
\[
1_F \cdot \check{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.
\]
This shows that \( g = 1_F f \) is \((\mathcal{M}, \mathcal{B})\) – measurable and that \( \{ f \neq g \} \subset F \) has \( \check{\mu} \) – measure zero. Since \( f = g, \check{\mu} \) – a.e., \( \int_X f d\check{\mu} = \int_X g d\check{\mu} \) so to prove Eq. (10.78) it suffices to prove
\[
\int_X g d\check{\mu} = \int_X g d\mu.
\]
(10.78)
Because \( \check{\mu} = \mu \) on \( \mathcal{M} \), Eq. (10.78) is easily verified for non-negative \( \mathcal{M} \) – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 9.39 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 \((\text{Re } g)_\pm\) and \((\text{Im } g)_\pm\).

10.10 More Exercises

Exercise 10.19. 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 10.20 (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:
\[
\rho(A, C) \leq \rho(A, B) + \rho(B, C)
\]
for all \( A, B, C \in \mathcal{M} \).

2. Define \( A \sim B \) iff \( \rho(A \triangle B) = 0 \) 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 \( \check{\rho}([A], [B]) := \rho(A, B) \) is gives a well defined metric on \( \mathcal{M}/\sim \).

Exercise 10.21. 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 10.22. Now suppose that \( A \) is some index set and for each \( \lambda \in A \), \( \mu_\lambda : \mathcal{M} \to [0, \infty] \) is a measure on \( \mathcal{M} \). Define \( \mu : \mathcal{M} \to [0, \infty] \) by \( \mu(A) = \sum_{\lambda \in A} \mu_\lambda(A) \) for each \( A \in \mathcal{M} \). Show that \( \mu \) is also a measure.

Exercise 10.23. Let \((X, \mathcal{M}, \mu)\) be a measure space and \( \{ A_n \}_{n=1}^\infty \subset \mathcal{M} \), show
\[
\mu(\{ A_n \text{ a.a.} \}) \leq \limsup_{n \to \infty} \mu(A_n)
\]
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 10.24 (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 10.25. 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 10.26. Suppose \( \{ a_n \}_{n=-\infty}^\infty \subset \mathbb{C} \) is a summable sequence (i.e. \( \sum_{n=-\infty}^{\infty} |a_n| < \infty \)), then \( f(\theta) := \sum_{n=-\infty}^{\infty} a_n e^{i n \theta} \) is a continuous function for \( \theta \in \mathbb{R} \) and
\[
a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-i n \theta} d\theta.
\]

Exercise 10.27. 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 10.28. 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

$$(1 - z)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n - 1)!!}{2^n n!} z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} z^n$$

for $|z| < 1$.

Exercise 10.29. Prove Lemma 10.79.
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.

11.1 Multiplicative System Theorems

**Notation 11.1** 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 $|f_n(\omega)| \leq M$ 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 11.2 (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 a $\sigma$ – 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 bounded simple functions. Since every bounded $\mathcal{B}$ – measurable function may be written as a bounded limit of such simple functions (see Theorem 9.39), it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.

**Step 4.** ($\sigma(\mathbb{M}) \subset \mathcal{B}$.) Let $\varphi_n(x) = 0 \lor [(nx) \land 1]$ (see Figure 11.1 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 5.69 we may find polynomial functions, $p_l(x)$ such that $p_l \rightarrow \varphi_n$ 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, $p_l \circ (f-a) \rightarrow F_n$ uniformly as $l \rightarrow \infty$, from which 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. 11.1. Plots of $\varphi_1$, $\varphi_2$ and $\varphi_3$.](image)
Second proof.* (This proof may safely be skipped.) This proof will make use of Dynkin’s \( \pi - \lambda \) Theorem \([6.54]\). Let
\[
\mathcal{L} := \{ A \subseteq \Omega : 1_A \in \mathbb{H} \}.
\]
We then have \( \Omega \in \mathcal{L} \) since \( 1_{\Omega} = 1 \in \mathbb{H} \), if \( A, B \in \mathcal{L} \) with \( A \subseteq B \) 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 \( \pi \) system.

Let \( \varphi_n (x) = 0 \vee [(nx) \wedge 1] \) (see Figure 11.1 above) so that \( \varphi_n (x) \uparrow 1_{x>0} \).

Given \( f_1, f_2, \ldots, f_k \in \mathcal{M} \) and \( a_1, \ldots, a_k \in \mathbb{R} \), let
\[
F_n := \prod_{i=1}^k \varphi_n (f_i - a_i)
\]
and let
\[
M := \sup_{i=1,\ldots,k} \sup_{\omega} |f_i (\omega) - a_i|.
\]
By the Weierstrass approximation Theorem \([6.69]\) we may find polynomial functions, \( p_i (x) \) such that \( p_i \rightarrow \varphi_n \) uniformly on \( [-M, M] \). Since \( p_i \) is a polynomial it is easily seen that \( \prod_{i=1}^k p_i \circ (f_i - a_i) \in \mathbb{H} \). Moreover,
\[
\prod_{i=1}^k p_i \circ (f_i - a_i) \rightarrow F_n \text{ uniformly as } t \rightarrow \infty,
\]
from which 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 \( \cap_{i=1}^k \{f_i > a_i\} \in \mathbb{H} \) or equivalently that \( \cap_{i=1}^k \{f_i > a_i\} \in \mathcal{L} \). Therefore \( \mathcal{L} \) contains the \( \pi - \lambda \) system, \( \mathcal{P} \), consisting of finite intersections of sets of the form, \( \{f > a\} \) with \( f \in \mathcal{M} \) and \( a \in \mathbb{R} \).

As a consequence of the above paragraphs and the \( \pi - \lambda \) Theorem \([6.54]\) \( \mathcal{L} \) contains \( \sigma (\mathcal{P}) = \sigma (\mathcal{M}) \). In particular it follows that \( 1_A \in \mathbb{H} \) for all \( A \in \sigma (\mathcal{M}) \).

Since any positive \( \sigma (\mathcal{M}) \) – measurable function may be written as an increasing limit of simple functions (see Theorem \([3.39]\)), it follows that \( \mathbb{H} \) contains all non-negative bounded \( \sigma (\mathcal{M}) \) – measurable functions. Finally, since any bounded \( \sigma (\mathcal{M}) \) – measurable functions may be written as the difference of two such non-negative simple functions, it follows that \( \mathbb{H} \) contains all bounded \( \sigma (\mathcal{M}) \) – measurable functions.

**Corollary 11.3.** 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 \( \mathcal{M} = \{1\} \cup \{A : A \in \mathcal{P}\} \). Then \( \mathcal{M} \subseteq \mathbb{H} \) is a multiplicative system and the proof is completed with an application of Theorem \([11.2]\).

**Example 11.4.** Suppose \( \mu \) and \( \nu \) are two probability measure on \( (\Omega, \mathcal{B}) \) such that
\[
\int_{\Omega} f \, d\mu = \int_{\Omega} f \, d\nu
\]
for all \( f \) in a multiplicative subset, \( \mathcal{M} \), of bounded measurable functions on \( \Omega \). Then \( \mu = \nu \) on \( \sigma (\mathcal{M}) \). Indeed, apply Theorem \([11.2]\) with \( \mathbb{H} \) being the bounded measurable functions on \( \Omega \) such that Eq. \((11.1)\) holds. In particular if \( \mathcal{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\} \) with \( \mathcal{P} \) being a multiplicative class we learn that \( \mu = \nu \) on \( \sigma (\mathcal{M}) = \sigma (\mathcal{P}) \).

**Exercise 11.1.** Let \( \Omega := \{1, 2, 3, 4\} \) and \( \mathcal{M} := \{1_A, 1_B\} \) where \( A := \{1, 2\} \) and \( B := \{2, 3\} \).

a) Show \( \sigma (\mathcal{M}) = 2^\Omega \).

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. \((11.1)\) holds for all \( f \in \mathcal{M} \).

**Moral:** the assumption that \( \mathcal{M} \) is multiplicative can not be dropped from Theorem \([11.2]\).

Here is a complex version of Theorem \([11.2]\).

**Theorem 11.5 (Complex Multiplicative System Theorem).** Suppose \( \mathbb{H} \) is a complex linear subspace of the bounded complex functions on \( \Omega \). \( 1 \in \mathbb{H} \) is closed under complex conjugation, and \( \mathbb{H} \) is closed under bounded convergence. If \( \mathbb{M} \subseteq \mathbb{H} \) is a multiplicative system which is closed under conjugation, then \( \mathbb{H} \) contains all bounded complex valued \( \sigma (\mathcal{M}) \)-measurable functions.

**Proof.** Let \( \mathcal{M}_0 = \text{span}_\mathbb{C} (\mathcal{M} \cup \{1\}) \) be the complex span of \( \mathcal{M} \). As the reader should verify, \( \mathcal{M}_0 \) is an algebra, \( \mathcal{M}_0 \subseteq \mathbb{H} \), \( \mathcal{M}_0 \) is closed under complex conjugation and \( \sigma (\mathcal{M}_0) = \sigma (\mathcal{M}) \). Let
\[
\mathbb{H}^c := \{f \in \mathbb{H} : f \text{ is real valued} \}
\]
and
\[
\mathbb{M}^c := \{f \in \mathcal{M}_0 : f \text{ is real valued} \}.
\]
Then \( \mathbb{H}^c \) is a real linear space of bounded real valued functions which is closed under bounded convergence and \( \mathcal{M}_0^c \subset \mathbb{H}^c \). Moreover, \( \mathcal{M}_0^c \) is a multiplicative system (as the reader should check) and therefore by Theorem 11.2 \( \mathbb{H}^c \) contains all bounded \( \sigma (\mathcal{M}_0^c) \) – measurable real valued functions. Thus \( \mathcal{M}_0^c \) and \( \mathcal{M}_0 \) are complex linear spaces closed under complex conjugation, for any \( f \in \mathcal{M}_0 \) or \( f \in \mathcal{M}_0 \), 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 \( \mathcal{M}_0 \) respectively. Therefore \( \mathcal{M}_0 = \mathcal{M}_0^c + i\mathcal{M}_0^c \), \( \sigma (\mathcal{M}_0^c) = \sigma (\mathcal{M}_0) = \sigma (\mathcal{M}) \), and \( \mathbb{H} = \mathbb{H}^c + i\mathbb{H}^c \). Hence if \( f : \Omega \to \mathbb{C} \) is a bounded \( \sigma (\mathcal{M}) \) – measurable function, then \( f = \text{Re} f + i\text{Im} f \in \mathbb{H} \) since \( \text{Re} f \) and \( \text{Im} f \) are in \( \mathbb{H}^c \).
Lemma 11.6. 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

\[
f_n(x) := \begin{cases} 
0 & \text{on } (-\infty, a] \cup [b + \frac{1}{n}, \infty) \\
\frac{n(x-a)}{n} & \text{if } a \leq x \leq a + \frac{1}{n} \\
1 & \text{if } a + \frac{1}{n} \leq x \leq b \\
1 - \frac{n(b-x)}{n} & \text{if } b \leq x \leq b + \frac{1}{n} 
\end{cases}
\]

This shows \(\mathcal{B}_R \subset \sigma(e_{\lambda} : \lambda > 0)\). As \(e_{\lambda}\) is continuous and hence Borel measurable for all \(\lambda > 0\) we automatically know that \(\sigma(e_{\lambda} : \lambda > 0) \subset \mathcal{B}_R\).

**Remark 11.8.** A slight modification of the above proof actually shows if \(\{\lambda_n\} \subset (0, \infty)\) with \(\lim_{n \to \infty} \lambda_n = 0\), then \(\sigma(e_{\lambda_n} : n \in \mathbb{N}) = \mathcal{B}_R\).

**Corollary 11.9.** Each of the following \(\sigma\)-algebras on \(\mathbb{R}^d\) are equal to \(\mathcal{B}_{\mathbb{R}^d}\):

1. \(\mathcal{M}_1 := \sigma(\cup_{i=1}^n \{x \to f_i(x) : f_i \in C_c(\mathbb{R})\})\),
2. \(\mathcal{M}_2 := \sigma(x \to f_1(x) \ldots f_d(x) : f_i \in C_c(\mathbb{R}))\),
3. \(\mathcal{M}_3 := \sigma(C_c(\mathbb{R}^d))\), and
4. \(\mathcal{M}_4 := \sigma(\{x \to e^{i\lambda x} : \lambda \in \mathbb{R}^d\})\).

**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 11.6 there exists \(f_n \in C_c(\mathbb{R})\) such that \(\lim_{n \to \infty} f_n = 1_{(a,b)}\). Therefore it follows that \(x \to 1_{(a,b)}(x)\) is \(\mathcal{M}_1\) measurable for each \(i\). Moreover if \(-\infty < a_i < b_i < \infty\) for each \(i\), then we may conclude that

\[
x \to \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = 1_{(a_1,b_1] \times \ldots \times (a_d,b_d]}(x)
\]

is \(\mathcal{M}_1\) measurable as well and hence \((a_1,b_1] \times \ldots \times (a_d,b_d] \in \mathcal{M}_1\). As such sets generate \(\mathcal{B}_{\mathbb{R}^d}\) we may conclude that \(\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_1\).

\(\mathcal{M}_2\) case. As above, we may find \(f_{i,n} \to 1_{(a_i,b_i)}\) as \(n \to \infty\) for each \(1 \leq i \leq d\) and therefore,

\[
1_{(a_1,b_1] \times \ldots \times (a_d,b_d]}(x) = \lim_{n \to \infty} f_{1,n}(x_1) \ldots f_{d,n}(x_d) \text{ for all } x \in \mathbb{R}^d.
\]
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 \( \mathcal{B}_{\mathbb{R}^d} = \mathcal{M}_2 \subset \mathcal{M}_3 \subset \mathcal{B}_{\mathbb{R}} \).

\( \mathcal{M}_4 \) case. Let \( \pi_j : \mathbb{R}^d \to \mathbb{R} \) be projection onto the \( j \)th factor, then for \( \lambda > 0 \), \( e_\lambda \circ \pi_j (x) = e^{i \lambda x_j} \). It then follows that

\[
\sigma (e_\lambda \circ \pi_j : \lambda > 0) = \sigma \left( (e_\lambda \circ \pi_j)^{-1} (W) : \lambda > 0, W \in \mathcal{B}_\mathbb{C} \right) = \sigma (\pi_j^{-1} (\sigma ((e_\lambda^{-1} (W)) : \lambda > 0, W \in \mathcal{B}_\mathbb{C}))) = \pi_j^{-1} (\mathcal{B}_\mathbb{R})
\]

wherein we have used Lemma \[11.7\] for the last equality. Since \( \sigma (e_\lambda \circ \pi_j : \lambda > 0) \subset \mathcal{M}_4 \) for each \( j \) we must have

\[
\mathcal{B}_{\mathbb{R}^d} = \mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R} = \sigma (\pi_j : 1 \leq j \leq d) \subset \mathcal{M}_4.
\]

\[\text{Alternative proof.}\] By Lemma \[11.15\] below there exists \( g_n \in \text{Trig}(\mathbb{R}) \) such that \( \lim_{n \to \infty} g_n = 1_{(a, b]} \). Since \( x \to g_n (x_i) \) is in the span \( \{ x \to e^{i \lambda x} : \lambda \in \mathbb{R}^d \} \) 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 \( \mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_4 \).

\[\text{Corollary 11.10.}\] 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. \( \mathbb{M}_1 := \{ x \to f_1 (x_1) \cdots f_d (x_d) : f_i \in C_c (\mathbb{R}) \} \)
2. \( \mathbb{M}_2 := C_c (\mathbb{R}^d) \), or
3. \( \mathbb{M}_3 := \{ x \to e^{i \lambda x} : \lambda \in \mathbb{R}^d \} \)

then \( \mathbb{H} \) contains all bounded complex Borel measurable functions on \( \mathbb{R}^d \).

\[\text{Proof.}\] 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

\[
1 = \lim_{n \to \infty} f_n (x_1) \cdots f_n (x_d).
\]

In case 3, \( 1 \in \mathbb{M}_3 \subset \mathbb{H} \) as well. The result now follows from Theorem \[11.5\] and Corollary \[11.9\].

\[\text{Proposition 11.11 (Change of Variables Formula).}\] 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

\[
\int_{u(a)}^{u(b)} f (x) \, dx = \int_{a}^{b} f (u(t)) \dot{u} (t) \, dt.
\]

Moreover, Eq. \[11.2\] is also valid if \( f : [c, d] \to \mathbb{R} \) is measurable and

\[
\int_{a}^{b} |f (u(t))| \, |\dot{u} (t)| \, dt < \infty.
\]

\[\text{Proof.}\] Let \( \mathbb{H} \) denote the space of bounded measurable functions such that Eq. \[11.2\] 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 \[11.10\] 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,

\[
\int_{a}^{b} f (u(t)) \dot{u} (t) \, dt = \int_{a}^{b} F' (u(t)) \dot{u} (t) \, dt
\]

\[
= \int_{a}^{b} \frac{d}{dt} F (u(t)) \, dt = F (u(b)) - F (u(a)) = \int_{u(a)}^{u(b)} F' (x) \, dx = \int_{u(a)}^{u(b)} f (x) \, dx.
\]

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. \[11.3\] holds. For \( M < \infty \), let \( f_M (x) = f (x) \cdot 1_{|f(x)| \leq M} \) - a bounded measurable function. Therefore applying Eq. \[11.2\] with \( f \) replaced by \( |f_M| \) shows,

\[
\left| \int_{u(a)}^{u(b)} |f_M (x)| \, dx \right| \leq \int_{a}^{b} |f_M (u(t))| \, |\dot{u} (t)| \, dt \leq \int_{a}^{b} |f_M (u(t))| \, |\dot{u} (t)| \, dt.
\]

Using the MCT, we may let \( M \uparrow \infty \) in the previous inequality to learn

\[
\left| \int_{u(a)}^{u(b)} f (x) \, dx \right| \leq \int_{a}^{b} |f (u(t))| \, |\dot{u} (t)| \, dt < \infty.
\]

Now apply Eq. \[11.2\] 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. (11.2) remains valid.

**Exercise 11.2.** 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

$$\int_{\mathbb{R}} f(x) \, dx = \int_{\mathbb{R}} f(u(t)) \, \dot{u}(t) \, dt \tag{11.4}$$

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 11.12.** 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

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^d} e^{ix \cdot \lambda} \, d\mu(x) \text{ for all } \lambda \in \mathbb{R}^d.$$

**Corollary 11.13.** 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. $\int_{\mathbb{R}^d} f_1(x_1) \ldots f_d(x_d) \, dv(x_d) = \int_{\mathbb{R}^d} f_1(x_1) \ldots f_d(x_d) \, d\mu(x)$ for all $f_i \in \mathcal{C}_c(\mathbb{R})$.
2. $\int_{\mathbb{R}^d} f(x) \, dv(x) = \int_{\mathbb{R}^d} f(x) \, d\mu(x)$ for all $f \in \mathcal{C}_c(\mathbb{R}^d)$.
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

$$\int_{\mathbb{R}^d} f \, d\mu = \int_{\mathbb{R}^d} f \, dv. \tag{11.5}$$

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 [11.10], 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. (11.5) 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 [11.13] by:

$$\int_{\mathbb{R}^d} e^{ix \cdot \lambda} \, d\mu(x) = \int_{\mathbb{R}^d} e^{ix \cdot \lambda} \, dv(x) \text{ for all } \lambda \in U, \tag{11.6}$$

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 [11.6] implies $\hat{\nu} = \hat{\mu}$. You are asked to carry out this argument in Exercise [11.3] making use of the following lemma.

**Lemma 11.14 (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

$$h(z + ib) = \sum_{n=0}^\infty c_n(b) z^n \text{ for all } |z| < \varepsilon. \tag{11.7}$$

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. (11.7) 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

$$\sum_{n=0}^\infty c_n(y) z^n = h(x + iy) = 0 \text{ for all } |z| < \delta,$$

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 11.15 (This may be omitted.).** Suppose that $-\infty < a < b < \infty$ and let $\text{Trig}(\mathbb{R}) \subset C(\mathbb{R}, \mathbb{C})$ be the complex linear span of $\{ x \to e^{ix \cdot \lambda} : \lambda \in \mathbb{R} \}$. Then there exists $f_n \in C_c(\mathbb{R}, [0, 1])$ and $g_n \in \text{Trig}(\mathbb{R})$ such that $\lim_{n \to \infty} f_n(x) = 1_{(a,b)}(x)$.

**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 $g_n \in \text{Trig}(\mathbb{R})$, 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 $\{ g_n \} \subset \text{Trig}(\mathbb{R})$ 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 [11.3]. If $\varepsilon > 0$ is given, we may apply Theorem [5.75] to find $A \subset_f \mathbb{Z}$ such that
\[ h(z + ib) = \mathbb{E} \left[ e^{ibX}e^{izX} \right] - \mathbb{E} \left[ e^{ibY}e^{izY} \right] = \sum_{n=0}^{\infty} c_n(b) z^n \quad (11.8) \]

where
\[ c_n(b) := \frac{1}{n!} \left( \mathbb{E} \left[ e^{ibX} X^n \right] - \mathbb{E} \left[ e^{ibY} Y^n \right] \right). \quad (11.9) \]

3. Conclude from item 2. that \( X \overset{d}{=} Y \), i.e. that \( \text{Law}_P(X) = \text{Law}_P(Y) \).

**Exercise 11.4.** Let \((\Omega, \mathcal{B}, P)\) be a probability space and \(X, Y : \Omega \to \mathbb{R}\) be a pair of random variables such that
\[ \mathbb{E} \left[ f(X)g(Y) - E[f(X)g(X)] \right] \]
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 11.2 to show \(\mathbb{H}\) is the vector space of all bounded Borel measurable functions. Then take \(h(x, y) = 1_{\{x=y\}}\).

**Exercise 11.5 (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 \(\mathcal{A}\) – 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 11.16.** Suppose that \((\Omega, \mathcal{B}, P)\) is a probability space, \(\{X_n\}_{n=1}^{\infty}\) is a collection of random variables on \(\Omega\), and \(\mathcal{B}_\infty := \sigma(\{X_1, X_2, X_3, \ldots\})\). 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 \(\mathcal{B}_{\mathcal{B}_\infty}\) – measurable function \(G : \mathbb{R}^n \to \mathbb{R}\) such that \(\mathbb{E}[|f - G(X_1, \ldots, X_n)|] < \varepsilon\). Moreover we may assume that \(\sup_{x \in \mathbb{R}^n} |G(x)| \leq M := \sup_{x \in \Omega} |f(\omega)|\).

**Proof.** Apply Exercise 11.5 with \(\mathcal{A} := \bigcup_{n=1}^{\infty} \mathcal{A}(X_1, \ldots, X_n)\) 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 \(\mathcal{A}(X_1, \ldots, X_n)\) – measurable for some \(n \in \mathbb{N}\). It now follows by the factorization Lemma 9.40 that \(\varphi = G(X_1, \ldots, X_n)\) for some \(\mathcal{B}_{\mathcal{B}_\infty}\) – measurable function \(G : \mathbb{R}^n \to \mathbb{R}\). If necessary, replace \(G\) by \([G \land \mathcal{M}] \cup (-M)\) in order to insure \(\sup_{x \in \mathbb{R}^n} |G(x)| \leq M\).

**Exercise 11.6 (Density of \(\mathcal{A}\) in \(\mathcal{B} = \sigma(\mathcal{A})\)).** Keeping the same notation as in Exercise 11.5 but now take \(f = 1_B\) for some \(B \in \mathcal{B}\) and given \(\varepsilon > 0\), write
Exercise 11.8. Suppose \( 1 \) 1.1.2.1 Obsolete stuff follows.

\[
\varphi = \sum_{i=0}^{n} \lambda_i 1_{A_i}, \quad \text{where } \lambda_0 = 0, \{\lambda_i\}_{i=1}^{n} \text{ is an enumeration of } \varphi(\Omega) \setminus \{0\}, \text{ and } A_i := \{\varphi = \lambda_i\}. \text{ Show: 1.}
\]

\[
E[1_B - \varphi] = P(A_0 \cap B) + \sum_{i=1}^{n} [1 - \lambda_i] P(B \cap A_i) + \lambda_i P(A_i \setminus B) \quad (11.10)
\]

\[
\geq P(A_0 \cap B) + \sum_{i=1}^{n} \min \{P(B \cap A_i), P(A_i \setminus B)\}. \quad (11.11)
\]

2. 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

\[
E[1_B - \psi] = P(A_0 \cap B) + \sum_{i=1}^{n} \min \{P(B \cap A_i), P(A_i \setminus B)\} \leq E[1_B - \varphi].
\]

Observe that \( \psi = 1_D \) where \( D = \cup_{i=1}^{n} A_i \in A \) and so you have shown; for every \( \varepsilon > 0 \) there exists a \( D \in A \) such that

\[
P(B \Delta D) = E[1_B - 1_D] < \varepsilon.
\]

Exercise 11.7. Suppose that \( \{(X_i, B_i)\}_{i=1}^{n} \) are measurable spaces and for each \( i, M_i \) is a multiplicative system of real bounded measurable functions on \( X_i \) such that \( \sigma(M_i) = B_i \) and there exist \( \chi_n \in M_i \) such that \( \chi_n \rightarrow 1 \) boundedly as \( n \rightarrow \infty \). Define \( f_i: X_i \rightarrow \mathbb{R} \) by \( f_i \otimes \cdots \otimes f_n: X_1 \times \cdots \times X_n \rightarrow \mathbb{R} \) be defined by

\[
(f_1 \otimes \cdots \otimes f_n)(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n).
\]

Show

\[
M_1 \otimes \cdots \otimes M_n := \{f_1 \otimes \cdots \otimes f_n: f_i \in M_i \text{ for } 1 \leq i \leq n\}
\]
is a multiplicative system of bounded measurable functions on \((X := X_1 \times \cdots \times X_n, \mathcal{B} := B_1 \otimes \cdots \otimes B_n)\) such that \( \sigma(M_1 \otimes \cdots \otimes M_n) = \mathcal{B} \).

11.3 A Strengthening of the Multiplicative System Theorem

Notation 11.18 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 \( \omega \in \Omega \), \( \text{then } f := \lim_{n \to \infty} f_n \in \mathbb{H} \).

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 11.19. 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 11.2.

Theorem 11.20. 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 11.19 reduces this theorem to Theorem 11.2.
11.4 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

**Notation 11.21** Given a collection of bounded functions, \( M \), from a set, \( \Omega \), to \( \mathbb{R} \), let \( M_{\uparrow} (M_{\downarrow}) \) denote the bounded monotone increasing (decreasing) limits of functions from \( M \). More explicitly a bounded function, \( f : \Omega \to \mathbb{R} \) is in \( M_{\uparrow} \) respectively \( M_{\downarrow} \) iff there exists \( f_n \in M \) such that \( f_n \uparrow f \) respectively \( f_n \downarrow f \).

**Theorem 11.22** (Bounded Approximation Theorem*). Let \((\Omega, \mathcal{B}, \mu)\) be a finite measure space and \( M \) be an algebra of bounded \( \mathbb{R} \) – valued measurable functions such that:

1. \( \sigma (M) = \mathcal{B} \),
2. 1 \( \in M \), and
3. \(|f| \in M \) for all \( f \in M \).

Then for every bounded \( \sigma (M) \) measurable function, \( g : \Omega \to \mathbb{R} \), and every \( \varepsilon > 0 \), there exists \( f \in M_{\downarrow} \) and \( h \in M_{\uparrow} \) such that \( f \leq g \leq h \) and \( \mu (h - f) < \varepsilon \).

**Proof.** Let us begin with a few simple observations.

1. \( M \) is a “lattice” – if \( f, g \in M \) then

\[
 f \lor g = \frac{1}{2} (f + g + |f - g|) \in M
\]

and

\[
 f \land g = \frac{1}{2} (f + g - |f - g|) \in M.
\]

2. If \( f, g \in M_{\downarrow} \) or \( f, g \in M_{\uparrow} \) then \( f + g \in M_{\downarrow} \) or \( f + g \in M_{\uparrow} \) respectively.
3. If \( \lambda \geq 0 \) and \( f \in M_{\downarrow} \) then \( \lambda f \in M_{\downarrow} \) (\( \lambda f \in M_{\uparrow} \)).
4. If \( f \in M_{\uparrow} \) then \( -f \in M_{\downarrow} \) and visa versa.
5. If \( f_n \in M_{\uparrow} \) and \( f_n \uparrow f \) where \( f : \Omega \to \mathbb{R} \) is a bounded function, then \( f \in M_{\uparrow} \).

Indeed, by assumption there exists \( f_{n,i} \in M \) such that \( f_{n,i} \uparrow f_n \) as \( i \to \infty \).

By observation (1), \( g_n := \max \{ f_{ij} : i, j \leq n \} \in M \). Moreover it is clear that \( g_n \leq \max \{ f_k : k \leq n \} = f_n \leq f \) and hence \( g_n \uparrow g := \lim_{n \to \infty} g_n \leq f \). Since \( f_{ij} \leq g \) for all \( i, j \), it follows that \( f_n = \lim_{j \to \infty} f_{ij} \leq g \) and consequently that \( f = \lim_{n \to \infty} f_n \leq g \leq f \). So we have shown that \( g_n \uparrow f \in M_{\uparrow} \).

Now let \( \mathbb{H} \) denote the collection of bounded measurable functions which satisfy the assertion of the theorem. Clearly, \( M \subset \mathbb{H} \) and in fact it is also easy to see that \( M_{\uparrow} \) and \( M_{\downarrow} \) are contained in \( \mathbb{H} \) as well. For example, if \( f \in M_{\uparrow} \), by

\[
\sum_{n=1}^{\infty} v_n = \lim_{N \to \infty} \sum_{n=1}^{N} v_n \in M_{\uparrow} \text{ (using observations 2. and 5.)}
\]

and for \( N \in \mathbb{N} \), let

\[
u^N := \sum_{n=1}^{N} u_n \in M_{\uparrow} \text{ (using observation 2).}
\]

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 \( u^N \leq g - g_1 \leq v \). Moreover,
\[
\mu \left( v - u^N \right) = \sum_{n=1}^{N} \mu \left( v_n - u_n \right) + \sum_{n=N+1}^{\infty} \mu \left( v_n \right) \leq \sum_{n=1}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \mu \left( v_n \right)
\]
\[
\leq \varepsilon + \sum_{n=N+1}^{\infty} \mu \left( v_n \right).
\]

However, since
\[
\sum_{n=1}^{\infty} \mu \left( v_n \right) \leq \sum_{n=1}^{\infty} \mu \left( \delta_n + \varepsilon 2^{-n} \right) = \sum_{n=1}^{\infty} \mu \left( \delta_n \right) + \varepsilon \mu \left( \Omega \right)
\]
\[
= \sum_{n=1}^{\infty} \mu \left( g - g_1 \right) + \varepsilon \mu \left( \Omega \right) < \infty,
\]
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, it 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 \( g = (g - g_1) + g_1 \in \mathbb{H} \). \( \blacksquare \)
12

Multiple and Iterated Integrals

12.1 Iterated Integrals

Notation 12.1 (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 \textit{iterated integrals} of \(f\) (when they make sense) are:

\[
\int_X \mathrm{d}\mu(x) \int_Y \mathrm{d}\nu(y) f(x, y) := \int_X \left[ \int_Y f(x, y) \mathrm{d}\nu(y) \right] \mathrm{d}\mu(x)
\]

and

\[
\int_Y \mathrm{d}\nu(y) \int_X \mathrm{d}\mu(x) f(x, y) := \int_Y \left[ \int_X f(x, y) \mathrm{d}\mu(x) \right] \mathrm{d}\nu(y).
\]

Notation 12.2 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 \((\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_C)\) - 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.

12.2 Tonelli’s Theorem and Product Measure

Theorem 12.3. Suppose \((X, \mathcal{M}, \mu)\) and \((Y, \mathcal{N}, \nu)\) are \(\sigma\)-finite measure spaces and \(f\) is a nonnegative \((\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_B)\) - measurable function, then for each \(y \in Y\),

\[
x \to f(x, y) \text{ is } \mathcal{M} - \mathcal{B}_{[0, \infty]} \text{ measurable},
\]

for each \(x \in X\),

\[
y \to f(x, y) \text{ is } \mathcal{N} - \mathcal{B}_{[0, \infty]} \text{ measurable},
\]

and

\[
x \to \int_Y f(x, y) \mathrm{d}\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0, \infty]} \text{ measurable},
\]

\[
y \to \int_X f(x, y) \mathrm{d}\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0, \infty]} \text{ measurable},
\]

and

\[
\int_X \mathrm{d}\mu(x) \int_Y \mathrm{d}\nu(y) f(x, y) = \int_Y \mathrm{d}\nu(y) \int_X \mathrm{d}\mu(x) f(x, y).
\]

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. (12.1) and (12.2) hold. Moreover

\[
\int_Y f(x, y) \mathrm{d}\nu(y) = \int_Y 1_A(x)1_B(y) \mathrm{d}\nu(y) = 1_A(x)\nu(B),
\]

so that Eq. (12.3) holds and we have

\[
\int_X \mathrm{d}\mu(x) \int_Y \mathrm{d}\nu(y) f(x, y) = \nu(B)\mu(A).
\]

Similarly,

\[
\int_Y \mathrm{d}\nu(y) \int_X \mathrm{d}\mu(x) f(x, y) = \nu(B)\mu(A)
\]

from which it follows that Eqs. (12.4) and (12.5) 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 \((\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_B)\) - measurable functions on \(X \times Y\) such that Eqs. (12.1) – (12.5) 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 11.3 that \(\mathbb{H}\) is the space
of all bounded \( (M \otimes N, B) \) – measurable functions on \( X \times Y \). Moreover, if \( f : X \times Y \to [0, \infty] \) is a \( (M \otimes N, B) \) – measurable function, let \( f_M = M \wedge f \) so that \( f_M \uparrow f \) as \( M \to \infty \). Then Eqs. (12.1) – (12.5) 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 \( X_n \in M, Y_n \in N \) such that \( X_n \uparrow X, Y_n \uparrow Y, \mu(X_n) < \infty \) and \( \nu(Y_n) < \infty \) for all \( n \in \mathbb{N} \). Then define \( \mu_m(A) = \mu(X_m \cap A) \) and \( \nu_n(B) = \nu(Y_n \cap B) \) for all \( A \in M \) and \( B \in N \) 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. (12.1) – (12.5) with \( \mu \) replaced by \( \mu_m \) and \( \nu \) by \( \nu_n \) for all \( (M \otimes N, B) \) – measurable functions, \( f : X \times Y \to [0, \infty] \). The validity of Eqs. (12.1) – (12.5) 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 \( u \in L^+(X, M) \),

\[
\int_X u d\mu_m = \int_X u 1_{X_m} d\mu \uparrow \int_X u d\mu \quad \text{as} \quad m \to \infty,
\]

and for all \( v \in L^+(Y, N) \),

\[
\int_Y v d\nu_n = \int_Y v 1_{Y_n} d\nu \uparrow \int_Y v d\nu \quad \text{as} \quad n \to \infty.
\]

\[\square\]

**Corollary 12.4.** 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

\[
\pi(E) = \int_X d\mu(x) \int_Y d\nu(y) 1_E(x, y) = \int_Y d\nu(y) \int_X d\mu(x) 1_E(x, y) \quad (12.7)
\]

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 7.1 Proposition 4.27 with \( E = \mathcal{M} \times \mathcal{N} \). For the existence, it suffices to observe, using the monotone convergence theorem, that \( \pi \) defined in Eq. (12.7) 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. (12.6).

\[\square\]

**Notation 12.5.** The measure \( \pi \) is called the product measure of \( \mu \) and \( \nu \) and will be denoted by \( \mu \otimes \nu \).

Theorem 12.6 (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 \( x \in X \),

\[
\int_Y f(\cdot, y) d\nu(y) \in L^+(X, \mathcal{M}), \quad \int_X f(x, \cdot) d\mu(x) \in L^+(Y, \mathcal{N})
\]

and

\[
\int_{X \times Y} f \, d\pi = \int_X d\mu(x) \int_Y d\nu(y) f(x, y) \quad (12.8)
\]

\[
= \int_Y d\nu(y) \int_X d\mu(x) f(x, y). \quad (12.9)
\]

**Proof.** By Theorem 12.3 and Corollary 12.3 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 9.39 one deduces the theorem for general \( f \in L^+(X \times Y, \mathcal{M} \otimes \mathcal{N}) \).

**Example 12.7.** In this example we are going to show, \( I := \int_R e^{-x^2/2} dm(x) = \sqrt{2\pi} \). To this end we observe, using Tonelli’s theorem, that

\[
I^2 = \left[ \int_R e^{-x^2/2} dm(x) \right]^2 = \int_R e^{-y^2/2} \left[ \int_R e^{-x^2/2} dm(x) \right] d\nu(y)
\]

\[
= \int_{R^2} e^{-(x^2+y^2)/2} dm^2(x, y)
\]

where \( m^2 = m \otimes m \) is “Lebesgue measure” on \( (R^2, B_{R^2} = B_R \otimes B_R) \). From the monotone convergence theorem,

\[
I^2 = \lim_{R \to \infty} \int_{D_R} e^{-(x^2+y^2)/2} dm^2(x, y)
\]

where \( D_R = \{(x, y) : x^2 + y^2 < R^2\} \). Using the change of variables theorem described in Section 12.5 below\footnote{Alternatively, you can easily show that the integral \( \int_{D_R} f dm^2 \) agrees with the multiple integral in undergraduate analysis when \( f \) is continuous. Then use the change of variables theorem from undergraduate analysis.} we find

\[
\int_{D_R} e^{-(x^2+y^2)/2} dm^2(x, y) = \int_{(0, R) \times (0, 2\pi)} e^{-r^2/2} r dr d\theta
\]

\[
= 2\pi \int_0^R e^{-r^2/2} r dr = 2\pi \left(1 - e^{-R^2/2}\right).
\]
From this we learn that
\[ I^2 = \lim_{R \to \infty} 2\pi \left( 1 - e^{-R^2/2} \right) = 2\pi \]
as desired.

### 12.3 Fubini’s Theorem

**Notation 12.8** If \((X, \mathcal{M}, \mu)\) is a measure space and \(f : X \to \mathbb{C}\) is any measurable function, let
\[
\int_X f \, d\mu := \begin{cases} \int_X f \, d\mu & \text{if } \int_X |f| \, d\mu < \infty \\ 0 & \text{otherwise.} \end{cases}
\]

**Theorem 12.9 (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:

1. \(\int_{X \times Y} |f| \, d\pi < \infty\), i.e. \(f \in L^1(\pi)\),
2. \(\int_X \left( \int_Y |f(x,y)| \, d\nu(y) \right) \, d\mu(x) < \infty\) and
3. \(\int_Y \left( \int_X |f(x,y)| \, d\mu(x) \right) \, d\nu(y) < \infty\).

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. \(y\), \(\int_Y f(\cdot, y) \, d\nu(y) \in L^1(\mu)\), \(\int_X f(x, \cdot) \, d\mu(x) \in L^1(\nu)\) and Eqs. \((12.8)\) and \((12.9)\) are still valid after putting a bar over the integral symbols.

**Proof.** The equivalence of Eqs. \((12.10)\) – \((12.12)\) is a direct consequence of Tonelli’s Theorem \[12.6\]. Now suppose \(f \in L^1(\pi)\) is a real valued function and let
\[
E := \left\{ x \in X : \int_Y |f(x,y)| \, d\nu(y) = \infty \right\}. \tag{12.13}
\]

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
\[
\int_Y f(x,y) \, d\nu(y) = \int_Y 1_{E^c}(x) f(x,y) \, d\nu(y) = \int_Y 1_{E^c}(x) [f_{+}(x,y) - f_{-}(x,y)] \, d\nu(y)
\]
\[
= \int_Y 1_{E^c}(x) f_{+}(x,y) \, d\nu(y) - \int_Y 1_{E^c}(x) f_{-}(x,y) \, d\nu(y).
\]

(12.14)

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 \(x \to \int_Y f(x,y) \, d\nu(y)\) is \(\mathcal{M}\)–measurable, being the difference of two measurable functions. Moreover
\[
\int_X \left[ \int_Y |f(x,y)| \, d\nu(y) \right] \, d\mu(x) \leq \int_X \left[ \int_Y |f(x,y)| \, d\nu(y) \right] \, d\mu(x) < \infty,
\]
which shows \(\int_Y f(x,y) \, d\nu(y) \in L^1(\mu)\). Integrating Eq. \((12.14)\) on \(x\) and using Tonelli’s theorem repeatedly implies,
\[
\int_X \left[ \int_Y f(x,y) \, d\nu(y) \right] \, d\mu(x) = \int_X d\mu(x) \int_Y d\nu(y) 1_{E^c}(x) f_{+}(x,y) - \int_X d\mu(x) \int_Y d\nu(y) 1_{E^c}(x) f_{-}(x,y)
\]
\[
= \int_Y d\nu(y) \int_X d\mu(x) 1_{E^c}(x) f_{+}(x,y) - \int_Y d\nu(y) \int_X d\mu(x) 1_{E^c}(x) f_{-}(x,y)
\]
\[
= \int_Y d\nu(y) \int_X d\mu(x) f_{+}(x,y) - \int_Y d\nu(y) \int_X d\mu(x) f_{-}(x,y)
\]
\[
= \int_{X \times Y} f_{+} \, d\pi - \int_{X \times Y} f_{-} \, d\pi = \int_{X \times Y} (f_{+} - f_{-}) \, d\pi = \int_{X \times Y} f \, d\pi \tag{12.15}
\]
which proves Eq. \((12.8)\) holds.

Now suppose that \(f = u + iv\) is complex valued and again let \(E\) be as in Eq. \((12.13)\). Just as above we still have \(E \in \mathcal{M}\) and \(\mu(E) = 0\) and
\[
\int_Y f(x,y) \, d\nu(y) = \int_Y 1_{E^c}(x) f(x,y) \, d\nu(y) = \int_Y 1_{E^c}(x) [u(x,y) + iv(x,y)] \, d\nu(y)
\]
\[
= \int_Y 1_{E^c}(x) u(x,y) \, d\nu(y) + i \int_Y 1_{E^c}(x) v(x,y) \, d\nu(y).
\]
This measure and its completion will be denoted by $M(X)$.

The previous theorems generalize to products of any finite number of $\sigma$–finite measure spaces.

**Theorem 12.10.** Suppose $\{(X_i, M_i, \mu_i)\}_{i=1}^n$ are $\sigma$–finite measure spaces and $X := X_1 \times \cdots \times X_n$. Then there exists a unique measure $(\pi)$ on $(X, M_1 \otimes \cdots \otimes M_n)$ such that

$$\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n) \text{ for all } A_i \in M_i. \quad (12.16)$$

(This measure and its completion will be denoted by $\mu_1 \otimes \cdots \otimes \mu_n$.) If $f : X \to [0, \infty)$ is a $M_1 \otimes \cdots \otimes M_n$–measurable function then

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \cdots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) f(x_1, \ldots, x_n) \quad (12.17)$$

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

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \cdots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) f(x_1, \ldots, x_n) \quad (12.18)$$

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 12.6 and 12.9. 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 $X_i = \prod_{j \neq i} X_j$, $M_i := \otimes_{j \neq i} M_j$, and $\mu_i := \otimes_{j \neq i} \mu_j$ be the product measure on $(X_i, M_i)$ which is assumed to exist by the induction hypothesis. Also let $M := M_1 \otimes \cdots \otimes M_n$ 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 $M$–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 $M_i$–measurable. Thus by the induction hypothesis, the right side of Eq. (12.17) is well defined.

2. If $\sigma \in S_n$ (the permutations of $\{1, 2, \ldots, n\}$) we may define a measure $\pi$ on $(X, M)$ by:

$$\pi(A) := \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \cdots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) 1_A(x_1, \ldots, x_n). \quad (12.19)$$

It is easy to check that $\pi$ is a measure which satisfies Eq. (12.16). Using the $\sigma$–finiteness assumptions and the fact that

$$\mathcal{P} := \{A_1 \times \cdots \times A_n : A_i \in M_i \text{ for } 1 \leq i \leq n\}$$

is a $\pi$–system such that $\mathcal{P} = M$, it follows from Exercise 6.8 that there is only one such measure satisfying Eq. (12.16). Thus the formula for $\pi$ in Eq. (12.19) is independent of $\sigma \in S_n$.

3. From Eq. (12.19) and the usual simple function approximation arguments we may conclude that Eq. (12.17) is valid.

Now suppose that $f \in L^1(X, M, \pi)$.

4. Using step 1 it is easy to check that

$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} f(x_1, \ldots, x_i, \ldots, x_n) d\mu_i(x_i)$$

is $M_i$–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 $M_i$–measurable and therefore

$$E := \left\{ (x_1, \ldots, \hat{x}_i, \ldots, x_n) : \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)| d\mu_i(x_i) < \infty \right\} \in M_i.$$
It now follows by induction that the right side of Eq. (12.18) is well defined.

6. Let \( i = \sigma n \) and \( T : X \to X_i \times X^i \) be the obvious identification;

\[
T(x_i, (x_1, \ldots, x_n)) = (x_1, \ldots, x_n).
\]

One easily verifies \( T \) is \( \mathcal{M}/\mathcal{M}_i \otimes \mathcal{M}^i \) measurable (use Corollary 9.19 repeatedly) and that \( \pi \circ T^{-1} = \mu_i \otimes \mu^i \) (see Exercise 6.8).

7. Let \( f \in L^1(\pi) \). Combining step 6. with the abstract change of variables Theorem (Exercise 10.11) implies

\[
\int_X f d\pi = \int_{X_i \times X^i} (f \circ T) d(\mu_i \otimes \mu^i).
\]

By Theorem 12.9 we also have

\[
\int_{X_i \times X^i} (f \circ T) d(\mu_i \otimes \mu^i) = \int_{X_i} \, d\mu_i(x_i) \int_{X^i} f(x_i, x^i) = \int_{X_i} \, d\mu_i(x_i) f(x_1, \ldots, x_n).
\]

Then by the induction hypothesis,

\[
\int_{X_i} \, d\mu_i(x_i) \int_{X^i} d\mu_i(x^i) f(x_1, \ldots, x_n) = \prod_{j \neq i} \int_{X_i} \, d\mu_j(x_j) \int_{X^i} d\mu_i(x_i) f(x_1, \ldots, x_n)
\]

(12.22)

where the ordering the integrals in the last product are inconsequential. Combining Eqs. (12.20) – (12.22) 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 12.11.** In this example we will show

\[
\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} \, dx = \frac{\pi}{2}.
\]

(12.23)

To see this write \( \frac{1}{x} = \int_0^\infty e^{-tx} \, dt \) and use Fubini-Tonelli to conclude that

\[
\int_0^M \frac{\sin x}{x} \, dx = \int_0^\infty \left[ \int_0^M e^{-tx} \sin x \, dt \right] \, dx
\]

\[
= \int_0^\infty \left[ \int_0^M e^{-tx} \sin x \, dx \right] \, dt
\]

\[
= \int_0^\infty \frac{1}{1 + t^2} \left( 1 - e^{-Mt} \sin M - t e^{-Mt} \cos M \right) \, dt
\]

\[
\to \int_0^\infty \frac{1}{1 + t^2} \, dt = \frac{\pi}{2} \text{ as } M \to \infty,
\]

wherein we have used the dominated convergence theorem (for instance, take \( g(t) := \frac{1}{1 + t^2} (1 + t e^{-t} + e^{-t}) \) to pass to the limit.

The next example is a refinement of this result.

**Example 12.12.** We have

\[
\int_0^\infty \frac{\sin x}{x} \, e^{-Ax} \, dx = \frac{1}{2} \pi - \arctan A \text{ for all } A > 0
\]

(12.24)

and for \( A, M \in [0, \infty) \),

\[
\left| \int_0^M \frac{\sin x}{x} \, e^{-Ax} \, dx - \frac{1}{2} \pi + \arctan A \right| \leq C e^{-MA} \frac{M}{M}
\]

(12.25)

where \( C = \max_{x \geq 0} \frac{1 + e^{-x}}{1 + e^{-x}} = \frac{1}{2 \sqrt{2} - 2} \approx 1.2 \). In particular Eq. (12.23) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

\[
|\sin x| = \int_0^x \cos y \, dy \leq \int_0^x |\cos y| \, dy \leq \int_0^x 1 \, dy = |x|
\]

so \( |\sin x| \leq 1 \) for all \( x \neq 0 \). Making use of the identity

\[
\int_0^\infty e^{-tx} \, dt = 1/x
\]

and Fubini’s theorem,
\[
\int_0^M \sin x x e^{-Ax} dx = \int_0^M dx \sin x e^{-Ax} \int_0^\infty e^{-tx} dt \\
= \int_0^\infty dt \int_0^M dx \sin x e^{-(A+t)x} \\
= \int_0^\infty \frac{1 - \cos M + (A + t) \sin M e^{-M(A+t)}}{(A + t)^2 + 1} dt \\
= \int_0^\infty \frac{1}{(A + t)^2 + 1} dt - \int_0^\infty \cos M + (A + t) \sin M e^{-M(A+t)} dt \\
= \frac{1}{2} \pi - \arctan A - \varepsilon(M,A) \\
\tag{12.26}
\]

where

\[
\varepsilon(M,A) = \int_0^\infty \frac{\cos M + (A + t) \sin M}{(A + t)^2 + 1} e^{-M(A+t)} dt.
\]

Since

\[
\left| \frac{\cos M + (A + t) \sin M}{(A + t)^2 + 1} \right| \leq \frac{1 + (A + t)}{(A + t)^2 + 1} \leq C,
\]

\[
|\varepsilon(M,A)| \leq \int_0^\infty e^{-M(A+t)} dt = C \frac{e^{-MA}}{M}.
\]

This estimate along with Eq. \ref{12.26} proves Eq. \ref{12.25} from which Eq. \ref{12.23} follows by taking \( A \to \infty \) and Eq. \ref{12.24} follows (using the dominated convergence theorem again) by letting \( M \to \infty \).

**Lemma 12.13.** Suppose that \( X \) is a random variable and \( \varphi : \mathbb{R} \to \mathbb{R} \) is a \( C^1 \) function such that \( \lim_{x \to -\infty} \varphi(x) = 0 \) and either \( \varphi'(x) \geq 0 \) for all \( x \) or \( \int_{-\infty}^0 |\varphi'(x)| dx < \infty \). Then

\[
\mathbb{E}[\varphi(X)] = \int_{-\infty}^\infty \varphi'(y) P(X > y) dy.
\]

Similarly if \( X \geq 0 \) and \( \varphi : [0, \infty) \to \mathbb{R} \) is a \( C^1 \) function such that \( \varphi(0) = 0 \) and either \( \varphi' \geq 0 \) 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 \( x \in \mathbb{R} \),

\[
\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y) dy.
\]

Under the stated assumptions on \( \varphi \), we may use either the monotone or the dominated convergence theorem to let \( M \to \infty \) in Eq. \ref{12.27} to find,

\[
\varphi(x) = \int_{-\infty}^x \varphi'(y) dy = \int_{-\infty}^1 1_{y < x} \varphi'(y) dy \text{ for all } x \in \mathbb{R}.
\]

Therefore,

\[
\mathbb{E}[\varphi(X)] = \mathbb{E} \left[ \int_{-\infty}^1 1_{y < X} \varphi'(y) dy \right] = \int_{-\infty}^1 \mathbb{E}[1_{y < X}] \varphi'(y) dy = \int_{-\infty}^\infty \varphi'(y) P(X > y) dy,
\]

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 12.14.** Here are a couple of examples involving Lemma 12.13.

1. Suppose \( X \) is a random variable, then

\[
\mathbb{E}[e^X] = \int_{-\infty}^\infty P(X > y) e^y dy = \int_0^\infty P(X > \ln u) du, \tag{12.28}
\]

where we made the change of variables, \( u = e^y \), to get the second equality.

2. If \( X \geq 0 \) and \( p \geq 1 \), then

\[
\mathbb{E} X^p = p \int_0^\infty y^{p-1} P(X > y) dy. \tag{12.29}
\]

### 12.4 Fubini’s Theorem and Completions*

**Notation 12.15** 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

\[ Ey := \{ x \in X : (x,y) \in E \}. \]

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 12.16.** Suppose \( (X,M,\mu) \) and \( (Y,N,\nu) \) are complete \( \sigma \)-finite measure spaces. Let \( (X \times Y,L,\lambda) \) be the completion of \( (X \times Y,M \otimes N,\mu \otimes \nu) \). If \( f \) is \( L \)-measurable and (a) \( f \geq 0 \) or (b) \( f \in L^1(\lambda) \) then \( f_x \) is \( N \)-measurable for \( \mu \) a.e. \( x \) and \( f^y \) is \( M \)-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

\[
\int_{X \times Y} f \, d\lambda = \int_Y \int_X f \, d\mu \, d\nu = \int_X \int_Y f \, d\nu \, d\mu.
\]

**Proof.** If \( E \in \mathcal{M} \otimes \mathcal{N} \) is a \( \mu \otimes \nu \) null set (i.e. \( (\mu \otimes \nu)(E) = 0 \)), then

\[
0 = (\mu \otimes \nu)(E) = \int_X \nu(x)E(x) \, d\mu(x) = \int_X \mu(E_y) \, d\nu(y).
\]

This shows that

\[
\mu \{ x : \nu(x) \neq 0 \} = 0 \quad \text{and} \quad \nu \{ y : \mu(E_y) \neq 0 \} = 0,
\]

i.e. \( \nu(x) = 0 \) for \( \mu \) a.e. \( x \) and \( \mu(E_y) = 0 \) for \( \nu \) a.e. \( y \). If \( h \) is \( L \) measurable and \( h = 0 \) for \( \lambda \)-a.e., then there exists \( E \in \mathcal{M} \otimes \mathcal{N} \) such that \( \{ (x, y) : h(x, y) \neq 0 \} \subseteq E \) and \( (\mu \otimes \nu)(E) = 0 \). Therefore \( |h(x, y)| \leq 1_E(x, y) \) and \( (\mu \otimes \nu)(E) = 0 \).

Since

\[
\{ h_x \neq 0 \} = \{ y \in Y : h(x, y) \neq 0 \} \subseteq \mathit{\lambda}E \quad \text{and} \quad \{ h_y \neq 0 \} = \{ x \in X : h(x, y) \neq 0 \} \subseteq \mathit{\lambda}E_y
\]

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 \( \int_X h(x, y) \, d\nu(y) \) exists and equals \( 0 \) for \( \mu \) a.e. \( x \) and similarly that \( \int_Y h(x, y) \, d\mu(x) \) exists and equals \( 0 \) for \( \nu \) a.e. \( y \). Therefore

\[
0 = \int_{X \times Y} hd\lambda = \int_Y \left( \int_X hd\mu \right) \, d\nu = \int_X \left( \int_Y hd\nu \right) \, d\mu.
\]

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 12.6, \( f = g + h \) 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

\[
\int_X f(x, y) \, d\mu(x) = \int_X g(x, y) \, d\mu(x).
\]

**12.5 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 12.17** Let

\[
m^d := m \otimes \cdots \otimes m \quad \text{on} \quad \mathcal{B}_{\mathbb{R}^d} = \mathcal{B}_\mathbb{R} \otimes \cdots \otimes \mathcal{B}_\mathbb{R}
\]

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 12.18.** A function \( f : \mathbb{R}^d \to \mathbb{R} \) is Lebesgue measurable if \( f^{-1}(\mathcal{B}_\mathbb{R}) \subseteq \mathcal{L}_d \).

**Notation 12.19** 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 12.20. 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 $J_i \in \mathcal{B}_R$ 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 $\mathcal{B}_R$ we find that
\[
m^d(x + A) = m(x_1 + J_1) \cdots m(x_d + J_d) = m(J_1) \cdots m(J_d) = m^d(A)
\]
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 12.1. □

Exercise 12.1. 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 subvectors $V \subset H$.

Theorem 12.21 (Change of Variables Theorem). Let $\Omega \subset_0 \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_0 \mathbb{R}^d$ be a $C^1$ - diffeomorphism.\footnote{That is $T : \Omega \to T(\Omega) \subset_0 \mathbb{R}^d$ is a continuously differentiable bijection and the inverse map $T^{-1} : T(\Omega) \to \Omega$ is also continuously differentiable.} See Figure 12.1. Then for any Borel measurable function, $f : T(\Omega) \to [0, \infty],$
\[
\int_\Omega f(T(x)) \left| \det T'(x) \right| dx = \int_{T(\Omega)} f(y) dy,
\]
where $T'(x)$ is the linear transformation on $\mathbb{R}^d$ defined by $T'(x)v := \left. \frac{d}{dt} \right|_{t=0} T(x + tv)$. More explicitly, viewing vectors in $\mathbb{R}^d$ as columns, $T'(x)$ may be represented by the matrix
\[
T'(x) = \begin{bmatrix}
\partial_1 T_1(x) & \cdots & \partial_d T_1(x) \\
\vdots & \ddots & \vdots \\
\partial_1 T_d(x) & \cdots & \partial_d T_d(x)
\end{bmatrix},
\]
i.e. the $i$ - $j$ - matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \ldots, T_d(x))^T$ and $\partial_i = \partial/\partial x_i$,

Remark 12.22. Theorem 12.21 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = \left| \det T'(x) \right| dx$. 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 10.12. 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 $\det T' = \left| T' \right|$ and
\[
\int_{[a,b]} 1_{T((a,\beta])} (T(x)) \left| T'(x) \right| dx = \int_{[a,b]} 1_{(\alpha,\beta)}(x) \left| T'(x) \right| dx = \int_\alpha^\beta \left| T'(x) \right| dx.
\]
If $T'(x) > 0$ on $[a,b]$, then
\[
\int_a^b \left| T'(x) \right| dx = \int_a^\beta T'(x) dx = T(\beta) - T(\alpha)
\]
\[
= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y) dy
\]
while if $T'(x) < 0$ on $[a,b]$, then

![Fig. 12.1. The geometric setup of Theorem 12.21](image-url)
\[
\int_{\alpha}^{\beta} |T'(x)| \, dx = -\int_{\alpha}^{\beta} T'(x) \, dx = T(\alpha) - T(\beta)
\]

\[
= m(T([\alpha, \beta])) = \int_{T([\alpha, \beta])} 1_{T([\alpha, \beta])} (y) \, dy.
\]

Combining the previous three equations shows

\[
\int_{[a,b]} f(T(x)) |T'(x)| \, dx = \int_{T([a,b])} f(y) \, dy
\]

whenever \( f \) is of the form \( f = 1_{T([\alpha, \beta])} \) with \( a < \alpha < \beta < b \). An application of Dynkin's multiplicative system Theorem \( 11.20 \) then implies that Eq. \( 12.32 \) 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 \( \Omega = \bigcup_{n=1}^{N} (a_n, b_n) \) where \( a_n, b_n \in \mathbb{R} \cup \{\pm \infty\} \) for \( n = 1, 2, \ldots, N \) 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

\[
\int_{\Omega} 1_{(a_n, b_n)} \cdot (f \circ T) \cdot |T'| \, dm = \lim_{k \to \infty} \int_{\Omega} 1_{T([\alpha_k, \beta_k])} \cdot f \circ T \cdot |T'| \, dm
\]

\[
= \int_{T(\Omega)} 1_{T([\alpha_k, \beta_k])} \cdot f \, dm
\]

Summing this equality on \( n \), then shows Eq. \( 12.30 \) 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. \( 12.31 \).

**Case 1.** Suppose \( T(x) \) has the form

\[
T(x) = (x_1, T_2(x), \ldots, T_d(x))
\]

or

\[
T(x) = (T_1(x), \ldots, T_{d-1}(x), x_i)
\]

for some \( i \in \{1, \ldots, d\} \). For definiteness we will assume \( T \) is as in Eq. \( 12.33 \), the case of \( T \) in Eq. \( 12.34 \) 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

\[
i_t(w) := (w_1, \ldots, w_{i-1}, t, w_{i+1}, \ldots, w_{d-1})
\]

\( \Omega \) be the (possibly empty) open subset of \( \mathbb{R}^{d-1} \) defined by

\[
\Omega := \{w \in \mathbb{R}^{d-1} : (w_1, \ldots, w_{i-1}, t, w_{i+1}, \ldots, w_{d-1}) \in \Omega \}
\]

and \( T_1 : \Omega \to \mathbb{R}^{d-1} \) be defined by

\[
T_1(w) = (T_2(w), \ldots, T_d(w))
\]

see Figure 12.2 Expanding \( \det T'(w_i) \) along the first row of the matrix \( T'(w_i) \)

\[
\begin{pmatrix}
\omega_1 & \omega_2 & \omega_3
\end{pmatrix}
\]

Fig. 12.2. 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

\[
\det |T'(w_i)| = |\det T'_t(w)|
\]

Now by the Fubini-Tonelli Theorem and the induction hypothesis,
\[ \int f \circ T \, \det T' \, dm = \int_{\mathbb{R}^d} 1 \cdot f \circ T \, \det T' \, dm \]
\[ = \int \mathbb{1}_{\Omega} (w_t) (f \circ T) (w_t) \, |\det T' (w_t)| \, dw \, dt \]
\[ = \int_{\mathbb{R}^d} \left[ \int (f \circ T) (w_t) \, |\det T' (w_t)| \, dw \right] \, dt \]
\[ = \int_{\mathbb{R}^d} \left[ \int f (t, T_t (w)) \, |\det T_t' (w)| \, dw \right] \, dt \]
\[ = \int \left[ \int_{T_t (\Omega_t)} f (t, z) \, dz \right] \, dt = \int \left[ \int_{\mathbb{R}^{d-1}} 1_{T (\Omega)} (t, z) f (t, z) \, dz \right] \, dt \]
\[ = \int \frac{f (y)}{T (\Omega)} \, dy \]

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. (12.30) is true locally.) Suppose that \( T : \Omega \rightarrow \mathbb{R}^d \) 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

\[ \int_{W} f \circ T \, \det T' \, dm = \int_{T (W)} f \, dm \]

holds for all Borel measurable function, \( f : T(W) \rightarrow [0, \infty] \). 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 \( i^{th} - \) 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,

\[ S (x) := (x_1, T_2 (x), \ldots, T_d (x)). \quad (12.35) \]

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 \( W \subset \Omega \) and \( S (W) \subset \mathbb{R}^d \)

and \( S : W \rightarrow S (W) \) is a \( C^1 \) - diffeomorphism. Let \( R : S (W) \rightarrow T (W) \subset \mathbb{R}^d \)
to be the \( C^1 \) - diffeomorphism defined by

\[ R(z) := T \circ S^{-1} (z) \] for all \( z \in S (W) \).

Because

\[ (T_1 (x), \ldots, T_d (x)) = T (x) = R (S (x)) = R ((x_1, T_2 (x), \ldots, T_d (x))) \]

for all \( x \in W \), if

\[ (z_1, z_2, \ldots, z_d) = S (x) = (x_1, T_2 (x), \ldots, T_d (x)) \]

then

\[ R (z) = (T_1 (S^{-1} (z)), z_2, \ldots, z_d). \quad (12.36) \]

Observe that \( S \) is a map of the form in Eq. (12.33). \( R \) is a map of the form in Eq. (12.34), \( T' (x) = R' (S (x)) S' (x) \) (by the chain rule) and (by the multiplicative property of the determinant)

\[ |\det T' (x)| = |\det R' (S (x))| \cdot |\det S' (x)| \quad \forall \, x \in W. \]

So if \( f : T(W) \rightarrow [0, \infty] \) is a Borel measurable function, two applications of the results in Case 1. shows,

\[ \int_{W} f \circ T \cdot |\det T'| \, dm = \int_{W} (f \circ R \cdot |\det R'|) \circ S \cdot |\det S'| \, dm \]
\[ = \int_{S(W)} f \circ R \cdot |\det R'| \, dm = \int_{R(S(W))} f \, dm \]
\[ = \int_{T(W)} f \, dm \]

and Case 2. is proved.

**Case 3.** (General Case.) Let \( f : \Omega \rightarrow [0, \infty] \) be a general non-negative Borel measurable function and let

\[ \Omega_n := \{ x \in \Omega : \text{dist}(x, \Omega^c) \geq 1/n \quad \text{and} \quad |x| \leq n \}. \]

Then each \( \Omega_n \) is a compact subset of \( \Omega \) and \( \Omega_n \uparrow \Omega \) as \( n \rightarrow \infty \). Using the compactness of \( \Omega_n \) and case 2, for each \( n \in \mathbb{N} \), there is a finite open cover \( \mathcal{W}_n \) of \( \Omega_n \) such that \( W \subset \Omega \) and Eq. (12.30) holds with \( \Omega \) replaced by \( W \) for each \( W \in \mathcal{W}_n \). Let \( \{ W_i \}_{i=1}^{\infty} \) be an enumeration of \( \bigcup_{n=1}^{\infty} \mathcal{W}_n \) and set \( W_1 = W = W_1 \) and \( W_i = W_i \setminus (W_1 \cup \cdots \cup W_{i-1}) \) for all \( i \geq 2 \). Then \( \Omega = \bigcup_{i=1}^{\infty} W_i \) and by repeated use of case 2.,
Then by the chain rule and the fundamental theorem of calculus,

\[ \int f \circ T \, \det T' \, dm = \sum_{i=1}^{\infty} \int_{W_i} f \circ T \cdot |\det T'| \, dm \]

\[ = \sum_{i=1}^{\infty} \int_{T(W_i)} \left( (T \circ W_i) f \right) \circ T \cdot |\det T'| \, dm \]

\[ = \sum_{i=1}^{\infty} \int_{T(W_i)} 1_{T(W_i)} \cdot f \, dm = \sum_{i=1}^{\infty} \int_{T(W_i)} 1_{T(W_i)} \cdot f \, dm \]

\[ = \int_{T(\Omega)} f \, dm. \]

\[ \square \]

**Remark 12.23.** When \( d = 1 \), one often learns the change of variables formula as

\[ \int_{a}^{b} f(T(x)) \, T'(x) \, dx = \int_{T(a)}^{T(b)} f(y) \, dy \]  \hspace{1cm} \hspace{1cm} (12.37) \]

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. (12.37) is equivalent to Eq. (12.30) with \( \Omega = (a, b) \). On the other hand if \( T' < 0 \) on \((a, b)\) then \( T((a, b)) = (T(b), T(a)) \) and Eq. (12.37) is equivalent to

\[ \int_{T(a)}^{T(b)} f(y) \, dy = - \int_{T(b)}^{T(a)} f(y) \, dy \]

which is again implies Eq. (12.30). On the other hand Eq. (12.37) is more general than Eq. (12.30) since it does not require \( T \) to be injective. The standard proof of Eq. (12.37) 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,

\[ \int_{a}^{b} f(T(x)) \, T'(x) \, dx = \int_{a}^{b} F'(T(x)) \, T'(x) \, dx = \int_{a}^{b} \frac{d}{dx} [F(T(x))] \, dx \]

\[ = F(T(x)) \bigg|_{a}^{b} = \int_{T(a)}^{T(b)} f(y) \, dy. \]

An application of Dynkin’s multiplicative systems theorem now shows that Eq. (12.37) 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 \((a, b)\).

### 12.5 Lebesgue Measure on \( \mathbb{R}^d \) and the Change of Variables Theorem

**Exercise 12.2.** Continuing the setup in Theorem 12.21 show that \( f \in L^1(\Omega, dm) \) iff

\[ \int_{\Omega} |f \circ T| \, |\det T'| \, dm < \infty \]

and if \( f \in L^1(\Omega, dm) \), then Eq. (12.30) holds.

**Example 12.24.** Continuing the setup in Theorem 12.21 if \( A \in \mathcal{B}_d \), then

\[ m(T(A)) = \int_{\mathbb{R}^d} 1_{T(A)}(y) \, dy = \int_{\mathbb{R}^d} 1_{T(A)}(T(x)) |\det T'(x)| \, dx \]

\[ = \int_{\mathbb{R}^d} 1_{A}(x) \, |\det T'(x)| \, dx \]

wherein the second equality we have made the change of variables, \( y = T(x) \). Hence we have shown

\[ d(m \circ T) = |\det T'| \, dm. \]

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.

\[ m(T(A)) = |\det T| m(A) \text{ for all } A \in \mathcal{B}_d. \]  \hspace{1cm} \hspace{1cm} (12.38)

This equation also shows that \( m \circ T \) and \( m \) have the same null sets and hence the equality in Eq. (12.38) 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. \( T^T T = I \)), then \( |\det T| = \pm 1 \) and hence \( m \) is invariant under all rotations. This is not obvious from the definition of \( m^d \) as a product measure!

**Example 12.25.** 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 12.26 (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,
Example 12.28. In this example, we will evaluate the integral

\[
T'(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}
\]

and therefore

\[
dx = |\det T'(r, \theta)| dr d\theta = r dr d\theta.
\]

Observing that \(R^2 \setminus T((0, \infty) \times (0, 2\pi)) = \ell := \{(x, 0) : x \geq 0\}\) has \(m^2 - \) measure zero, it follows from the change of variables Theorem 12.21 that

\[
\int_{R^2} f(x) dx = \int_{0}^{2\pi} d\theta \int_{0}^{\infty} dr \, r \cdot f(r \cos \theta, \sin \theta)
\]

(12.39)

for any Borel measurable function \(f : R^2 \rightarrow [0, \infty]\).

Example 12.27 (Holomorphic Change of Variables). Suppose that \(f : \Omega \subset C \cong R^2 \rightarrow C\) 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

\[
dudv = \left| \det \begin{bmatrix} U_y & U_x \\ V_y & V_x \end{bmatrix} \right| dx dy = |U_x V_y - U_y V_x| dx dy.
\]

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

\[
U_x V_y - U_y V_x = U_x^2 + V_y^2 = |f'|^2.
\]

Therefore

\[
dudv = |f'(x + iy)|^2 dx dy.
\]

Example 12.28. In this example, we will evaluate the integral

\[
I := \iint_{\Omega} (x^4 - y^4) dx dy
\]

where

\[
\Omega = \{(x, y) : 1 < x^2 - y^2 < 2, 0 < xy < 1\},
\]

see Figure 12.3. We are going to do this by making the change of variables,

\[
(u, v) := T(x, y) = (x^2 - y^2, xy),
\]

in which case

\[
dudv = \left| \det \begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix} \right| dx dy = 2 (x^2 + y^2) dx dy
\]

Notice that

\[
(x^4 - y^4) = (x^2 - y^2)(x^2 + y^2) = u (x^2 + y^2) = \frac{1}{2} u dudv.
\]

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 \(\Omega = \pm D \cup D\) and \(T(\pm D) = (1, 2) \times (0, 1)\). The change of variables theorem then implies

\[
I_{\pm} := \iint_{\pm D} (x^4 - y^4) dx dy = \frac{1}{2} \iint_{(1, 2) \times (0, 1)} u dudv = \frac{1}{2} \frac{u^2}{2} \cdot 1 = \frac{3}{4}
\]

and therefore

\[
I = I_+ + I_- = 2 \cdot (3/4) = 3/2.
\]

Exercise 12.3 (Spherical Coordinates). Let \(T : (0, \infty) \times (0, \pi) \times (0, 2\pi) \rightarrow R^3\) be defined by

\[
T(r, \varphi, \theta) = (r \sin \varphi \cos \theta, r \sin \varphi \sin \theta, r \cos \varphi)
\]

\[
= r (\sin \varphi \cos \theta, \sin \varphi \sin \theta, \cos \varphi),
\]

see Figure 12.4. By making the change of variables \(x = T(r, \varphi, \theta)\), show
12.6 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 12.30. If $T$ is a real $d \times d$ matrix, then there exists $D = \text{diag}(\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$ and two orthogonal matrices $R$ and $S$ such that $T = RDS^\text{tr}$.

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 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$ such that $T^*u_j = \lambda_j^2 u_j$. Let $1 \leq k \leq d$ be the first index such that $\lambda_{k+1} = 0$ wherein by convention $\lambda_{d+1} = 0$. Then $\{v_j := \frac{1}{\lambda_j} T u_j\}_{j=1}^k$ 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 $D = \text{diag}(\lambda_1, \ldots, \lambda_d)$ and let $R$ and $S$ be the orthogonal matrices defined by

$$Re_j = v_j \text{ and } Se_j = u_j \text{ for all } 1 \leq j \leq d.$$ 

We then have

$$TSe_j = Tu_j = \lambda_j v_j = \lambda_j Re_j = R\lambda_j e_j = RDe_j \text{ for all } 1 \leq j \leq d$$

and hence $TS = RD$, i.e. $T = RDS^\text{tr}$. \hfill $\blacksquare$

Theorem 12.31. 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 $T, m$ are still translation invariant measures which are $|\det T|$ times each other.

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 $D = \text{diag}(\lambda_1, \ldots, \lambda_d)$ with $\lambda_i \geq 0$, then $D[0, 1]^d = [0, \lambda_1] \times \cdots \times [0, \lambda_d]$ so that

$$k = km\left([0, 1]^d\right) = m\left(D[0, 1]^d\right) = \lambda_1 \cdots \lambda_d = \det D.$$ 

3. For the general case we use Lemma 12.30 to write $T = RDS^\text{tr}$ and then find,

$$mT = mRDS^\text{tr} = mDS^\text{tr} = \det(D)mS^\text{tr} = \det(D)m = |\det T| m.$$ 

\hfill $\blacksquare$
Theorem 12.32 (Linear Change of Variables Theorem). If $T \in \text{GL}(d, \mathbb{R}) = \text{GL}(\mathbb{R}^d)$ – the space of $d \times d$ invertible matrices, then the change of variables formula,

$$\int_{\mathbb{R}^d} f(y)dy = \int_{\mathbb{R}^d} \left|\det T\right| f \circ T(x)dx,$$  \hspace{1cm} (12.41)

holds for all Riemann integrable functions $f : \mathbb{R}^d \to \mathbb{R}$.

Proof. As usual it suffices to show Eq. (12.41) holds when $f = 1_A$ for some $A \in \mathcal{B}_{\mathbb{R}^d}$. In this case,

$$\int_{\mathbb{R}^d} f \circ T(x)dx = \int_{\mathbb{R}^d} 1_A \circ T(x)dx = \int_{\mathbb{R}^d} 1_{T^{-1}(A)}(x)dx = m(T^{-1}(A)) = |\det T^{-1}| m(A)$$

$$= \frac{1}{|\det T|} \int_{\mathbb{R}^d} 1_A(x)dx = \frac{1}{|\det T|} \int_{\mathbb{R}^d} f(x)dx.$$  \hspace{1cm} (12.42)

The next exercise gives an alternative proof of Theorem 12.31.

Exercise 12.4 (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:

$$|\det T| \int_{\mathbb{R}^d} 1_R \circ T(x)dx = m(R) = \int_{\mathbb{R}^d} 1_R(y)dy$$  \hspace{1cm} (12.43)

for each of the following linear transformations:

1. Suppose that $i < k$ and

$$T(x_1, x_2, \ldots, x_d) = (x_1, \ldots, x_{i-1}, x_k, x_{i+1}, \ldots, x_{k-1}, x_i, x_{k+1}, \ldots, x_d),$$

i.e. $T$ swaps the $i$ and $k$ coordinates of $x$.

2. $T(x_1, \ldots, x_k, \ldots, x_d) = (x_1, \ldots, c x_k, \ldots, x_d)$ where $c \in \mathbb{R} \setminus \{0\}$.

3. $T(x_1, x_2, \ldots, x_d) = (x_1, \ldots, x_i + c x_k, \ldots, x_d)$ where $c \in \mathbb{R}$.

Hint: you should use Fubini’s theorem along with the one dimensional change of variables theorem.

12.6.1 $\delta$ – function localization proof of the change of variables

Theorem 12.21

The proof we give here comes from [31] who attributes the idea to Cornea. Recall that we are trying to prove Eq. (12.30) which states,

$$\int_{\Omega} f(T(x)) |\det T'(x)| dx = \int_{T(\Omega)} f(y) dy.$$  \hspace{1cm} (12.43)

The key heuristic 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

$$1 = \int_{T(\Omega)} \delta(y_0 - y) dy = \int_{\Omega} \delta(y_0 - T(x)) |\det T'(x)| dx.$$  \hspace{1cm} (12.43)

If we can verify Eq. (12.43), then multiplying this equation by $f(y_0)$ and integrating the result, would give,

$$\int_{T(\Omega)} f(y_0) dy_0 = \int_{\Omega} dx \int_{T(\Omega)} dy_0 f(y_0) \delta(y_0 - T(x)) |\det T'(x)|$$

$$= \int_{\Omega} dx \int_{T(\Omega)} dy_0 f(y_0) \delta(y_0 - T(x)) |\det T'(x)|$$

$$= \int_{\Omega} dx f(T(x)) |\det T'(x)|$$

which is the desired change of variables formula. So let us explain why Eq. (12.43) should be true.

As $y_0 - T(x) = 0$ iff $x = x_0 := T^{-1}(y_0)$, we should have

$$\int_{\Omega} \delta(y_0 - T(x)) |\det T'(x)| dx = \int_{\Omega} \delta(y_0 - T(x)) |\det T'(x)| dx$$

$$= |\det T'(x_0)| \int_{\Omega} \delta(y_0 - T(x)) dx$$

and so Eq. (12.43) is equivalent to showing

$$\int_{\Omega} \delta(y_0 - T(x)) dx = \frac{1}{|\det T'(x_0)|} \forall x_0 \in \Omega.$$  \hspace{1cm} (12.44)

There are some mistakes in the arguments given in this reference which we have taken the opportunity to correct in the exposition below.

If you are uncomfortable with the heuristic discussion to follows you may skip it and jump directly to Proposition 12.33.
The “reasons” to expect Eq. (12.44) is correct; 1) 
\[ T(x) \sim T(x_0) + T'(x_0)(x - x_0) = y_0 + T'(x_0)(x - x_0) \]
for \( x \) near \( x_0 \), 2) \( \delta(y_0 - T(x)) \) is supported on 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
\[
\int \delta(y_0 - T(x)) \, dx = \int \delta(T'(x_0)(x - x_0)) \, dx = \int_{\mathbb{R}^d} \delta(T'(x_0)w) \, dw
\]
\[
= \frac{1}{|\det[T'(x_0)]|} \int_{\mathbb{R}^d} \delta(z) \, dz = \frac{1}{|\det[T'(x_0)]|}.
\]

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 \( \text{supp} \, \delta_1 \subset B \) - 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. (12.44).

**Proposition 12.33.** If for each \( r > 0 \), \( J_r : T(\Omega) \to [0, \infty] \) be defined by
\[
J_r(y) := \int_{\Omega} \delta_r(T(x) - y) \, dx,
\]
then \( \lim_{r \downarrow 0} J_r(y) = 1/|\det[T'(T^{-1}(y))]| \) locally uniformly in \( y \in \Omega \). In more detail we are claiming to any compact subset, \( K \subset T(\Omega) \),
\[
\lim_{r \downarrow 0} \sup_{y \in K} \left| J_r(y) - \frac{1}{|\det[T'(T^{-1}(y))]|} \right| = 0.
\]

We will give the proof of this result a little later in this section after Lemma 12.34. First let us see how to use Propositions 12.33 to prove Theorem 12.21.

**Proof of Theorem 12.21.** Let \( f \in C_c(T(\Omega), [0, \infty)) \) and \( K := \text{supp}(f) \). Then by Tonelli’s theorem,
\[
\int_{T(\Omega)} f(y) J_r(y) \, dy = \int_{T(\Omega)} f(x) \, d\delta_r(T(x) - y) \, f(y) = \int_{\Omega} f \ast \delta_r(T(x)) \, dx.
\]
By Propositions 12.33 and dominated convergence theorem,
\[
\lim_{r \downarrow 0} \int_{T(\Omega)} f(y) J_r(y) \, dy = \int_{T(\Omega)} f(y) \, dy.
\]
By simple approximate \( \delta \) - function arguments we know \( f \ast \delta_r \rightarrow f \) uniformly on \( \Omega \) and moreover \( \text{supp}(f \ast \delta_r) \subset K_r := \{ x \in \mathbb{R}^d : d_K(x) \leq r \} \).

Hence if we let
\[
\rho = \min \left( \frac{1}{2} \text{dist}(K, T(\Omega)), 1 \right),
\]
then for all \( 0 < r \leq \rho \), \( \text{supp} \left( (f \ast \delta_r) \circ T \right) \subset T^{-1}(K_{\rho}) \) (a compact subset of \( \Omega \)) and \( (f \ast \delta_r) \circ T \rightarrow f \circ T \) uniformly as \( r \downarrow 0 \). Thus again by Dominated convergence theorem we conclude
\[
\lim_{r \downarrow 0} \int_{T(\Omega)} f(y) \, dy = \int_{T(\Omega)} f(T(x)) \, dx.
\]
Combining Eqs. (12.47), (12.48), and (12.50) implies
\[
\int_{\Omega} f(T(x)) \, dx = \int_{T(\Omega)} f(y) \frac{1}{|\det T'(T^{-1}(y))|} \, dy
\]
from which Theorem 12.21 follows upon replacing \( f(y) \) by \( f(y) \cdot |\det T'(T^{-1}(y))| \).

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 12.33.

**Lemma 12.34.** Let \( K \) be a compact subset of \( T(\Omega) \), \( 0 < \rho \leq 1 \) be as in Eq. (12.44), \( K_{\rho} := \{ x \in \mathbb{R}^d : d_K(x) \leq \rho \} \)
and
\[
M := \sup_{y \in K_{\rho}} \left\| (T^{-1})'(y) \right\|_{\text{op}} < \infty.
\]

Then
\[
\sup_{y \in K} \left\| T^{-1}(y+w) - T^{-1}(y) \right\| \leq M \|w\| \quad \forall w \in \rho B,
\]
\[
\inf_{y \in K} \left\| T'(T^{-1}(y))w \right\| \geq \frac{1}{M} \|w\| \quad \forall w \in \mathbb{R}^d,
\]
and \( \lim_{r \downarrow 0} \varepsilon_K(r) = 0 \) where
\[
\varepsilon_K(r) := \sup_{y \in K} \sup_{\|w\| \leq \rho M} \left\| \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y - T'(T^{-1}(y))w \right] \right\|_{\text{op}}.
\]

**Proof.** To simplify notation in the proof, let \( S := T^{-1} : T(\Omega) \to \Omega \). For \( y \in K \) and \( \|w\| \leq \rho \) we have by the fundamental theorem of calculus,
\[ S(y + w) - S(y) = \left[ \int_0^1 S'(y + sw) \, ds \right] w \]  \hspace{1cm} (12.55)

which along with standard estimates gives Eq. (12.52).

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

\[ \left| T'(S(y))^{-1} \right| = \| S'(y) \| \leq M \| z \| \quad \forall \, y \in K \text{ and } z \in \mathbb{R}^d. \]

Taking \( z = T'(S(y)) w \) in this inequality leads directly to Eq. (12.53).

Let \( \rho \in (0, \rho] \) be chosen so that \( M \rho < \text{dist} \left( S(K), \Omega' \right) \). In proving the last assertion we always assume that \( r \in (0, \rho) \). By the fundamental theorem of calculus, for \( y \in K \) and \( w \in rMB \), we have

\[ \frac{1}{r} \left[ T'(S(y) + rw) - y \right] = \left[ \frac{1}{r} \int_0^r T'((S(y) + \sigma w)) \, d\sigma \right] w - T'(S(y)) w \]

\[ = \frac{1}{r} \int_0^r \left[ T'((S(y) + \sigma w)) - T'(S(y)) \right] w \, d\sigma \]

\[ = \int_0^r \left[ T'((S(y) + \sigma w)) - T'(S(y)) \right] \frac{w}{r} \, d\sigma. \]

It is now a simple matter to use this identity and simple estimates to show

\[ \varepsilon_K(r) \leq \sup_{y \in K} \sup_{\|w\| \leq rM} \| T'((S(y) + w)) - T'(S(y)) \|_\infty \cdot M \]

which tends to 0 as \( r \downarrow 0 \) by uniform continuity of \( T' \) on

\[ \{ S(K) \}_{MB} = \left\{ x \in \mathbb{R}^d : d_{S(K)}(x) \leq M \rho \right\}. \]

We are now ready for the proof of Proposition 12.33.

**Proof of Proposition 12.33**

Let \( K \) be a compact subset of \( T(\Omega) \), \( 0 < \rho \leq 1 \) be as in Eq. (12.49), and let \( y \in K \) and \( 0 < r \leq \rho \) and let us continue to use the notation in Lemma 12.34. If \( x \in \Omega \) is such that \( \delta_r(T(x) - y) > 0 \), then \( \| T(x) - y \| \leq r \) or equivalently that \( T(x) \in y + rB \), i.e. \( x \in T^{-1}(y + rB) \). Now the estimate in Eq. (12.52) implies\(^6\)

\[ T^{-1}(y + rB) \subset T^{-1}(y) + rMB \quad \forall \, y \in K \text{ and } r \leq \rho \]

and hence \( J_r(y) \) defined in Eq. (12.45) may be described by;

\[ J_r(y) := \int_{T^{-1}(y) + rMB} \delta_r(T(x) - y) \, dx \quad \forall \, y \in K. \]

We now make the affine change of variables, \( x = T^{-1}(y) + rw \) in the above integral to find

\[ J_r(y) := \int_{w \in MB} \delta_1 \left( \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) \, dw \quad \forall \, y \in K. \]

Using the dominated convergence theorem along with the continuity of \( \delta_1 \) it follows that

\[ \lim_{r \downarrow 0} J_r(y) = \int_{MB} \lim_{r \downarrow 0} \delta_1 \left( \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) \, dw \]

\[ = \int_{MB} \delta_1 \left( \lim_{r \downarrow 0} \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) \, dw \]

\[ = \int_{MB} \delta_1 \left( T'(T^{-1}(y)) \right) \, dw. \hspace{1cm} (12.56) \]

According to Eq. (12.53), \( \| T'(T^{-1}(y)) \| > 1 \) if \( \|w\| > M \) and therefore,

\[ \int_{MB} \delta_1 \left( T'(T^{-1}(y)) \right) \, dw = \int_{\mathbb{R}^d} \delta_1 \left( T'(T^{-1}(y)) \right) \, dw \]

\[ = \int_{\mathbb{R}^d} \delta_1(z) \frac{1}{|\det T'(T^{-1}(y))|} \, dz \]

\[ = \frac{1}{|\det T'(T^{-1}(y))|} \text{ wherein we have made the linear change of variables, } z = T'(T^{-1}(y)) \, w \text{ in the second equality}\(^7\) and used } \int_{\mathbb{R}^d} \delta_1(z) \, dz = 1 \text{ in the last equality.}\]

\[ w \rightarrow \delta_1 \left( \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) = \delta_1(M^{-1}w) \]

which is supported in \( MB \) in general and not in \( B \).

\(^6\) At this point in \([31]\), it is implicitly asserted that \( w \rightarrow \delta_1 \left( \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) \) is supported in \( B \). This is however false in general. For example if \( M > 1 \) and \( T(x) = M^{-1}x \), then

\[ w \rightarrow \delta_1 \left( \frac{1}{r} \left[ T(T^{-1}(y) + rw) - y \right] \right) \]

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. (12.56) is uniform over \( y \in K \). However, for \( y \in K \),

\[
\left| J_r (y) - \int_{MB} \delta_1 \left( T' \left( T^{-1} (y) \right) \right) w \, dw \right|
\]

\[
= \int_{MB} \left| \delta_1 \left( \frac{1}{r} \left[ T \left( T^{-1} (y) + r \omega \right) \right] - y \right) \right| - \delta_1 \left( T' \left( T^{-1} (y) \right) \right) w \, dw
\]

\[
\leq \int_{MB} \left| \delta_1 \left( \frac{1}{r} \left[ T \left( T^{-1} (y) + r \omega \right) - y \right] \right) \right| - \delta_1 \left( T' \left( T^{-1} (y) \right) \right) w \, dw
\]

\[
\leq \| \nabla \delta_1 \|_\infty \cdot m \left( MB \right) \cdot \varepsilon_K (r) \to 0 \text{ as } r \downarrow 0
\]

wherein \( \varepsilon_K (r) \) is as in Eq. (12.54). The proof is complete since we have already seen in Lemma 12.34 that \( \lim_{r \downarrow 0} \varepsilon_K (r) = 0 \).

**12.6.2 Radon-Nykodym proof.**

**Remark 12.35 (A proof using the Radon–Nykodym theorem).** As usual let \( T : \Omega \to T (\Omega) \) be a \( C^1 \) – diffeomorphism 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

\[
\int_{T (\Omega)} f \circ T \, dm = \int_{\Omega} f (m \circ T^{-1}) \quad (12.57)
\]

and

\[
\int_{T (\Omega)} g \circ T^{-1} \, dm = \int_{\Omega} g (m \circ T) \cdot \quad (12.58)
\]

2. Referring to the math 140 notes, show \( |m \circ T (A)| \leq K m (A) \) and similarly \( |m \circ T^{-1} (A)| \leq K m (A) \). Therefore by the easiest version of the the Radon – Nykodym there are bounded non-negative functions, \( \alpha \) and \( \beta \) such that

\[
\frac{d (m \circ T)}{dm} = \alpha : \Omega \to \mathbb{R} \text{ and } \frac{d (m \circ T^{-1})}{dm} = \beta : T (\Omega) \to \mathbb{R}.
\]

In other words we now have,

\[
\int_{\Omega} f \circ Tdm = \int_{T (\Omega)} f \beta dm \text{ and } \int_{T (\Omega)} g \circ T^{-1} dm = \int_{\Omega} g \alpha dm.
\]

(12.59)

(12.60)

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

\[
\beta \circ T \alpha = 1 \text{ a.e. } \iff \alpha = \frac{1}{\beta \circ T}.
\]

(12.61)

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 \( \delta_r \ast \alpha \to \alpha \) in \( L^1_{\text{loc}} \) as \( r \downarrow 0 \) where \( \delta_r (x) := \frac{1}{m (B_r (0))} m (B_r (x)) \).

4. For \( x \in \Omega \) and \( y \in y \in B_r (x) = x + B_r (0) \),

\[
T (y) - T (x) = \int_0^1 T' (x + t (y - x)) (y - x) \, dt
\]

\[
= \left[ T' (x) + \int_0^1 [T' (x + t (y - x)) - T' (x)] \, dt \right] (y - x)
\]

\[
= T' (x) \left( I + \varepsilon (x, y) \right) (y - x)
\]

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 \( \lim_{r \to 0} \varepsilon_0 (r) = \varepsilon_0 (0) = 0 \).

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

\[
T (B_r (x)) \subset T (x) + T' (x) B_{r (1 + \varepsilon (r))} (0).
\]

(12.62)
5. We now give two arguments showing \( \alpha (x) = \frac{d(m \circ T)}{dm} (x) \leq |\det T'(x)| \) for a.e. \( x \in \Omega \).

a) From Eq. (12.62) it follows that
\[
\frac{m (T (B_r (x))))}{m (B_r (x))} \leq \frac{m (T (x) + T'(x) B_{r(1 + \varepsilon (r))} (0))}{m (B_r (x))}
= |\det T'(x)| \cdot (1 + \varepsilon (r))^d .
\]
Letting \( r \downarrow 0 \) and using the Lebesgue differentiation theorem (see Theorem ??) which is a rather deep result!

\[
\alpha (x) = \frac{d (m \circ T)}{dm} (x) \leq |\det T'(x)| \text{ for a.e. } x \in \Omega . \tag{12.63}
\]
Applying this result with \( T \) replaced by \( T^{-1} \) then shows,

\[
\beta (y) = \frac{d (m \circ T^{-1})}{dm} (y) \leq |\det (T^{-1})'| (y) \text{ for a.e. } y \in T (\Omega). \tag{12.64}
\]

b) Alternatively let us observe that
\[
\frac{m (T (B_r (x))))}{m (B_r (x))} = \frac{1}{m (B_r (x))} \int_{B_r (0)} \alpha (x + y) dy = \delta_r \ast \alpha (x).
\]
By the easier approximate identity Theorem ?? we know \( \delta_r \ast \alpha \rightarrow \alpha \) in \( L^1_{loc} \) and so there exists \( r_n \downarrow 0 \) such that \( \delta_{r_n} \ast \alpha \rightarrow \alpha \) a.e. as \( n \rightarrow \infty \). Thus we again learn that for a.e. \( x \),

\[
\alpha (x) = \lim_{n \rightarrow \infty} \delta_{r_n} \ast \alpha (x) = \lim_{n \rightarrow \infty} \frac{m (T (B_{r_n} (x))))}{m (B_{r_n} (x))} \leq \lim_{n \rightarrow \infty} |\det T'(x)| \cdot (1 + \varepsilon (r_n))^d = |\det T'(x)|.
\]

6. We now use \( \alpha = \frac{1}{\beta \circ T} \) from Eq. (12.61) along with the inequality in Eq. (12.64) to learn
\[
\alpha (x) = \frac{1}{\beta \circ T (x)} \geq \frac{1}{|\det (T^{-1})' (T (x))|}.
\]
On the other hand, since \( T^{-1} \circ T = I \), it follows by the chain rule that \( (T^{-1})' (T (x)) T'(x) = I \) and therefore
\[
\frac{1}{|\det (T^{-1})' (T (x))|} = |\det T'(x)|
\]
and we may conclude \( \alpha (x) \geq |\det T'(x)| \). This result along with the inequality in Eq. (12.63) shows
\[
\alpha (x) = |\det T'(x)|.
\]

7. Using this result back in Eq. (12.60) with \( g = f \circ T \) for some function on \( f : T (\Omega) \rightarrow \mathbb{R} \) 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. (12.62):
\[
T (x) + T'(x) B_{r(1 + \varepsilon (r))} (0) \subset T (B_r (x)) \subset T (x) + T'(x) B_{r(1 + \varepsilon (r))} (0).
\]
If we had done this we could have avoided discussing \( \beta \) altogether. Indeed, the method of step 5. would then give
\[
|\det T'(x)| \cdot (1 + \varepsilon (r))^d \leq \frac{m (T (B_r (x))))}{m (B_r (x))} \leq |\det T'(x)| \cdot (1 + \varepsilon (r))^d
\]
which upon letting \( r \downarrow 0 \) would have shown \( \frac{d |m \circ T|}{dm} (x) = |\det T'(x)| \).

### 12.7 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, \( S_{d-1} \) and \( \Phi : \mathbb{R}^d \setminus \{ 0 \} \rightarrow (0, \infty) \times S^{d-1} \) be defined by \( \Phi (x) := (|x|, |x|^{-1} x) \). The inverse map, \( \Phi^{-1} : (0, \infty) \times S^{d-1} \rightarrow \mathbb{R}^d \setminus \{ 0 \} \), is given by \( \Phi^{-1} (r, \omega) = r \omega \). Since \( \Phi \) and \( \Phi^{-1} \) are continuous, they are both Borel measurable. For \( E \in S_{d-1} \) and \( a > 0 \), let
\[
E_a := \{ r \omega : r \in (0, a] \text{ and } \omega \in E \} = \Phi^{-1}((0, a] \times E) \in B_{\mathbb{R}^d}.
\]

**Definition 12.36.** For \( E \in S_{d-1} \), 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 \( E \in S_{d-1} \), then \( E_1 = \Phi^{-1}((0, 1] \times E) \in B_{\mathbb{R}^d} \) 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
Therefore we expect the area of $f$ measurable function then below. On the other hand

$$
\int_{(1,1+\varepsilon)\cdot E} = \{r\omega : r \in (1,1+\varepsilon) \text{ and } \omega \in E\}
$$

should be approximately given by $m\left((1,1+\varepsilon)\cdot E\right) \cong \sigma(E)\varepsilon$, see Figure 12.5 below. On the other hand

$$
m((1,1+\varepsilon)\cdot E) = m(E_{1+\varepsilon} \setminus E_1) = \{(1+\varepsilon)^d - 1\} \cdot m(E_1).
$$

Therefore we expect the area of $E$ should be given by

$$
\sigma(E) = \lim_{\varepsilon \downarrow 0} \frac{(1+\varepsilon)^d - 1}{\varepsilon} \cdot m(E_1) = d \cdot m(E_1).
$$

The following theorem is motivated by Example 12.26 and Exercise 12.3.

**Theorem 12.37 (Polar Coordinates).** If $f : \mathbb{R}^d \to [0,\infty]$ is a $(B_{\mathbb{R}^d},\mathcal{B})$-measurable function then

$$
\int_{\mathbb{R}^d} f(x)dm(x) = \int_{(0,\infty) \times S^{d-1}} f(r\omega)r^{d-1}drd\sigma(\omega). \tag{12.65}
$$

In particular if $f : \mathbb{R}^+ \to \mathbb{R}^+$ is measurable then

$$
\int_{\mathbb{R}^d} f(|x|)dx = \int_0^\infty f(r)V(r) \tag{12.66}
$$

where $V(r) = m(B(0,r)) = r^d m(B(0,1)) = d^{-1} \sigma(S^{d-1}) r^d$.

**Proof.** By Exercise 10.11

$$
\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d \setminus \{0\}} (f \circ \Phi^{-1}) \circ \Phi dm = \int_{(0,\infty) \times S^{d-1}} (f \circ \Phi^{-1}) \, d(\Phi_* m) \tag{12.67}
$$

and therefore to prove Eq. (12.65) we must work out the measure $\Phi_* m$ on $B_{(0,\infty)} \otimes B_{S^{d-1}}$ defined by

$$
\Phi_* m(A) := m(\Phi^{-1}(A)) \quad \forall A \in B_{(0,\infty)} \otimes B_{S^{d-1}}. \tag{12.68}
$$

If $A = (a,b] \times E$ with $0 < a < b$ and $E \in B_{S^{d-1}}$, 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,

$$
(\Phi_* m)((a,b] \times E) = m(bE_1 \setminus aE_1) = m(bE_1) - m(aE_1) = b^d m(E_1) - a^d m(E_1) = \int_a^b r^{d-1}dr. \tag{12.69}
$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$
\rho(J) = \int_J r^{d-1}dr \quad \forall J \in B_{(0,\infty)}, \tag{12.70}
$$

Eq. (12.69) may be written as

$$
(\Phi_* m)((a,b] \times E) = \rho((a,b]) \cdot \sigma(E) = (\rho \otimes \sigma)((a,b] \times E). \tag{12.71}
$$

Since $E = \{(a,b] \times E : 0 < a < b \text{ and } E \in B_{S^{d-1}}\}$ is a $\pi$ class (in fact it is an elementary class) such that $\sigma(E) = B_{(0,\infty)} \otimes B_{S^{d-1}}$, it follows from the $\pi - \lambda$ Theorem and Eq. (12.71) that $\Phi_* m = \rho \otimes \sigma$. Using this result in Eq. (12.67) gives

$$
\int_{\mathbb{R}^d} f dm = \int_{(0,\infty) \times S^{d-1}} (f \circ \Phi^{-1}) \, d(\rho \otimes \sigma)
$$

which combined with Tonelli’s Theorem 12.6 proves Eq. (12.67). □

**Corollary 12.38.** The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1} \subset \mathbb{R}^d$ is

$$
\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{12.72}
$$

where $\Gamma$ is the gamma function is as in Example 10.47 and 10.50.
We continue to work inductively this way to define \( \pi \). As can be seen from Figure 12.6, so that
\[ x_n = \sin \frac{x}{2} \]
for \( n \geq 1 \) we let \( \pi \), which proves Eq. (12.72).

\[ I_d(1) = \int_0^\infty dr \int_0^\infty \sigma(S^{d-1}) \int_0^\infty r^d e^{-r^2} dr. \]

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
\[ \int_0^\infty r^d e^{-r^2} dr = \int_0^\infty \frac{1}{2} u^{-1/2} e^{-u^{1/2}} du = \frac{1}{2} \Gamma(d/2). \] (12.73)

Combining the last two equations with Lemma 12.29 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. (12.72).

12.8 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
\[ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = T_2(\theta, \varphi_1), \]
as can be seen from Figure 12.6 so that
\[ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} T_2(\theta, \varphi_1) \\ r \cos \varphi_1 \end{pmatrix} = \begin{pmatrix} r \sin \varphi_1 \cos \theta \\ r \sin \varphi_1 \sin \theta \\ r \cos \varphi_1 \end{pmatrix} =: T_3(\theta, \varphi_1, r, \varphi_1). \]

We continue to work inductively this way to define
\[ T_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r, \varphi_{n-1}) \]
and more generally,
\[ x_1 = r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3 \]
\[ x_2 = r \sin \varphi_2 \sin \varphi_3 \cos \varphi_4 \]
\[ x_3 = r \sin \varphi_3 \sin \varphi_4 \cos \varphi_5 \]
\[ \vdots \]
\[ x_{n-2} = r \sin \varphi_{n-2} \sin \varphi_{n-3} \cos \varphi_{n-4} \]
\[ x_{n-1} = r \sin \varphi_{n-2} \sin \varphi_{n-3} \cos \varphi_{n-4} \]
\[ x_n = r \cos \varphi_{n-2}. \] (12.74)

By the change of variables formula,
\[ \int_{\mathbb{R}^n} f(x) dm(x) = \int_0^\infty dr \int_{0 \leq \varphi_i \leq \pi, 0 \leq \theta \leq 2\pi} d\varphi_1 \ldots d\varphi_{n-2} \varphi_i =: T_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) \]
\[ \times f(T_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r)) \] (12.75)
where 
\[ \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) := |\det T_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r)|. \]

**Proposition 12.39.** The Jacobian, \( \Delta_n \), is given by 
\[ \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) = r^{n-1} \sin^{n-2} \varphi_{n-2} \sin^2 \varphi_{n-2} \sin \varphi_1. \quad (12.76) \]

If \( f \) is a function on \( rS^{n-1} \) – the sphere of radius \( r \) centered at 0 inside of \( \mathbb{R}^n \), then 
\[ \int_{rS^{n-1}} f(x) d\sigma(x) = r^{n-1} \int_{S^{n-1}} f(r\omega) d\sigma(\omega) = \int_{0 \leq \varphi_i \leq \pi, 0 \leq \theta \leq 2\pi} f(T_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r)) \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) d\varphi_1 \ldots d\varphi_{n-2} d\theta \]
\[ = \int_{0 \leq \varphi_i \leq \pi, 0 \leq \theta \leq 2\pi} \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) d\varphi_1 \ldots d\varphi_{n-2} d\theta \quad (12.77) \]

**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 \( \frac{\partial T_n}{\partial \xi}(\theta, \varphi_1, \ldots, \varphi_{n-2}, \rho) \) we have
\[ \Delta_{n+1}(\theta, \varphi_1, \ldots, \varphi_{n-2}, \varphi_{n-1}, r) = \left| \begin{array}{cccc}
\frac{\partial T_n}{\partial \rho} & \frac{\partial T_n}{\partial \varphi_1} & \cdots & \frac{\partial T_n}{\partial \varphi_{n-2}} \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0
\end{array} \right| \]
\[ = r \left( \cos^2 \varphi_{n-1} + \sin^2 \varphi_{n-1} \right) \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, \rho) \]
\[ = r \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, \rho \sin \varphi_{n-1}), \]
\[ \text{i.e.} \]
\[ \Delta_{n+1}(\theta, \varphi_1, \ldots, \varphi_{n-2}, \varphi_{n-1}, r) = r \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r \sin \varphi_{n-1}). \quad (12.78) \]
To arrive at this result we have expanded the determinant along the bottom row. Starting with \( \Delta_2(\theta, r) = r \) already derived in Example 12.26, Eq. [12.78] implies,
\[ \Delta_3(\theta, \varphi_1, r) = r \Delta_2(\theta, r \sin \varphi_1) = r^2 \sin \varphi_1 \]
\[ \Delta_4(\theta, \varphi_1, \varphi_2, r) = r \Delta_3(\theta, r \sin \varphi_2) = r^3 \sin^2 \varphi_2 \sin \varphi_1 \]
\[ \vdots \]
\[ \Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) = r^{n-1} \sin^{n-2} \varphi_{n-2} \ldots \sin^2 \varphi_1 \sin \varphi_1 \]
which proves Eq. [12.76]. Equation [12.77] now follows from Eqs. [12.65], [12.75] and [12.76].

As a simple application, Eq. [12.77] implies
\[ \sigma(S^{n-1}) = \int_{0 \leq \varphi_i \leq \pi, 0 \leq \theta \leq 2\pi} \sin^{n-2} \varphi_{n-2} \ldots \sin^2 \varphi_2 \sin \varphi_1 d\varphi_1 \ldots d\varphi_{n-2} d\theta \]
\[ = 2\pi \prod_{k=1}^{n-2} \gamma_k = \sigma(S^{n-2}) \gamma_{n-2} \quad (12.79) \]
where \( \gamma_k := \int_0^\pi \sin^k \varphi d\varphi \). If \( k \geq 1 \), we have by integration by parts that,
\[ \gamma_k = \int_0^\pi \sin^k \varphi d\varphi = -\int_0^\pi \sin^{k-1} \varphi \cos \varphi \, d\cos \varphi = 2\delta_{k,1} + (k-1) \int_0^\pi \sin^{k-2} \varphi \cos^2 \varphi \, d\varphi \]
\[ = 2\delta_{k,1} + (k-1) \int_0^\pi \sin^{k-2} \varphi \left( 1 - \sin^2 \varphi \right) d\varphi = 2\delta_{k,1} + (k-1) [\gamma_{k-2} - \gamma_k] \]
and hence \( \gamma_k \) satisfies \( \gamma_0 = \pi, \gamma_1 = 2 \) and the recursion relation
\[ \gamma_k = \frac{k-1}{k} \gamma_{k-2} \text{ for } k \geq 2. \]
Hence we may conclude
\[ \gamma_0 = \pi, \gamma_1 = 2, \gamma_2 = \frac{1}{2}, \gamma_3 = \frac{2}{3}, \gamma_4 = \frac{3}{4}, \gamma_5 = \frac{4}{5}, \gamma_6 = \frac{5}{6}, \gamma_7 = \frac{6}{7} \]
and more generally by induction that
\[ \gamma_{2k} = \frac{(2k-1)!!}{(2k)!!} \quad \text{and} \quad \gamma_{2k+1} = \frac{2k+1}{(2k)!!}. \]
Indeed,
\[ \gamma_{2(k+1)+1} = \frac{2k+2}{2k+3} \gamma_{2k+1} = \frac{2k+2}{2k+3} \frac{(2k)!!}{(2k+1)!!} = 2 \frac{[2(k+1)]!!}{(2k+1)!!} \]
and
\[ \gamma_{2(k+1)} = \frac{2k+1}{2k+1} \gamma_{2k} = \frac{2k+1}{2k+2} \frac{(2k-1)!!}{(2k)!!} = \frac{2(k+1)!!}{(2k+2)!!}. \]
The recursion relation in Eq. (12.79) may be written as
\[ \sigma(S^n) = \sigma(S^{n-1}) \gamma_{n-1} \quad (12.80) \]
which combined with \( \sigma(S^1) = 2\pi \) implies
there exists an $X$ which is verified inductively using Eq. (12.80). Indeed, let $\sigma$ and more generally that

$$\sigma(S) = 2\pi \cdot 2 \cdot \gamma_2 = 2\pi \cdot \frac{1}{2} \pi = \frac{2^2\pi^2}{2!!}, \quad \sigma(S') = 2\pi \cdot \frac{2^2\pi^2}{2!!} \cdot \gamma_3 = -\frac{2^3\pi^2}{3!!}, \quad \sigma(S^2) = 2\pi \cdot 2 \cdot \frac{1}{2} \pi \cdot \frac{2}{3} = \frac{2^4\pi^2}{5!!},$$

and more generally that

$$\sigma(S^{2n}) = \frac{2(2\pi)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{(2\pi)^{n+1}}{(2n)!!} \quad (12.81)$$

which is verified inductively using Eq. (12.80). Indeed,

$$\sigma(S^{2n+1}) = \sigma(S^{2n})\gamma_{2n} = \frac{2(2\pi)^n}{(2n-1)!!} \frac{(2n-1)!!}{(2n)!!} = \frac{(2\pi)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)}) = \sigma(S^{2n+2}) = \sigma(S^{2n+1})\gamma_{2n+1} = \frac{(2\pi)^{n+1}}{(2n)!!} - \frac{2(2n+1)!!}{(2n+1)!!} = \frac{2(2\pi)^{n+1}}{(2n+1)!!}.$$

Using

$$2n!! = 2n(2n-1) \ldots (2 \cdot 1) = 2^n n!$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{(n+1)!}$ which shows that Eqs. (12.65) and (12.81) are in agreement. We may also write the formula in Eq. (12.81) as

$$\sigma(S^n) = \begin{cases} 
\frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\
\frac{(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ odd}
\end{cases}$$

12.9 Gaussian Random Vectors

**Definition 12.40 (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

$$\mathbb{E}[e^{i\lambda \cdot X}] = \exp \left( -\frac{1}{2} Q \lambda \cdot \lambda + i \mu \cdot \lambda \right) \text{ for all } \lambda \in \mathbb{R}^d. \quad (12.82)$$

We will write $X \overset{\text{d}}{=} N(Q, \mu)$ to denote a Gaussian random vector such that Eq. (12.82) holds.

Notice that if there exists a random variable satisfying Eq. (12.82) then its law is uniquely determined by $Q$ and $\mu$ because of Corollary 11.13. In the exercises below you will develop some basic properties of Gaussian random vectors – see Theorem 12.44 for a summary of what you will prove.

**Exercise 12.5.** Show that $Q$ must be non-negative in Eq. (12.82).

**Definition 12.41.** Given a Gaussian random vector, $X$, we call the pair, $(Q, \mu)$ appearing in Eq. (12.82) the characteristics of $X$. We will also abbreviate the statement that $X$ is a Gaussian random vector with characteristics $(Q, \mu)$ by writing $X \overset{\text{d}}{=} N(Q, \mu)$.

**Lemma 12.42.** Suppose that $X \overset{\text{d}}{=} N(Q, \mu)$ and $A : \mathbb{R}^d \to \mathbb{R}^m$ is a $m \times d$ - real matrix and $\alpha \in \mathbb{R}^m$, then $AX + \alpha \overset{\text{d}}{=} N(AQA^T, A\alpha + \alpha)$. In short we might abbreviate this by saying, $AN(Q, \mu) + \alpha \overset{\text{d}}{=} N(AQA^T, A\alpha + \alpha)$.

**Proof.** Let $\xi \in \mathbb{R}^m,$ then

$$\mathbb{E} \left[ e^{i\zeta \cdot (AX+\alpha)} \right] = e^{i\zeta \cdot \alpha} \mathbb{E} \left[ e^{i\zeta \cdot X} \right] = e^{i\zeta \cdot \alpha} \exp \left( -\frac{1}{2} AQA^T \xi \cdot \xi + i \mu \cdot A^T \xi \right)$$

$$= e^{i\zeta \cdot \alpha} \exp \left( -\frac{1}{2} AQA^T \xi \cdot \xi + i (A\mu + \alpha) \cdot \xi \right)$$

from which it follows that $AX + \alpha \overset{\text{d}}{=} N(AQA^T, A\alpha + \alpha)$.

**Exercise 12.6.** 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. (12.82) with $Q = I$ and $\mu = 0$. Also show

$$\mu_i = \mathbb{E} N_i \text{ and } \delta_{ij} = \text{Cov}(N_i, N_j) \text{ for all } 1 \leq i, j \leq d. \quad (12.83)$$

**Hint:** use Exercise 10.15 and (of course) Fubini’s theorem.
Exercise 12.7. 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 12.6. Show that $X$ is Gaussian by showing Eq. (12.82) holds with $Q = AA^t$ ($A^t$ is the transpose of the matrix $A$) and $\mu = \mu$. Also show that

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \text{Cov}(X_i, X_j) \text{ for all } 1 \leq i, j \leq m.$$ (12.84)

Remark 12.43 (Spectral Theorem). Recall that if $Q$ is a real symmetric $d \times d$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u_j^d\}_{j=1}^d$, such that $Q u_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{R}$. Moreover, $\lambda_j \geq 0$ for all $j$ is equivalent to $Q$ being non-negative. When $Q \geq 0$ we may define $Q^{1/2}$ by

$$Q^{1/2} u_j := \sqrt{\lambda_j} u_j \text{ for } 1 \leq j \leq d.$$

Notice that $Q^{1/2} \geq 0$ 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

$$Q^{-1/2} u_j := \frac{1}{\sqrt{\lambda_j}} u_j \text{ for } 1 \leq j \leq d$$

so that $Q^{-1/2} = \left[Q^{1/2}\right]^{-1}$.

Exercise 12.8. 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 12.6. By Exercise 12.7, we know that $X = Q^{1/2} N + \mu$ is a Gaussian random vector satisfying Eq. (12.82). Use the multi-dimensional change of variables formula to show

$$\text{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}} \exp\left(-\frac{1}{2Q^{-1}} (y - \mu) \cdot (y - \mu)\right) dy.$$

Let us summarize some of what the preceding exercises have shown.

Theorem 12.44. 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. (12.82). Moreover for such an $X$,

$$\text{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}} \exp\left(-\frac{1}{2Q^{-1}} (y - \mu) \cdot (y - \mu)\right) dy$$

where $Q$ and $\mu$ may be computed from $X$ using,

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \text{Cov}(X_i, X_j) \text{ for all } 1 \leq i, j \leq m.$$ (12.85)

When $Q$ is degenerate, i.e. $\text{Nul}(Q) \neq \{0\}$, then $X = Q^{1/2} N + \mu$ is still a Gaussian random vectors satisfying Eq. (12.82). However now the Law$_P(X)$ is a measure on $\mathbb{R}^d$ which is concentrated on the non-trivial subspace, $\text{Nul}(Q)^\perp$ – the details of this are left to the reader for now.

Exercise 12.9 (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 $\|x\| := \sqrt{x^t \cdot x}$ and $m := \max \{Qx \cdot x : \|x\| = 1\}$ be the largest eigenvalue of $Q$. Then $\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right] < \infty$ for every $\varepsilon < \frac{1}{m}$.

Remark 12.45. We can in fact compute $\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right]$ exactly – see Eq. (12.86) 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. $Qu_i = q_i u_i$ with $q_i \geq 0$ for $1 \leq i \leq d$. Further let $Z_i := N \cdot u_i$ and observe that $(Z_1, \ldots, Z_d)$ 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, $u_1^t$, $u_2^t$, $\ldots$, $u_d^t$ respectively. From this it follows that $Z \overset{d}{=} N(R^t R, 0) = N(I, 0)$ is still a standard normal random vector. Moreover,

$$\|X\|^2 = \|AN + \mu\|^2 = \|AN\|^2 + \|\mu\|^2 + 2AN \cdot \mu$$

$$= \sum_{i=1}^d \left(q_i Z_i^2 + 2\sqrt{q_i} \mu_i Z_i\right) + \|\mu\|^2$$

where $\mu_i := \mu \cdot u_i$. Therefore using the independence of the $\{Z_i\}_{i=1}^d$,

$$\mathbb{E}\left[e^{\varepsilon\|X\|^2}\right] = e^{\varepsilon\|\mu\|^2} \prod_{i=1}^d \mathbb{E}\left[e^{\varepsilon q_i Z_i^2 + 2\sqrt{q_i} \mu_i Z_i}\right].$$

Now observe that for any $\alpha < 1$, $t \in \mathbb{R}$, and $Z \overset{d}{=} N(0, 1)$, that

$$\mathbb{E}\left[e^{\alpha Z^2 + tZ}\right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\alpha z^2 + tz} e^{-\frac{1}{2}z^2} \, dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(1-2\alpha)z^2 + tz} \, dz$$

$$= \frac{1}{\sqrt{(1-2\alpha)}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}z^2 + t\frac{\alpha}{\sqrt{1-2\alpha}} z} \, dz$$

$$= \frac{1}{\sqrt{(1-2\alpha)}} \exp\left(\frac{1}{2} \frac{t^2}{(1-2\alpha)}\right).$$

wherin 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

---

For those who know about operator norms observe that $m = \|Q\|$ in this case.
Thus we have shown,

\[ \lambda \]

Definition 12.46 (Gaussian Random Vectors).

Because of Eq. (12.85), for all \( \varepsilon \)

\[ E \left[ e^{\varepsilon \|X\|^2} \right] = e^{\varepsilon \|\mu\|^2} \cdot \prod_{i=1}^{d} \frac{1}{\sqrt{1 - 2\varepsilon q_i}} \exp \left( \frac{(2\varepsilon \sqrt{q_i} \mu_i)^2}{2(1 - 2\varepsilon q_i)} \right) \]

\[ = e^{\varepsilon \|\mu\|^2} \cdot \frac{1}{\sqrt{\det(I - 2\varepsilon Q)}} \exp \left( 2\varepsilon^2 \sum_{i=1}^{d} q_i \mu_i^2 \right) \]

\[ = e^{\varepsilon \|\mu\|^2} \cdot \frac{1}{\sqrt{\det(I - 2\varepsilon Q)}} \exp \left( 2\varepsilon^2 \sum_{i=1}^{d} q_i \mu_i^2 \right) \].

To simplify this expression more observe that

\[ \varepsilon \|\mu\|^2 + 2\varepsilon^2 \sum_{i=1}^{d} \frac{q_i}{1 - 2\varepsilon q_i} \mu_i^2 = \sum_{i=1}^{d} \left[ 2\varepsilon^2 \frac{q_i}{1 - 2\varepsilon q_i} + \varepsilon \right] \mu_i^2 \]

\[ = \sum_{i=1}^{d} \frac{\varepsilon}{1 - 2\varepsilon q_i} \mu_i^2 = \varepsilon (1 - 2\varepsilon Q)^{-1} \mu \cdot \mu. \]

Thus we have shown,

\[ E \left[ e^{\varepsilon \|X\|^2} \right] = \frac{1}{\sqrt{\det(I - 2\varepsilon Q)}} \exp \left( \varepsilon \mu \cdot (1 - 2\varepsilon Q)^{-1} \mu \right). \] (12.86)

Remark 12.47. 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 \( d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \) and \( \nu = \frac{1}{2} (\delta_{-1} + \delta_1) \). Then \( (X, Y) : \Omega \to \mathbb{R}^2 \) is a random vector such that both components, \( X \) and \( Y \), are Gaussian random variables but \( (X, Y) \) is not a Gaussian random vector.

Exercise 12.10. Prove the assertion made in Remark 12.47. Hint: explicitly compute \( E \left[ e^{i(\lambda_1 X + \lambda_2 Y)} \right] \).

12.9.1 *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 12.44.

Remark 12.48. Recall that if \( Q \) is a real symmetric \( N \times N \) matrix, then the spectral theorem asserts there exists an orthonormal basis, \( \{u_i\}_{i=1}^{N} \), such that \( Q u_j = \lambda_j u_j \) for some \( \lambda_j \in \mathbb{R} \). Moreover, \( \lambda_j \geq 0 \) for all \( j \) is equivalent to \( Q \) being non-negative. Hence if \( Q \geq 0 \) and \( f : \{\lambda_j : j = 1, 2, \ldots, N\} \to \mathbb{R} \), we may define \( f(Q) \) to be the unique linear transformation on \( \mathbb{R}^N \) such that \( f(Q) u_j = \lambda_j u_j \).

Example 12.49. When \( Q \geq 0 \) and \( f(x) := \sqrt{x} \), we write \( Q^{1/2} \) or \( \sqrt{Q} \) for \( f(Q) \). Notice that \( Q^{1/2} \geq 0 \) and \( Q = Q^{1/2} Q^{1/2} \).

Example 12.50. When \( Q \) is symmetric and

\[ f(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \]

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 \( \lambda_i \neq 0 \) for all \( i \). When \( Q \) is not invertible,

\[ Q^{-1} := f(Q) = Q_{\operatorname{ran}(Q)}^{-1} P, \] (12.88)

where \( P : \mathbb{R}^N \to \mathbb{R}^N \) is orthogonal projection onto the \( \operatorname{Ran}(Q) \). Observe that

\[ P = g(Q) \text{ where } g(x) = 1_{x\neq 0}. \]

Lemma 12.51. For any \( Q \geq 0 \) we can find a matrix, \( A \), such that \( Q = AA^t \). In fact it suffices to take \( A = Q^{1/2} \).

Proposition 12.52. Suppose \( X \overset{d}{=} N(Q, c) \) (see Definition 12.44) 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
\[ \int_{\mathbb{R}^N} f(x) \, d\mu(x) = \frac{1}{Z} \int_{c + \text{Ran}(Q)} f(x) \exp \left( -\frac{1}{2} Q^{-1} (x - c) \cdot (x - c) \right) \, dx \]

where \( dx \) is now “Lebesgue measure” on \( c + \text{Ran}(Q) \), \( Q^{-1} \) is defined as in Eq. \((12.88)\), and \( Z := \sqrt{\det(2\pi Q)}_{\text{Ran}(Q)} \).

**Proof.** Let \( k = \dim \text{Ran}(Q) \) and choose a linear transformation, \( U : \mathbb{R}^k \to \mathbb{R}^n \), such that \( \text{Ran}(U) = \text{Ran}(Q) \) and \( U : \mathbb{R}^k \to \text{Ran}(Q) \) is an isometric isomorphism. Letting \( A := Q^{1/2}U \), we have

\[ AA^* = Q^{1/2}UU^*Q^{1/2} = Q^{1/2}P_{\text{Ran}(Q)}Q^{1/2} = Q. \]

Therefore, if \( Y = (I_{k \times k}, 0) \), then \( X = AY + c \overset{d}{=} N(Q, c) \) by Lemma \( 12.42 \). Observe that \( X - c = Q^{1/2}UY \) takes values in \( \text{Ran}(Q) \) and hence the Law of \( (X - c) \) is a probability measure on \( \mathbb{R}^N \) which is concentrated on \( \text{Ran}(Q) \). From this it follows that \( \mu = P \circ X^{-1} \) is a probability measure on \( \mathbb{R}^N \) which is concentrated on the affine space, \( c + \text{Ran}(Q) \). At any rate from Theorem \( 12.44 \) we have

\[ \int_{\mathbb{R}^N} f(x) \, d\mu(x) = \int_{\mathbb{R}^k} f(Ay + c) \left( \frac{1}{2\pi} \right)^{k/2} e^{-\frac{1}{2} |y|^2} dy \]

\[ = \int_{\mathbb{R}^k} f \left( Q^{1/2}Uy + c \right) \left( \frac{1}{2\pi} \right)^{k/2} e^{-\frac{1}{2} |y|^2} dy. \]

Since

\[ Q^{1/2}Uy + c = U^{1/2}Q^{1/2}Uy + c, \]

we may make the change of variables, \( z = U^{1/2}Q^{1/2}Uy \), using

\[ dz = \sqrt{\det Q}_{\text{Ran}(Q)} \, dy = \prod_{i : \lambda_i \neq 0} \lambda_i^{\frac{1}{2}} \, dy \]

and

\[ |y|^2 = \left( U^{1/2}Q^{1/2}U \right)^{-2} z^2 = \left( U^{1/2}Q^{-1/2}Uz \right)^2 = \left( Q^{-1/2}Uz, Q^{-1/2}Uz \right)_{\mathbb{R}^N} \]

\[ = \left( Q^{-1}Uz, Uz \right)_{\mathbb{R}^N}, \]

to find

\[ \int_{\mathbb{R}^N} f(x) \, d\mu(x) = \int_{\mathbb{R}^k} f(Uz + c) \left( \frac{1}{2\pi} \right)^{k/2} e^{-\frac{1}{2} \left( U^{1/2}Q^{1/2}U \right)^{-1} z^2} \, dz \]

\[ = \int_{\mathbb{R}^k} f(Uz + c) \left( \frac{1}{2\pi} \right)^{k/2} \frac{1}{\sqrt{\det Q}_{\text{Ran}(Q)}} e^{-\frac{1}{2} \left( Q^{-1}Uz, Uz \right)_{\mathbb{R}^N}} \, dz \]

\[ = \int_{\mathbb{R}^k} f(Uz + c) \frac{1}{\sqrt{\det(2\pi Q)_{\text{Ran}(Q)}}} e^{-\frac{1}{2} \left( Q^{-1}Uz, Uz \right)_{\mathbb{R}^N}} \, dz. \]

This completes the proof, since \( z = Uz + c \in c + \text{Ran}(Q) \) is by definition distributed as Lebesgue measure on \( c + \text{Ran}(Q) \) when \( z \) is distributed as Lebesgue measure on \( \mathbb{R}^k \).

### 12.10 Kolmogorov’s Extension Theorems

In this section we will extend the results of Section \( 6.3 \) to spaces which are not simply products of discrete spaces. We begin with a couple of results involving the topology on \( \mathbb{R}^N \).

#### 12.10.1 Regularity and compactness results

**Theorem 12.53 (Inner-Outer Regularity).** Suppose \( \mu \) is a probability measure on \( (\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N}) \), then for all \( B \in \mathcal{B}_{\mathbb{R}^N} \) we have

\[ \mu(B) = \inf \{ \mu(V) : B \subset V \text{ and } V \text{ is open} \} \]  \hfill (12.89)

and

\[ \mu(B) = \sup \{ \mu(K) : K \subset B \text{ with } K \text{ compact} \}. \]  \hfill (12.90)

**Proof.** In this proof, \( C, C_i \) 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 \( (12.89) \) and also that

\[ \mu(B) = \sup \{ \mu(C) : C \subset B \text{ with } C \text{ closed} \}. \]  \hfill (12.91)

Moreover, by MCT, we know that if \( C \) is closed and \( K_n := C \cap \{ x \in \mathbb{R}^N : |x| \leq n \} \), then \( \mu(K_n) \uparrow \mu(C) \). This observation along with Eq. \( (12.91) \) shows Eq. \( (12.90) \) 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 prove the latter assertion, given a closed subset, $C \subset \mathbb{R}^N$, and $\varepsilon > 0$, let

$$C_\varepsilon := \bigcup_{x \in C} B (x, \varepsilon)$$

where $B (x, \varepsilon) := \{ y \in \mathbb{R}^N : |y - x| < \varepsilon \}$. 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 $C \subset \mathcal{F}$.

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

$$\mu (V \setminus C) \leq \mu (V_1 \setminus C_1) + \mu (V_2 \setminus C_2) \leq \mu (V_1 \setminus C_1) + \mu (V_2 \setminus C_2) < 2 \varepsilon.$$ 

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 $A := \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$ if $A_n \in \mathcal{F}$ with $A_0 \cap A_m = \emptyset$ 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 $C^N := \bigcup_{n \leq N} C_n$ and $V := \bigcup_{n=1}^{\infty} V_n$. Then

$$C^N \subset A \subset V$$

and

$$\mu (V \setminus C^N) \leq \sum_{n=0}^{\infty} \mu (V_n \setminus C_n) \leq \sum_{n=0}^{N} \mu (V_n \setminus C_n) + \sum_{n=N+1}^{\infty} \mu (V_n)$$

$$\leq \sum_{n=0}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} [\mu (A_n) + \varepsilon 2^{-n}]$$

$$= \varepsilon + \sum_{n=N+1}^{\infty} \mu (A_n).$$

The last term is less that $2 \varepsilon$ for $N$ sufficiently large because $\sum_{n=1}^{\infty} \mu (A_n) = \mu (A) < \infty$.

**Notation 12.54** Let $I := [0, 1]$, $Q = I^N$, $\pi_j : Q \to I$ be the projection map, $\pi_j (x) = x_j$ (where $x = (x_1, x_2, \ldots, x_j, \ldots)$ for all $j \in \mathbb{N}$, and $B_0 := \sigma (\pi_j : j \in \mathbb{N})$ be the product $\sigma$–algebra on $Q$. Let us further say that a sequence $(x (m))_{m=1}^{\infty} \subset Q$, where $x (m) = (x_1 (m), x_2 (m), \ldots)$, 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 12.55 (Baby Tychonoff’s Theorem).** The infinite dimensional cube, $Q$, is compact, i.e. every sequence $(x (m))_{m=1}^{\infty}$ 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 12.56 (Finite Intersection Property).** Suppose that $K_n \subset Q$ are sets which are, (i) closed under taking sequential limits, and (ii) have the finite intersection property, i.e. $\cap_{m=1}^{\infty} K_m = \emptyset$ for all $m \in \mathbb{N}$, then $\cap_{m=1}^{\infty} K_m = \emptyset$.

**Proof.** By assumption, for each $n \in \mathbb{N}$, there exists $x (n) \in \cap_{m=1}^{n} K_m$. Hence by Lemma 12.55 there exists a subsequence, $(n_k)$, 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 $\cap_{m=1}^{\infty} K_m = \emptyset$.

**12.10.2 Kolmogorov’s Extension Theorem and Infinite Product Measures**

**Theorem 12.57 (Kolmogorov’s Extension Theorem).** Let $I := [0, 1]$. For each $n \in \mathbb{N}$, let $\mu_n$ be a probability measure on $(I^n, B_{I^n})$ such that $\mu_n+1 (A \times I) = \mu_n (A)$. Then there exists a unique measure, $P$ on $(Q, B_Q)$ such that

$$P (A \times Q) = \mu_n (A) \quad (12.92)$$

for all $A \in B_{I^n}$ and $n \in \mathbb{N}$.

**Proof.** Let $A := \cup B_n$ where $B_n := \{ A \times Q : A \in B_{I^n} \} = \sigma (\pi_1, \ldots, \pi_n)$, where $\pi_i (x) = x_i$ if $x = (x_1, x_2, \ldots) \in Q$. Then define $P$ on $A$ by Eq. (12.92), which is easily seen (Exercise 12.11) to be a well defined finitely additive measure on $A$. So to finish the proof it suffices to show if $B_n \in A$ is a decreasing sequence such that

$$\inf_{n \to \infty} P (B_n) = \lim_{n \to \infty} P (B_n) = \varepsilon > 0,$$

then $B := \cap B_n \neq \emptyset$.

To simplify notation, we may reduce to the case where $B_n \in B_n$ for all $n$. To see this is permissible, let us choose $1 \leq n_1 < n_2 < n_3 < \ldots$ such that

$\varepsilon \downarrow 0$ for example, if $K_m = K_m' \times Q$ with $K_m'$ being a closed subset of $I^n$, then $K_m'$ is closed under sequential limits.
$B_k \in B_n$ for all $k$. (This is possible since $B_n$ is increasing in $n$.) We now define a new decreasing sequence of sets, \( \{ B_k \}_{k=1}^\infty \) as follows,

\[
\left( \hat{B}_1, \hat{B}_2, \ldots \right) = \left( Q, Q, B_1, B_1, B_2, B_2, B_3, B_3, \ldots \right).
\]

We then have $\hat{B}_n \in B_n$ for all $n$, $\lim_{n \to \infty} P(\hat{B}_n) = \varepsilon > 0$, and $B = \cap_{n=1}^\infty \hat{B}_n$.

Hence we may replace $B_n$ by $\hat{B}_n$ if necessary so as to have $B_n \in B_n$ for all $n$.

Since $B_n \in B_n$, there exists $B'_n \in B_n$ such that $B_n = B'_n \times Q$ for all $n$. Using the regularity Theorem 12.53 there are compact sets, $K'_n \subset B'_n \subset I^n$, such that $\mu_n(B'_n \setminus K'_n) \leq \varepsilon 2^{-m-1}$ for all $n \in \mathbb{N}$. Let $K_n := K'_n \times Q$, then $P(B_n \setminus K_n) \leq \varepsilon 2^{-m-1}$ for all $n$. Moreover,

\[
P(B_n \cap [\cap_{m=1}^n K_m]) = P(\cup_{m=1}^n [B_n \setminus K_m]) \leq \sum_{m=1}^n P(B_n \setminus K_m) \leq \sum_{m=1}^n \varepsilon 2^{-m-1} \leq \varepsilon / 2.
\]

So, for all $n \in \mathbb{N}$,

\[
P(\cap_{m=1}^n K_m) = P(B_n) - P(B_n \setminus [\cap_{m=1}^n K_m]) \geq \varepsilon - \varepsilon / 2 = \varepsilon / 2,
\]

and in particular, $\cap_{m=1}^n K_m \neq \emptyset$. An application of Corollary 12.56 now implies, $\emptyset \neq \cap_{m=1}^n K_m \subset \cap_{m=1}^n B_m$.

**Exercise 12.11.** Show that Eq. (12.92) defines a well defined finitely additive measure on $A := \cup B_n$.

The next result is an easy corollary of Theorem 12.57.

**Theorem 12.58.** Suppose \( \{ (X_n, \mathcal{M}_n) \}_{n \in \mathbb{N}} \) are standard Borel spaces (see Appendix 12.11 below), $X := \prod_{n \in \mathbb{N}} X_n$, $\pi_n : X \to X_n$ be the $n$th projection map, $\mathcal{B}_n := \sigma(\pi_k : k \leq n)$, $\mathcal{B} = \sigma(\pi_n : n \in \mathbb{N})$, and $T_n := X_{n+1} \times X_{n+2} \times \ldots$. Further suppose that for each $n \in \mathbb{N}$ we are given a probability measure, $\mu_n$ on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that

\[
\mu_n(A \times X_{n+1}) = \mu_n(A) \quad \text{for all } n \in \mathbb{N} \text{ and } A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n.
\]

Then there exists a unique probability measure, $P$, on $(X, \mathcal{B})$ such that $P(A \times T_n) = \mu_n(A)$ for all $A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$.

**Proof.** Since each $(X_n, \mathcal{M}_n)$ is measure theoretic isomorphic to a Borel subset of $I$, we may assume that $X_n \in B_I$ and $\mathcal{M}_n = (B_I)_X$ for all $n$. Given $A \in B_I$, let $\tilde{\mu}_n(A) := \mu_n(A \cap [X_1 \times \cdots \times X_n])$ a probability measure on $B_I$. Furthermore,

\[
\tilde{\mu}_{n+1}(A \times I) = \mu_{n+1}([A \times I] \cap [X_1 \times \cdots X_{n+1}]) = \mu_{n+1}((A \cap [X_1 \times \cdots X_n]) \times X_{n+1}) = \mu_n(A \cap [X_1 \times \cdots X_n]) = \tilde{\mu}_n(A).
\]

Hence by Theorem 12.57 there is a unique probability measure, $\bar{P}$, on $I^N$ such that

\[
\bar{P}(A \times I^n) = \tilde{\mu}_n(A) \quad \text{for all } n \in \mathbb{N} \text{ and } A \in B_I.
\]

We will now check that $P := \bar{P}|_{\otimes_{n=1}^\infty \mathcal{M}_n}$ is the desired measure. First off we have

\[
\bar{P}(X) = \lim_{n \to \infty} \bar{P}(X_1 \times \cdots \times X_n \times I^n) = \lim_{n \to \infty} \tilde{\mu}_n(X_1 \times \cdots \times X_n) = 1.
\]

Secondly, if $A \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$, we have

\[
P(A \times T_n) = \bar{P}(A \times T_n) = \bar{P}((A \times I^n) \cap X) = \bar{P}(A \times I^n) = \tilde{\mu}_n(A) = \mu_n(A).
\]

Here is an example of this theorem in action.

**Theorem 12.59 (Infinite Product Measures).** Suppose that $\{ \nu_n \}_{n=1}^\infty$ are a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$ and $\mathcal{B} := \otimes_{n=1}^\infty \mathcal{B}_n$ is the product $\sigma$-algebra on $\mathbb{R}^N$. Then there exists a unique probability measure, $\nu$, on $(\mathbb{R}^N, \mathcal{B})$, such that

\[
\nu(A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^N) = \nu_1(A_1) \cdots \nu_n(A_n) \quad \forall A_i \in \mathcal{B}_n \quad \forall n \in \mathbb{N}.
\]

Moreover, this measure satisfies,

\[
\int_{\mathbb{R}^N} f(x_1, \ldots, x_n) \, d\nu(x) = \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) \, d\nu_1(x_1) \cdots d\nu_n(x_n)
\]

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 12.58 with $\mu_n := \nu_1 \otimes \cdots \otimes \nu_n$ 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 \( (\mathbb{R}^n, B) \) such that

\[ \nu(A \times \mathbb{R}^n) = \mu_n(A) \text{ for all } A \in B_{\mathbb{R}^n} \text{ and } n \in \mathbb{N}. \]

Taking \( A = A_1 \times A_2 \times \cdots \times A_n \) with \( A_i \in B_{\mathbb{R}} \) then gives Eq. (12.93). For this measure, it follows that Eq. (12.94) holds when \( f = 1_{A_1 \times \cdots \times A_n} \). Thus by an application of Theorem 11.2 with \( \mathbb{M} = \{1_{A_1 \times \cdots \times A_n} : A_i \in B_{\mathbb{R}}\} \) and \( \mathbb{H} \) being the set of bounded measurable functions, \( f : \mathbb{R}^n \to \mathbb{R} \), for which Eq. (12.94) shows that Eq. (12.94) 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 20.71 below.

### 12.11 Appendix: Standard Borel Spaces

For more information along the lines of this section, see Royden \[39\] and Parthasarathy \[34\].

**Definition 12.60.** 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 12.61.** A measurable space, \((X, \mathcal{M})\) is said to be a **standard Borel space** if \((X, \mathcal{M}) \cong (B, B_\mathbb{R})\) where \( B \) is a Borel subset of \(((0, 1) : B_{(0, 1)}\).

**Definition 12.62 (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 12.72 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 \( \mathbb{R}^3, \) are all (measure theoretic) isomorphic to \((0, 1)\). Moreover we will also see that a countable product of standard Borel spaces is again a standard Borel space, see Corollary 12.69.

*On first reading, you may wish to skip the rest of this section.

**Lemma 12.63.** Suppose \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) are measurable spaces such that \( X = \biguplus_{n=1}^{\infty} X_n, \ Y = \biguplus_{n=1}^{\infty} Y_n, \) 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 \((X := \biguplus_{n=1}^{\infty} X_n, \bigotimes_{n=1}^{\infty} \mathcal{M}_n)\) are \((Y := \biguplus_{n=1}^{\infty} Y_n, \bigotimes_{n=1}^{\infty} \mathcal{N}_n)\) 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

\[ f^{-1}(B) = \bigcup_{n=1}^{\infty} f_n^{-1}(B \cap Y_n) = \bigcup_{n=1}^{\infty} f_n^{-1}(B \cap Y_n) \in \mathcal{M}. \]

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

\[ f(x) = (f_1(x_1), f_2(x_2), \ldots) \text{ with } x = (x_1, x_2, \ldots) \in X. \]

Again it is clear that \( f \) is bijective and measurable, since

\[ f^{-1}\left( \biguplus_{n=1}^{\infty} B_n \right) = \bigcup_{n=1}^{\infty} f_n^{-1}(B_n) \in \bigotimes_{n=1}^{\infty} \mathcal{N}_n \]

for all \( B_n \in \mathcal{M}_n \) and \( n \in \mathbb{N} \). Similar reasoning shows that \( f^{-1} \) is measurable as well.

**Proposition 12.64.** 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 \((0, 1) \cup A \) where \( A \) 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 12.63 to observe that

\[ (-1, 1) = \{0\} \cup \biguplus_{n=0}^{\infty} ((-2^{-n}, -2^{-n}] \cup [2^{-n-1}, 2^{-n}) \]

and

\[ [-1, 1] = \{0\} \cup \biguplus_{n=0}^{\infty} ([{-2^{-n}, -2^{-n-1}] \cup (2^{-n-1}, 2^{-n})]. \]

Similarly \((0, 1)\) is isomorphic to \((0, 1)\) because
(0, 1) = \sum_{n=0}^{\infty} (2^{-n}, 2^{-n}) 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 \( A = \{1\} \), then by Lemma 12.63 and what we have already proved, \((0, 1] \cup \{1\} \cong (0, 1)\). Similarly if \( N \in \mathbb{N} \) with \( N \geq 2 \) and \( A = \{2, \ldots, N+1\} \), then

\[
(0, 1] \cup A \cong (0, 1] \cup A = (0, 2^{-N+1}] \cup \left[ \sum_{n=1}^{N-1} (2^{-n}, 2^{-n-1}] \cup A \right.
\]

while

\[
(0, 1) = (0, 2^{-N+1}] \cup \left[ \sum_{n=1}^{N-1} (2^{-n}, 2^{-n-1}] \right. \cup \left\{ 2^{-n} : n = 1, 2, \ldots, N \right\}
\]

and so again it follows from what we have proved and Lemma 12.63 that \((0, 1] \cong (0, 1) \cup A\). Finally if \( A = \{2, 3, 4, \ldots\} \) is a countable set, we can show \((0, 1] \cong (0, 1) \cup A\) with the aid of the identities,

\[
(0, 1) = \left[ \sum_{n=1}^{\infty} (2^{-n}, 2^{-n-1}] \right. \cup \left\{ 2^{-n} : n \in \mathbb{N} \right\}
\]

and

\[
(0, 1] \cup A \cong (0, 1] \cup A = \left[ \sum_{n=1}^{\infty} (2^{-n}, 2^{-n-1}] \right. \cup A.
\]

Notation 12.65 Suppose \((X, \mathcal{M})\) is a measurable space and \(A\) is a set. Let \(\pi_a : X^A \to X\) denote projection 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 12.66. If \(\varphi : A \to B\) is a bijection of sets and \((X, \mathcal{M})\) is a measurable space, then \((X^A, \mathcal{M}^{\otimes A}) \cong (X^B, \mathcal{M}^{\otimes B})\).

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 \(f^{-1}(a) = a \circ \varphi^{-1}\). 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 \(\pi_{\varphi(a)}^{X^B}\) 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 12.67. Let \(\Omega := \{0, 1\}^n\), \(\pi_i : \Omega \to \{0, 1\}\) be projection onto the \(i\)th component, and \(B := \sigma(\pi_1, \pi_2, \ldots)\) be the product \(\sigma\)-algebra on \(\Omega\). Then \((\Omega, B) \cong ((0, 1], B_{(0, 1)})\).

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 \(r_1(x) = x\),

\[
\gamma_1(x) := 1_{\{x \geq 1\}} and r_2(x) := x - 2^{-1}\gamma_1(x) \in (0, 2^{-1}),
\]

then let \(r_2 := 1_{\{x \geq 1\}} and r_3 = r_2 - 2^{-2}\gamma_2 \in (0, 2^{-2})\). Working inductively, we construct \((\gamma_k(x), r_k(x))_{k=1}^\infty\) such that \(\gamma_k(x) \in \{0, 1\}\), and

\[
r_{k+1}(x) = r_k(x) - 2^{-k}\gamma_k(x) = x - \sum_{j=1}^k 2^{-j}\gamma_j(x) \in (0, 2^{-k})
\]

for all \(k\). Let us now define \(g : [0, 1] \to \Omega\) by \(g(x) := (\gamma_1(x), \gamma_2(x), \ldots)\). 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

\[
x = \sum_{j=1}^\infty 2^{-j}\gamma_j(x) and r_{k+1}(x) = \sum_{j=k+1}^\infty 2^{-j}\gamma_j(x).
\]

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. (12.96)) \(r_{k+1}(x) = 2^{-k}\) which would contradict Eq. (12.95). 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

$$A := \Omega \setminus \Omega_0 = \cup_{n=1}^{\infty} \{ \omega \in \Omega : \omega_j = 1 \text{ for } j \geq n \}$$

is a countable set, it follows that $A \in \mathcal{B}$ and therefore $\Omega_0 = \Omega \setminus A \in \mathcal{B}$. Hence we may now conclude that $g : (0,1), B_{(0,1)} \to (\Omega_0, B_{(0,1)})$ is a measurable bijection with measurable inverse given by $f|_{\Omega_0}$, i.e. $(0,1), B_{(0,1)} \cong (\Omega_0, B_{(0,1)})$. An application of Lemma 12.63 and Proposition 12.64 now implies

$$\Omega = \Omega_0 \cup A \cong [0,1) \cup \mathbb{N} \cong [0,1) \cong (0,1).$$

\cleardoublepage

**Corollary 12.68.** The following spaces are all isomorphic to $((0,1), B_{(0,1)}); (0,1)^d$ and $\mathbb{R}^d$ for any $d \in \mathbb{N}$ and $[0,1]^N$ and $\mathbb{R}^N$ in which both of these spaces are equipped with their natural product $\sigma$ – algebras.

**Proof.** In light of Lemma 12.63 and Proposition 12.64 we know that $[0,1)^d \cong \mathbb{R}^d$ and $(0,1)^N \cong [0,1) \cong \mathbb{R}$, respectively. So, using Proposition 12.67, it suffices to show $(0,1)^d \cong \Omega \cong (0,1)^N$ and to do this it suffices to show $\Omega^d \cong \Omega$ and $\Omega^N \cong \Omega$.

To reduce the problem further, let us observe that $\Omega^d \cong \{0,1\}^{N \times \{1,2,3\ldots,d\}}$ and $\Omega^N \cong \{0,1\}^N$. For example, let $\omega : \Omega \to \{0,1\}^N$ be defined by $g(\omega(i,j) = \omega(i,j)$ for all $\omega \in \Omega^d = \{0,1\}^N$ and $g$ is a bijection and since $\pi_{(i,j)} \circ g(\omega) = \pi_{(i,j)}(\omega) = (\omega(i,j))$, it follows that $g$ is measurable. The inverse, $g^{-1} : \{0,1\}^N \to \Omega^N$, to $g$ is given by $g^{-1}(\omega) = (\omega(i,j) = \alpha(i,j)$. To see this map is measurable, we have $\pi_i \circ g^{-1} : \{0,1\}^N \to \Omega = \{0,1\}^N$ is given $\pi_{(i,j)} \circ g^{-1}(\alpha) = g^{-1}(\alpha)(i,j) = g^{-1}(\alpha)(i,j) = g^{-1}(\alpha)(\omega(i,j)) = \pi_{(i,j)}(\omega) \cong \{0,1\}^{N\times \{1,2,3\ldots,d\}}$

\cleardoublepage

**Corollary 12.69.** Suppose that $(\mathbb{X}_n, M_n)$ for $n \in \mathbb{N}$ are standard Borel spaces, then $X := \prod_{n=1}^{\infty} \mathbb{X}_n$ equipped with the product $\sigma$ – algebra, $M := \otimes_{n=1}^{\infty} M_n$ 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 isomorphism $f_n : X_n \to A_n$. Then $f : X \to A := \prod_{n=1}^{\infty} A_n$ defined by $f(x_1, x_2, \ldots) = (f_1(x_1), f_2(x_2), \ldots)$ is easily seen to be a measure theoretic isomorphism when $A$ is equipped with the product $\sigma$ – algebra, $\otimes_{n=1}^{\infty} B_{A_n}$. So according to Corollary 12.68, to finish the proof it suffices to show $\otimes_{n=1}^{\infty} B_{A_n} = \mathcal{M}_A$ where $\mathcal{M} := \otimes_{n=1}^{\infty} B_{[0,1]}$ is the product $\sigma$ – algebra on $[0,1]^\mathbb{N}$.

The $\sigma$ – algebra, $\otimes_{n=1}^{\infty} B_{A_n}$, is generated by sets of the form, $B = \prod_{n=1}^{\infty} B_n$ where $B_n \in B_{A_n} \subset B_{[0,1]}$. On the other hand, the $\sigma$ – algebra, $\mathcal{M}_A$ is generated by sets of the form, $A \cap B$ where $B := \prod_{n=1}^{\infty} B_n$ with $B_n \in B_{[0,1]}$. Since

$$A \cap B = \prod_{n=1}^{\infty} (B \cap A_n) = \prod_{n=1}^{\infty} B_n$$

where $B_n = B \cap A_n$ is the generic element in $B_{A_n}$, we see that $\otimes_{n=1}^{\infty} B_{A_n}$ and $\mathcal{M}_A$ can both be generated by the same collections of sets, we may conclude that $\otimes_{n=1}^{\infty} B_{A_n} = \mathcal{M}_A$.

Our next goal is to show that any Polish space with its Borel $\sigma$ – algebra is a standard Borel space.

**Corollary 12.70** Let $Q := [0,1]^\mathbb{N}$ denote the (infinite dimensional) unit cube in $\mathbb{R}^\mathbb{N}$. For $a,b \in Q$ let

$$d(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n} |a_n - b_n| = \sum_{n=1}^{\infty} \frac{1}{2^n} |\pi_n(a) - \pi_n(b)|.$$

**Exercise 12.12.** Show $d$ is a metric and that the Borel $\sigma$ – algebra on $(Q,d)$ is the same as the product $\sigma$ – algebra. 

**Theorem 12.71.** To every separable metric space $(\mathbb{X}, \rho)$, there exists a continuous injective map $G : \mathbb{X} \to Q$ such that $G : \mathbb{X} \to G(\mathbb{X}) \subset Q$ is a homeomorphism. Moreover, if the metric, $\rho$, is also complete, then $G(\mathbb{X})$ is a $G_{\delta}$ – set, i.e. the $G_{\delta}$ is the countable intersection of open subsets of $(Q,d)$. In short, any separable metrizable space $\mathbb{X}$ is homeomorphic to a subset of $(Q,d)$ and if $\mathbb{X}$ is a Polish space then $\mathbb{X}$ 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 $D = \{a_n\}^{\infty}_{n=1}$ be a countable dense subset of $\mathbb{X}$ and define

$$G(x) = (\rho(x,a_1), \rho(x,a_2), \rho(x,a_3), \ldots) \in Q$$

where $\rho$ is the distance function.
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,
\[ \rho(x,y) \leq \rho(x,a_n) + \rho(y,a_n) = 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n) \leq 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n) + 2^n\gamma(x,y). \]
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 let \( \overline{d} \) be the metric on \( S \) defined by \( \overline{d}(G(x),G(y)) = \rho(x,y) \) for all \( x, y \in X \). Then \((S,\overline{d})\) is a complete metric (being the isometric image of a complete metric space) and by what we have just prove, \( \tau_{\overline{d}} = \tau_\rho \). Consequently, if \( u \in S \) and \( \varepsilon > 0 \) is given, we may find \( \delta'(<\varepsilon) \) such that \( B_{\overline{d}}(u,\delta'(\varepsilon)) \subset B_d(u,\varepsilon) \). Taking \( \delta(\varepsilon) = \min(\delta'(\varepsilon),\varepsilon) \), we have \( \text{diam}_{\overline{d}}(B_d(u,\delta(\varepsilon))) < \varepsilon \) and \( \text{diam}_{\overline{d}}(B_d(u,\delta'(\varepsilon))) < \varepsilon \).

\[ \text{diam}_{\overline{d}}(A) := \{\sup d(u,v) : u,v \in A\} \quad \text{and} \quad \text{diam}_d(A) := \{\sup d(u,v) : u,v \in A\}. \]

Let \( \bar{S} \) denote the closure of \( S \) inside of \((Q,d)\) and for each \( n \in \mathbb{N} \) let
\[ N_n := \{N \in \tau_\overline{d} : \text{diam}_d(N) \vee \text{diam}_\overline{d}(N \cap S) < 1/n\} \]
and let \( U_n := \cup N_n \in \tau_\overline{d} \). From the previous paragraph, it follows that \( S \subset U_n \) and therefore \( S \subset \bar{S} \cap (\bigcap_{n=1}^{\infty} U_n) \).

Conversely if \( u \in \bar{S} \cap (\bigcap_{n=1}^{\infty} U_n) \) and \( n \in \mathbb{N} \), there exists \( N_n \in N_n \) such that \( u \in N_n \). Moreover, since \( N_1 \cap \cdots \cap N_n \) is an open neighborhood of \( u \) in \( \bar{S} \), there exists \( u_n \in N_1 \cap \cdots \cap N_n \cap S \) for each \( n \in \mathbb{N} \). From the definition of \( N_n \), we have \( \lim_{n \to \infty} d(u,u_n) = 0 \) and \( \sigma(u_n,u_m) \leq \max\left(n^{-1},m^{-1}\right) \to 0 \) 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 \((S,d_\bar{S})\) 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, \( S = \bar{S} \cap (\bigcap_{n=1}^{\infty} U_n) \). This completes the proof because we may write \( S = (\cap_{n=1}^{\infty} U_{1/n}) \) where \( U_{1/n} := \{u \in Q : d(u,\bar{S}) < 1/n\} \) and therefore, \( S = (\cap_{n=1}^{\infty} U_n) \cap (\cap_{n=1}^{\infty} U_{1/n}) \) is a \( \mathcal{G}_\delta \) set.

**Corollary 12.72.** 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 [12.71] shows that \( X \) is homeomorphic to a measurable (in fact a \( \mathcal{G}_\delta \)) 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 \( X \).

### 12.12 More Exercises

**Exercise 12.13.** 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 \( \pi := F_{\mu_j}(\mathcal{M}_j \otimes \mathcal{N}_j) \), i.e. \( \pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3](F^{-1}(A)) \) for all \( A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{N}_3 \). 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 \((\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{N}_3,\mathcal{B}_{\mathbb{R}}) \) – 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 12.14.** Prove the second assertion of Theorem [12.20] 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 [6.14].
Exercise 12.15. (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 $\int_X \int_X 1_D(x,y)d\nu(y) \, dm(x) \neq \int_X \int_X 1_D(x,y)dm(x) \, d\nu(y)$ by explicitly computing both sides of this equation.

Exercise 12.16. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 12.17. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the Folland Problem 2.50 on p. 69 pertaining to the Fubini Theorem involving counting measures.)

Exercise 12.18. By explicitly computing both sides of this equation.

Exercise 12.20. Show $\int_X \int_X 1_D(x,y)dm(x) \, d\nu(y)$ by explicitly computing both sides of this equation.


Exercise 12.22. Folland Problem 2.60 on p. 77. Many of the following exercises are probably repeats of the exercises above.

Exercise 12.23. Folland Problem 2.46 on p. 69. Let $\sigma$ be Lebesgue measure on $[0,1]$ and therefore $\int f(x) \, d\sigma(x)$ is not defined as a Lebesgue integral.

Exercise 12.24. Folland Problem 2.61 on p. 77. That is show that

Exercise 12.25. Folland Problem 2.62 on p. 77. Fractional integration.

Exercise 12.26. Folland Problem 2.64 on p. 80. Rotation invariance of surface measure on $S^{n-1}$.

Exercise 12.27. Show, using Problem 12.25 that $\int_{S^{d-1}} \omega_i \omega_j \, d\sigma(\omega) = \frac{1}{d} \delta_{ij} \sigma(S^{d-1})$.

Hint: show $\int_{S^{d-1}} \omega_i^2 \, d\sigma(\omega)$ is independent of $i$ and therefore $\int_{S^{d-1}} \omega_i^2 \, d\sigma(\omega) = \frac{1}{d} \sum_{j=1}^{d} \int_{S^{d-1}} \omega_j^2 \, d\sigma(\omega)$.

12.13 Exercises

Many of the following exercises are probably repeats of the exercises above.

Exercise 12.28. Prove Theorem 12.28. Suggestion, to get started define $\pi(A) := \int_X \int_X d\mu(x_1) \cdot \int_X d\mu(x_n) 1_A(x_1, \ldots, x_n)$ and then show Eq. (??) holds. Use the case of two factors as the model of your proof.

Exercise 12.29. Let $(X_j, M_j, \mu_j)$ 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 $(M_1 \otimes M_2) \otimes M_3, M_1 \otimes M_2 \otimes M_3, (M_1 \otimes M_2) \otimes M_3)$ – measurable and $F^{-1}$ is $(M_1 \otimes M_2 \otimes M_3) –$ measurable. That is $F((X_1 \times X_2) \times X_3, (M_1 \otimes M_2) \otimes M_3) \to (X_1 \times X_2 \times X_3, M_1 \otimes M_2 \otimes M_3)$ is a measure theoretic isomorphism.

2. Let $\pi := F^*, [\{(\mu_1 \otimes \mu_2) \otimes \mu_3\}]$, i.e. $\pi(A) = [\{(\mu_1 \otimes \mu_2) \otimes \mu_3\}]$ for all $A \in M_1 \otimes M_2 \otimes M_3$. Then $\pi$ is the unique measure on $M_1 \otimes M_2 \otimes M_3$ 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 $A_i \in M_i$. 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 $(M_1 \otimes M_2 \otimes M_3, \mathcal{B}_{\mathbb{R}^d})$ – 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 12.30. Prove the second assertion of Theorem 12.30. 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 12.30.

Exercise 12.31. (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 12.32.** Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

**Exercise 12.33.** Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the \( \mathcal{M} \times \mathcal{B}_{\mathbb{R}} \) should be \( \mathcal{M} \otimes \mathcal{B}_{\mathbb{R}} \) in this problem.)

**Exercise 12.34.** Folland Problem 2.55 on p. 77. (Explicit integrations.)

**Exercise 12.35.** 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 12.36.** 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 12.37.** Folland Problem 2.57 on p. 77.

**Exercise 12.38.** Folland Problem 2.58 on p. 77.

**Exercise 12.39.** Folland Problem 2.60 on p. 77. Properties of the \( \Gamma \) function.

**Exercise 12.40.** Folland Problem 2.61 on p. 77. Fractional integration.

**Exercise 12.41.** Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on \( S^{n-1} \).

**Exercise 12.42.** Folland Problem 2.64 on p. 80. On the integrability of \( x^\alpha \| \log |x| \|^\beta \) for \( x \) near 0 and \( x \) near \( \infty \) in \( \mathbb{R}^n \).

**Exercise 12.43.** Show, using Problem 12.41 that

\[
\int_{S^{d-1}} \omega_i \omega_j \, d\sigma(\omega) = \frac{1}{d} \delta_{ij} \sigma(S^{d-1}).
\]

**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 13.1.** We say that \( A \) is independent of \( B \) if \( P(A|B) = P(A) \) or equivalently that

\[
P(A \cap B) = P(A) P(B).
\]

We further say a finite sequence of collections of sets, \( \{C_i\}_{i=1}^n \), are independent if

\[
P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j)
\]

for all \( A_i \in C_i \) and \( J \subset \{1, 2, \ldots, n\} \).

13.1 Basic Properties of Independence

If \( \{C_i\}_{i=1}^n \), are independent classes then so are \( \{C_i \cup \{\Omega\}\}_{i=1}^n \). Moreover, if we assume that \( \Omega \in C_i \) for each \( i \), then \( \{C_i\}_{i=1}^n \), are independent iff

\[
P(\cap_{j=1}^n A_j) = \prod_{j=1}^n P(A_j) \quad \text{for all} \quad (A_1, \ldots, A_n) \in C_1 \times \cdots \times C_n.
\]

**Theorem 13.2.** Suppose that \( \{C_i\}_{i=1}^n \) is a finite sequence of independent \( \pi \)-classes. Then \( \sigma(C_i) \) for each \( i \) are also independent.

**Proof.** As mentioned above, we may always assume without loss of generality that \( \Omega \in C_i \). Fix, \( A_j \in C_j \) for \( j = 2, 3, \ldots, n \). We will begin by showing that

\[
Q(A) := P(A \cap A_2 \cap \cdots \cap A_n) = P(A) P(A_2) \cdots P(A_n) \quad \text{for all} \quad A \in \sigma(C_1).
\]  
(13.1)

Since \( Q(\cdot) \) and \( P(A_2) \cdots P(A_n) P(\cdot) \) are both finite measures agreeing on \( \Omega \) and \( A \) in the \( \pi \)-system \( C_1 \), Eq. (13.1) is a direct consequence of Proposition 6.55. Since \( (A_2, \ldots, A_n) \in C_2 \times \cdots \times C_n \) were arbitrary we may now conclude that \( \sigma(C_1), C_2, \ldots, C_n \) are independent.

By applying the result we have just proved to the sequence, \( C_2, \ldots, C_n, \sigma(C_1) \) shows that \( \sigma(C_2), C_3, \ldots, C_n, \sigma(C_1) \) are independent. Similarly we show inductively that

\[
\sigma(C_j), C_{j+1}, \ldots, C_n, \sigma(C_1), \ldots, \sigma(C_{j-1})
\]

are independent for each \( j = 1, 2, \ldots, n \). The desired result occurs at \( j = n \).

**Corollary 13.3.** Suppose that \( \{A_j\}_{j=1}^m \) and \( \{A'_k\}_{k=1}^n \) are sets from \( \mathcal{B} \) and let \( A = A \left( \{A_j\}_{j=1}^m \right) \) and \( A' = A \left( \{A'_k\}_{k=1}^n \right) \), then \( A \) and \( A' \) are \( P \)-independent iff \( P(A_j \cap B_K) = P(A_j) P(B_K) \) for all \( J \subset \{1, 2, \ldots, m\} \) and \( K \subset \{1, 2, \ldots, n\} \) where \( A_j = \cap_{i \in J} A_i \) and \( A'_K = \cap_{k \in K} A'_k \).

**Proof.** Apply Theorem 13.2 with \( C_1 = \{A_j : J \subset \{1, 2, \ldots, m\}\} \) and \( C_2 = \{A'_K : K \subset \{1, 2, \ldots, n\}\} \).

**Definition 13.4.** Let \( (\Omega, \mathcal{B}, P) \) be a probability space and \( \{A_j\}_{j=1}^N \subset \mathcal{B} \) be a collection of events and for \( J \subset \{1, \ldots, N\} \), let \( A_J := A \left( \{A_j\}_{j \in J} \right) \). We say that there are no hidden dependencies among the \( \{A_j\}_{j=1}^N \) if for all proper subsets, \( J \subset \{1, \ldots, N\} \), the algebras \( A_J \) and \( A_{J'} \) are \( P \)-independent.

**Corollary 13.5 (Characterizing no hidden independence).** Let \( \{A_j\}_{j=1}^N \subset \mathcal{B} \). Then there are no hidden dependencies among the \( \{A_j\}_{j=1}^N \) iff

\[
P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j) \quad \text{for all} \quad J \subset \{1, \ldots, N\}.
\]  
(13.2)

**Proof.** We will show if there are no hidden dependencies among the \( \{A_j\}_{j=1}^N \), then we will show that Eq. (13.2) holds by induction on \( n := |J| \). When \( n = 1 \) there is nothing to prove. For \( n = 2 \), we have \( J = \{a, b\} \) and we let \( A = \{a\} \) so that \( b \in A' \). By the assumption that \( A_A \) and \( A_{A'} \) are \( P \)-independent it follows that \( P(A_a \cap A_b) = P(A_a) P(A_b) \) as \( A_a \in A_A \) and \( A_b \in A_{A'} \). When
n = 3 so that J = \{a, b, c\} we let A = \{a, b\} so that c \in A^c. As A_a \cap A_b \in \mathcal{A}_A\ A_c \in \mathcal{A}_{A^c}, it follows by the hypothesis that
\[ P( A_a \cap A_b \cap A_c ) = P( [A_a \cap A_b] \cap A_c ) = P( A_a \cap A_b ) P( A_c ) . \]
However, by the n = 2 case we also know that P( A_a \cap A_b ) = P( A_a ) P( A_b )
and so we have shown,
\[ P( A_a \cap A_b \cap A_c ) = P( A_a \cap A_b ) P( A_c ) = P( A_a ) P( A_b ) P( A_c ) . \]
The full simple inductive proof is now left to the reader.

For the converse direction, for any \( J \subset \{1, 2, \ldots, N\} \) let
\[ C_J := \{ A_A = \cap_{j \in A} A_j : A \subset J \} \]
which is a multiplicative class generating \( A_J \). From Eq. \[13.2\] it is now easy to verify that \( C_J \) and \( C_J \) are \( P \)-independent for all \( J \subset \{1, 2, \ldots, N\} \) and hence, by Theorem \[13.2\] so are the algebras, \( A_J \) and \( A_{J^c} \).

**Definition 13.6.** Let (\( \Omega, \mathcal{B}, P \)) be a probability space, \( \{ (S_i, S_i) \}_{i=1}^n \) be a collection of measurable spaces and \( Y_i : \Omega \to S_i \) be a measurable map for 1 \( \leq i \leq n \). The maps \( \{ Y_i \}_{i=1}^n \) are \( P \)-independent if \( \{ C_i \}_{i=1}^n \) are \( P \)-independent, where \( C_i := Y_i^{-1}(S_i) = \sigma(Y_i) \subset \mathcal{B} \) for 1 \( \leq i \leq n \).

**Theorem 13.7 (Independence and Product Measures).** Let (\( \Omega, \mathcal{B}, P \)) be a probability space, \( \{ (S_i, S_i) \}_{i=1}^n \) 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 \( \mu_i := P \circ Y_i^{-1} = \text{Law}_P(Y_i) \). Then \( \{ Y_i \}_{i=1}^n \) are independent iff
\[ \text{Law}_P(Y_1, \ldots, Y_n) = \mu_1 \otimes \cdots \otimes \mu_n , \]
where \( (Y_1, \ldots, Y_n) : \Omega \to S_1 \times \cdots \times S_n \) and
\[ \text{Law}_P(Y_1, \ldots, Y_n) = P \circ (Y_1, \ldots, Y_n)^{-1} : S_1 \otimes \cdots \otimes S_n \to [0, 1] \]
is the joint law of \( Y_1, \ldots, Y_n \).

**Proof.** Recall that the general element of \( C_i \) is of the form \( A_i = Y_i^{-1}(B_i) \) with \( B_i \in S_i \). Therefore for \( A_i = Y_i^{-1}(B_i) \in C_i \) we have
\[ P( A_1 \cap \cdots \cap A_n ) = P(( Y_1, \ldots, Y_n ) \in B_1 \times \cdots \times B_n ) \]
= \( (Y_1, \ldots, Y_n) \mu ) (B_1 \times \cdots \times B_n ) . \]
If \( (Y_1, \ldots, Y_n) \mu = \mu_1 \otimes \cdots \otimes \mu_n \) it follows that
\[ P( A_1 \cap \cdots \cap A_n ) = \mu_1 \otimes \cdots \otimes \mu_n (B_1 \times \cdots \times B_n ) \]
= \( \mu_1(B_1) \cdots \mu_n(B_n) = P(Y_1 \in B_1) \cdots P(Y_n \in B_n) \]
= \( P(A_1) \cdots P(A_n) \)
and therefore \( \{ C_i \} \) are \( P \)-independent and hence \( \{ Y_i \} \) are \( P \)-independent. Conversely if \( \{ Y_i \} \) are \( P \)-independent, i.e. \( \{ C_i \} \) are \( P \)-independent, then
\[ P(( Y_1, \ldots, Y_n ) \in B_1 \times \cdots \times B_n ) = P( A_1 \cap \cdots \cap A_n ) \]
= \( P(A_1) \cdots P(A_n) \)
= \( P(Y_1 \in B_1) \cdots P(Y_n \in B_n) \]
= \( \mu_1(B_1) \cdots \mu_n(B_n) \)
= \( \mu_1 \otimes \cdots \otimes \mu_n (B_1 \times \cdots \times B_n ) . \]
Since
\[ \pi := \{ B_1 \times \cdots \times B_n : B_i \in S_i \text{ for } 1 \leq i \leq n \} \]
is a \( P \)-system which generates \( S_1 \otimes \cdots \otimes S_n \) and
\[ (Y_1, \ldots, Y_n) \mu = \mu_1 \otimes \cdots \otimes \mu_n \text{ on } \pi , \]
it follows that \( (Y_1, \ldots, Y_n) \mu \) \( P \)-independent on all of \( S_1 \otimes \cdots \otimes S_n \).

**Remark 13.8.** When have a collection of not necessarily independent random functions, \( Y_i : \Omega \to S_i \), like in Theorem \[13.7\] it is not in general possible to recover the joint distribution, \( \pi := \text{Law}_P(Y_1, \ldots, Y_n) \), from the individual distributions, \( \mu_i = \text{Law}_P(Y_i) \) 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 \( (\mathbb{R}, \mathcal{B}) \), and \( (Y_1, Y_2) \) have joint distribution, \( \pi \), given by,
\[ \pi(C) = \int_C( x, x) d\mu(x) \text{ for all } C \in \mathcal{B} . \]
If we let \( \mu_i = \text{Law}_P(Y_i) \), then for all \( A \in \mathcal{B} \) we have
\[ \mu_1(A) = P(Y_1 \in A) = P( (Y_1, Y_2) \in A \times \mathbb{R} ) \]
= \( \pi( A \times \mathbb{R} ) = \int_{\mathbb{R}} 1_{A \times \mathbb{R}}( x, x) d\mu(x) = \mu(A) \).
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 \( (\mathbb{R}^2, \mathcal{B}) \) with the same marginals as \( \pi \), i.e. \( \pi( A \times \mathbb{R} ) = \mu( A) \) for all \( A \in \mathcal{B} \).

**Lemma 13.9.** Let \( (\Omega, \mathcal{B}, P) \) be a probability space, \( \{ (S_i, S_i) \}_{i=1}^n \) and \( \{ (T_i, T_i) \}_{i=1}^n \) 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
\[ \sigma(F_i \circ Y_i) = (F_i \circ Y_i)^{-1}(T_i) = Y_i^{-1}(F_i^{-1}(T_i)) \subset Y_i^{-1}(S_i) = C_i. \]
The fact that \( \{\sigma(F_i \circ Y_i)\}_{i=1}^n \) is independent now follows easily from the assumption that \( \{C_i\} \) are \( P \)-independent.

Example 13.10. If \( \Omega := \prod_{i=1}^n S_i, B := S_1 \otimes \cdots \otimes S_n, Y_i(s_1, \ldots, s_n) = s_i \) for all \((s_1, \ldots, s_n) \in \Omega, \) and \( C_i := Y_i^{-1}(S_i) \) for all \(i.\) Then the probability measures, \( P \) on \((\Omega, B)\) for which \( \{C_i\}_{i=1}^n \) are independent are precisely the product measures, \( P = \mu_1 \otimes \cdots \otimes \mu_n \) where \( \mu_i \) is a probability measure on \((S_i, S_i) \) for \(1 \leq i \leq n.\) Notice that in this setting,
\[ C_i := Y_i^{-1}(S_i) = \{s_1 \times \cdots \times s_i-1 \times B \times S_{i+1} \times \cdots \times S_n : B \in S_i \} \subset B. \]

Proposition 13.11. Suppose that \((\Omega, B, P)\) 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
\[ E\left[ \prod_{j=1}^n Z_j \right] = \prod_{j=1}^n E[Z_j]. \]

Proof. Let \( \mu_j := P \circ Z_j^{-1} : B \rightarrow [0, 1] \) be the law of \( Z_j \) for each \(j.\) Then we know \((Z_1, \ldots, Z_n)_* P = \mu_1 \otimes \cdots \otimes \mu_n.\) Therefore by Example 10.52 and Tonelli’s theorem,
\[ E\left[ \prod_{j=1}^n |Z_j| \right] = \int_{\mathbb{R}^n} \left| \prod_{j=1}^n |z_j| \right| d(\otimes_{j=1}^n \mu_j)(z) \]
\[ = \prod_{j=1}^n \int_{\mathbb{R}} |z_j| d\mu_j(z_j) = \prod_{j=1}^n E|Z_j| < \infty \]
which shows that \( \prod_{j=1}^n Z_j \) is integrable. Thus again by Example 10.52 and Fubini’s theorem,
\[ E\left[ \prod_{j=1}^n Z_j \right] = \int_{\mathbb{R}^n} \left( \prod_{j=1}^n |z_j| \right) d(\otimes_{j=1}^n \mu_j)(z) \]
\[ = \prod_{j=1}^n \int_{\mathbb{R}} z_j d\mu_j(z_j) = \prod_{j=1}^n E[Z_j]. \]

Theorem 13.12. Let \((\Omega, B, P)\) be a probability space, \( \{(S_i, S_i)\}_{i=1}^n \) be a collection of measurable spaces and \( Y_i : \Omega \rightarrow S_i \) be a measurable map for \(1 \leq i \leq n.\) Further let \( \mu_i := P \circ Y_i^{-1} = \text{Law}_P(\mu_i) \) and \( \pi := P \circ (Y_1, \ldots, Y_n)^{-1} : \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n \) be the joint distribution of
\[ (Y_1, \ldots, Y_n) : \Omega \rightarrow S_1 \times \cdots \times S_n. \]
Then the following are equivalent,
1. \( \{Y_i\}_{i=1}^n \) are independent,
2. \( \pi = \mu_1 \otimes \cdots \otimes \mu_n \)
3. for all bounded measurable functions, \( f : (S_1 \times \cdots \times S_n, \mathcal{S}_i \otimes \cdots \otimes \mathcal{S}_n) \rightarrow (\mathbb{R}_+, \mathcal{B}_\mathbb{R}), \)
\[ Ef(Y_1, \ldots, Y_n) = \int_{S_1 \times \cdots \times S_n} f(x_1, \ldots, x_n) \prod_{i=1}^n dx_i, \quad (13.3) \]
( where the integrals may be taken in any order),
4. \( E\left[ \prod_{i=1}^n f_i(Y_i) \right] = \prod_{i=1}^n E[f_i(Y_i)] \) for all bounded (non-negative) measurable functions, \( f_i : S_i \rightarrow \mathbb{R} \) or \( \mathbb{C}.\)

Proof. (1 \( \iff \) 2) has already been proved in Theorem 13.7. The fact that (2 \( \iff \) 3) now follows from Exercise 10.11 and Fubini’s theorem. Similarly, (3 \( \iff \) 4) follows from Exercise 10.11 and Fubini’s theorem after taking \( f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i). \) Lastly for (4 \( \iff \) 1), let \( A_i \subset S_i \) and take \( f_i := 1_{A_i}. \) In 4, to learn,
\[ P(\bigcap_{i=1}^n \{Y_i \in A_i\}) = E\left[ \prod_{i=1}^n 1_{A_i}(Y_i) \right] = \prod_{i=1}^n E[1_{A_i}(Y_i)] = \prod_{i=1}^n P(Y_i \in A_i) \]
which shows that the \( \{Y_i\}_{i=1}^n \) are independent.

Corollary 13.13. Suppose that \((\Omega, B, P)\) is a probability space and \( \{Y_j : \Omega \rightarrow \mathbb{R}\}_{j=1}^n \) is a sequence of random variables with countable ranges, say \( \Lambda \subset \mathbb{R}. \) Then \( \{Y_j\}_{j=1}^n \) are independent iff
\[ P(\bigcap_{j=1}^n \{Y_j = y_j\}) = \prod_{j=1}^n P(Y_j = y_j) \quad (13.4) \]
for all choices of \(y_1, \ldots, y_n \in \Lambda.\)

Proof. If the \( \{Y_j\} \) are independent then clearly Eq. (13.4) holds by definition as \( \{Y_j = y_j\} \in Y_j^{-1}(\mathcal{B}_{\mathbb{R}}). \) Conversely if Eq. (13.4) holds and \( f_i : \mathbb{R} \rightarrow [0, \infty) \) are measurable functions then,
\[ \mathbb{E} \left[ \prod_{i=1}^{n} f_i (Y_i) \right] = \sum_{y_1, \ldots, y_n \in A} \prod_{i=1}^{n} f_i (y_i) \cdot P (\cap_{j=1}^{n} \{ Y_j = y_j \}) \]
\[ = \sum_{y_1, \ldots, y_n \in A} \prod_{i=1}^{n} f_i (y_i) \cdot \prod_{j=1}^{n} P (Y_j = y_j) \]
\[ = \prod_{i=1}^{n} \sum_{y_i \in A} f_i (y_i) \cdot P (Y_j = y_j) \]
\[ = \prod_{i=1}^{n} \mathbb{E} [f_i (Y_i)] \]

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 13.12.

**Definition 13.14 (i.i.d.).** A sequence 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 \( (X_n, P = (X_n)_n, P \) for all \( n, k \). That is we should have

\[ P (X_n \in A) = P (X_k \in A) \quad \text{for all} \quad n, k, n \in \mathbb{N} \quad \text{and} \quad A \in \mathcal{B}. \]

If \( \{ X_n \}_{n=1}^{\infty} \) are i.i.d. random variables if

\[ P (X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{j=1}^{n} P (X_1 \in A_1) = \prod_{j=1}^{n} P (X_1 \in A_i) = \prod_{j=1}^{n} \mu (A_i) \quad (13.5) \]

where \( \mu = (X_1)_n, P \). The identity in Eq. (13.5) 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 13.19, the \( \{ Y_n \}_{n=1}^{\infty} \) there are i.i.d. with \( \text{Law}_P (Y_n) = P \circ Y_n^{-1} = \mu \) for all \( n \in \mathbb{N} \). We will give another proof of the existence of arbitrary sequences of independent random variables in Proposition 13.23 below which will rely on the following very special case.

**Exercise 13.1.** Suppose that \( \Omega = [0, 1), \mathcal{B} = \mathcal{B}_{[0,1)} \), and \( P = m \) is Lebesgue measure on \( \mathcal{B} \). Let \( Y_1 (\omega) := \omega_i \) 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

\[ \omega = \sum_{i=1}^{\infty} Y_i (\omega) 2^{-i} \quad \text{for all} \quad \omega_i \in \{ 0, 1 \}. \]

As long as \( \omega \neq k2^{-n} \) for some \( 0 < k \leq n \), the above equation uniquely determines the \( \{ Y_i (\omega) \}_n \). Owing to the fact that \( \sum_{i=1}^{\infty} 2^{-i} = 2^{-n} \), 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, \( \{ Y_i \}_{i=1}^{\infty} \), 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 \( (Y_1 (\omega), \ldots, Y_n (\omega)) \) is equivalent to knowing for which \( k \in \mathbb{N} \cap [0, 2^n) \) 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,

\[ P (\{ Y_{n+1} = 1 \} | [2^{-n}k, 2^{-n} (k + 1)]) = \frac{1}{2} = P (\{ Y_{n+1} = 0 \} | [2^{-n}k, 2^{-n} (k + 1)]). \]

See Section 13.3 if you need more guidance with this exercise.

**Exercise 13.2.** Let \( X, Y \) be two random variables on \( (\Omega, \mathcal{B}, P) \).

1. Show that \( X \) and \( Y \) are independent iff \( \text{Cov} (f (X), g (Y)) = 0 \) (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 \( \text{Cov} (X, Y) = 0 \).
3. Show by example that if \( X, Y \in L^2 (P) \) and \( \text{Cov} (X, Y) = 0 \) does not necessarily imply that \( X \) and \( Y \) are independent. **Hint:** try taking \( (X, Y) = (X, ZX) \) where \( X \) and \( Z \) are independent random variables such that \( \mathbb{E} Z = 0 \) is similar to Remark 12.47.

**Exercise 13.3 (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 \( \text{Cov} (f (X), g (X)) \geq 0 \). **Hint:** let \( Y \) be another random variable which has the same law as \( X \) and is independent of \( X \). Then consider

\[ \mathbb{E} [(f (Y) - f (X)) \cdot (g (Y) - g (X))]. \]

Let us now specialize to the case where \( S_i = \mathbb{R}^m_i \) and \( S_i = \mathbb{B}_{\mathbb{R}^m} \), for some \( m_i \in \mathbb{N} \).

**Theorem 13.15.** Let \( (\Omega, \mathcal{B}, P) \) be a probability space, \( m_j \in \mathbb{N}, S_j = \mathbb{R}^{m_j}, S_j = \mathbb{B}_{\mathbb{R}^{m_j}}, Y_j : \Omega \to S_j \) be random vectors, and \( \mu_j := \text{Law}_P (Y_j) = P \circ Y_j^{-1} : S_j \to [0, 1] \) for \( 1 \leq j \leq n \). The the following are equivalent:

1. \( \{ Y_j \}_{j=1}^{n} \) are independent,
2. \( \text{Law}_P (Y_1, \ldots, Y_n) = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n \),
3. for all bounded measurable functions, \( f : (S_1 \times \cdots \times S_n, S_1 \otimes \cdots \otimes S_n) \to (\mathbb{R}, \mathcal{B}_\mathbb{R}) \),

\[ \mathbb{E} [f (Y_1, \ldots, Y_n)] = \int_{S_1 \times \cdots \times S_n} f (x_1, \ldots, x_n) d \mu_1 (x_1) \ldots d \mu_n (x_n), \quad (13.6) \]

( 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. \[ P \left( \bigcap_{j=1}^{m} \{ Y_j \leq y_j \} \right) = \prod_{j=1}^{n} P \left( \{ Y_j \leq y_j \} \right) \] 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. \[ \mathbb{E} \left[ \prod_{j=1}^{n} \lambda_j (Y_j) \right] = \prod_{j=1}^{n} \mathbb{E} [\lambda_j (Y_j)] \] for all \( \lambda_j \in S_j = \mathbb{R}^{m_j} \).

**Proof.** The equivalence of 1. – 4. has already been proved in Theorem 13.12. 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 \( (-\infty, y_j] := (-\infty, (y_j)_1] \times \cdots \times (-\infty, (y_j)_m] \).

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 11.3 or Corollary 11.10. In each case, let \( \mathbb{M} \) denote the linear space of bounded measurable functions such that Eq. (13.6) 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 \( N = m_1 + \cdots + m_n \)

\[ y = (y_1, \ldots, y_n) = (y_1^1, y_2^2, \ldots, y_n^n) \in \mathbb{R}^N \]

\[ \lambda = (\lambda_1, \ldots, \lambda_n) = (\lambda_1^1, \lambda_2^2, \ldots, \lambda_n^n) \in \mathbb{R}^N \]

For showing 5. \( \implies \) 3. take \( \mathbb{M} = \{ 1_{(-\infty, y]} : y \in \mathbb{R}^N \} \).

For showing 6. \( \implies \) 3. take \( \mathbb{M} \) to be a those functions on \( \mathbb{R}^N \) which are of the form, \( f (y) = \prod_{j=1}^{n} f_j (y_j) \) with each \( f_j \in C_c (\mathbb{R}) \).

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 \( N = m_1 + \cdots + m_n \), \( \pi := \text{Law} (Y_1, \ldots, Y_n) \) and \( \lambda := (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^N \). Item 7. states,

\[ \hat{\pi} (\lambda) := \int_{\mathbb{R}^N} e^{i \lambda \cdot y} d\pi (y) = \mathbb{E} \left[ e^{i \sum_{j=1}^{n} \lambda_j \cdot Y_j} \right] \]

\[ = \prod_{j=1}^{n} \mathbb{E} [e^{i \lambda_j \cdot Y_j}] = \prod_{j=1}^{n} \int_{\mathbb{R}^d} e^{i \lambda_j \cdot y} d\mu_j (y_j) \]

\[ = (\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n) (\lambda) \]

and so \( \pi = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n \) by Corollary 11.13.

**Definition 13.16.** A collection of subsets of \( \mathcal{B} \), \( \{ C_t \}_{t \in T} \) is said to be **independent** iff \( \{ C_t \}_{t \in T} \) are independent for all finite subsets, \( \Lambda \subset T \). More explicitly, we are requiring

\[ P \left( \bigcap_{t \in \Lambda} A_t \right) = \prod_{t \in \Lambda} P (A_t) \]

whenever \( \Lambda \) is a finite subset of \( T \) and \( A_t \in \mathcal{C}_t \) for all \( t \in \Lambda \).

**Corollary 13.17.** If \( \{ C_t \}_{t \in T} \) is a collection of independent classes such that each \( C_t \) is a \( \pi \)-system, then \( \{ \sigma (C_t) \}_{t \in T} \) are independent as well.

**Definition 13.18.** A collections of random variables, \( \{ X_t : t \in T \} \) are **independent** iff \( \{ \sigma (X_t) : t \in T \} \) are independent.

**Example 13.19.** Suppose that \( \{ \mu_n \}_{n=1}^{\infty} \) is any sequence of probability measure on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \). Let \( \Omega = \mathbb{R}^N \), \( \mathcal{B} := \otimes_{n=1}^{\infty} \mathcal{B}_\mathbb{R} \) 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 \( \text{Law}_P (Y_n) = \mu_n \) for each \( n \).

**Lemma 13.20 (Independence of groupings).** Suppose that \( \{ B_t : t \in T \} \) 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

\[ \mathcal{B}_{T_s} = \bigvee_{t \in T_s} B_t := \sigma (\cup_{t \in T_s} B_t) . \]

Then \( \{ \mathcal{B}_{T_s} \}_{s \in S} \) is again independent family of \( \sigma \)-fields.

**Proof.** Let

\[ \mathcal{C}_s = \{ \cap_{t \in K} B_t : B_t \in \mathcal{B}_t, K \subset T_s \} . \]

It is now easily checked that \( \mathcal{B}_{T_s} \) is \( \sigma \) (\( \mathcal{C}_s \)) 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 13.17.
Remark 13.21. To better understand the last proof it is instructive to write out a special case in more detail. Suppose that \( \{ A, B, C, D, E \} \) are independent sub-\( \sigma \)– algebras relative to \( (\Omega, \mathcal{F}, P) \). This independence implies

\[
P(A \cap B \cap C \cap D \cap E) = P(A) P(B) P(C) P(D) P(E) = P(A \cap B) P(C \cap D \cap E)
\]

for all \( A \in A, B \in B, C \in C, D \in D, \) and \( E \in E \). Therefore,

\[
C_1 := \{ A \cap B : A \in A \text{ and } B \in B \} \quad \text{and} \quad C_2 := \{ C \cap D \cap E : C \in C, D \in D, \text{ and } E \in E \}
\]

are independent. As \( C_1 \) and \( C_2 \) are multiplicative we may conclude from Theorem 13.2 that \( A \cap B = \sigma(C_1) \) and \( C \cap D \cap E = \sigma(C_2) \) are independent as well.

Corollary 13.22. Suppose that \( \{ Y_n \}_{n=1}^\infty \) is a sequence of independent random variables (or vectors) and \( A_1, \ldots, A_m \) is a collection of pairwise disjoint subsets of \( \mathbb{N} \). Further suppose that \( f_i : \mathbb{R}^{A_i} \to \mathbb{R} \) is a measurable function for each \( 1 \leq i \leq m \), then \( Z_i := f_i (\{ Y_n \}_{n \in A_i}) \) is again a collection of independent random variables.

Proof. Notice that \( \sigma(Z_i) \subset \sigma(\{ Y_n \}_{n \in A_i}) = \sigma(\cup_{i \in A} \sigma(Y_i)) \). Since \( \sigma(\{ Y_i \}_{i \in A_i}) \) are independent by assumption, it follows from Lemma 13.20 that \( \{ \sigma(\{ Y_n \}_{n \in A_i}) \}_{i=1}^m \) are independent and therefore so is \( \{ \sigma(Z_i) \}_{i=1}^m \), i.e. \( \{ Z_i \}_{i=1}^m \) are independent.

Proposition 13.23. 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 \( \text{Law}_P(X_n) = \mu_n \) 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 13.1. Let \( \varphi : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \) be a bijection and set \( Z_{t,n} := Y_{\varphi(t,n)} \) so that \( \{ Z_{t,n} \}_{t,n=1}^\infty \) are i.i.d. Bernoulli random variables on \((0,1), \mathcal{B}_{[0,1)}, P = m\). For each \( n \in \mathbb{N} \), let \( U_n := \sum_{t=1}^\infty Z_{t,n} 2^{-t} \). According to Corollary 13.22, \( \{ U_n \}_{n=1}^\infty \) are independent random variables. Moreover the reader should verify that \( \text{Law}_P(U_n) = m \) 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 9.48 so that \( \text{Law}_P(G_n(U_n)) = \mu_n \) for each \( n \). Then \( X_n := G_n(U_n) \) is a desired collection of independent random variables on \((0,1), \mathcal{B}_{[0,1)}, P = m\). [See Section 13.9 for more details on this type of construction.]

The following theorem follows immediately from the definitions and Theorem 13.15.

Theorem 13.24. Let \( X := \{ X_t : t \in T \} \) be a collection of random variables. Then the following are equivalent:

1. The collection \( X \) is independent,
2. \[ P(\cap_{t \in A} \{ X_t \in A_t \}) = \prod_{t \in A} P(X_t \in A_t) \]
for all finite subsets, \( A \subset T \), and all \( \{ A_t \}_{t \in A} \subset \mathcal{B}_\mathbb{R} \),
3. \[ P(\cap_{t \in A} \{ X_t \leq x_t \}) = \prod_{t \in A} P(X_t \leq x_t) \]
for all finite subsets, \( A \subset T \), and all \( \{ x_t \}_{t \in A} \subset \mathbb{R} \).
4. For all \( \Gamma \subset_f T \) and \( f : \mathbb{R}^n \to \mathbb{R} \) which are bounded an measurable for all \( t \in \Gamma \),
\[
\mathbb{E}
\left[
\prod_{t \in \Gamma} f_t (X_t)
\right]
= \prod_{t \in \Gamma} \mathbb{E} f_t (X_t) = \int_{\mathbb{R}^n} \prod_{t \in \Gamma} f_t (x_t) \prod_{t \in \Gamma} d\mu_t (x_t).
\]
5. \[ \mathbb{E} \left( \prod_{t \in \Gamma} \exp (e^{a_t} X_t) \right) = \prod_{t \in \Gamma} \mathbb{E} \mu_t (\lambda) .\]
6. For all \( \Gamma \subset_f T \) and \( f : (\mathbb{R}^n)^\Gamma \to \mathbb{R} \),
\[
\mathbb{E} [ f(X_\Gamma) ] = \int_{(\mathbb{R}^n)^\Gamma} f(x) \prod_{t \in \Gamma} d\mu_t (x_t).
\]
7. For all \( \Gamma \subset_f T \), \( \text{Law}_P(X_\Gamma) = \otimes_{t \in \Gamma} \mu_t \).
8. \( \text{Law}_P(X) = \otimes_{t \in T} \mu_t \).

Moreover, if \( B_t \) is a sub-\( \sigma \)– algebra of \( \mathcal{B} \) for \( t \in T \), then \( \{ B_t \}_{t \in T} \) are independent iff for all \( \Gamma \subset_f T \),
\[
\mathbb{E}
\left[
\prod_{t \in \Gamma} X_t
\right]
= \prod_{t \in \Gamma} \mathbb{E} X_t \quad \text{for all } X_t \in L^\infty (\Omega, \mathcal{B}_t, P).
\]

Proof. The equivalence of 1. and 2. follows almost immediately form the definition of independence and the fact that \( \sigma(X_t) = \{ \{ X_t \in A \} : A \in \mathcal{B}_\mathbb{R} \} \). Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 13.17 with \( \mathcal{C}_t := \{ \{ X_t \leq a \} : a \in \mathbb{R} \} \) and making use the observations that \( \mathcal{C}_t \) is a \( \pi \)– system for all \( t \) and that \( \sigma(\mathcal{C}_t) = \sigma(X_t) \). The remaining equivalence are also easy to check.
13.2 Examples of Independence

13.2.1 An Example of Ranks

Lemma 13.25 (No Ties). Suppose that \( X \) and \( Y \) are independent random variables on a probability space \((\Omega, \mathcal{F}, P)\). If \( F(x) := P(X \leq x) \) 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

\[
P(X = Y) = \mathbb{E}[1_{\{X = Y\}}] = \int_{\mathbb{R}^2} 1_{\{x = y\}} d(\mu \otimes \nu)(x, y)
\]

\[
= \int_{\mathbb{R}} d\nu(y) \int_{\mathbb{R}} d\mu(x) 1_{\{x = y\}} = \int_{\mathbb{R}} \mu(\{y\}) d\nu(y)
\]

\[
= \int_{\mathbb{R}} 0 d\nu(y) = 0.
\]

Second Proof. For sake of comparison, let’s give a proof where we do not allow ourselves to use Fubini’s theorem. To this end let \( \{a_l := \frac{l}{N}\}_{l=-\infty}^{\infty} \) (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

\[
\{(x, x) : x \in \mathbb{R}\} \subset \bigcup_{l \in \mathbb{Z}} [(a_l, a_{l+1}] \times (a_l, a_{l+1})]
\]

and therefore,

\[
P(X = Y) \leq \sum_{l \in \mathbb{Z}} P(X \in (a_l, a_{l+1}]) \times Y \in (a_l, a_{l+1}) = \sum_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)]^2
\]

\[
\leq \sup_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)] \sum_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)] = \sup_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)].
\]

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

\[
P(X = Y) \leq \limsup_{l \to \infty} \sup_{l \in \mathbb{Z}} F\left(\frac{l+1}{N}\right) - F\left(\frac{l}{N}\right) = 0.
\]

Thus \( R_n = k \) if \( X_n \) is the \( k \)th – largest element in the list, \((X_1, \ldots, X_n)\).

For example if \((X_1, X_2, X_3, X_4, X_5, X_6, X_7, \ldots) = (9, -8, 3, 7, 23, 0, -11, \ldots)\), 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 \( i = 1, 2, \ldots, 5 \) 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. \((*, *, X_4, *, X_5)\). Similarly, \( R_2 = 3 \) 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 \( i = 1, 2, \ldots, n \), then \( X_n < X_{n-1} < \cdots < X_1 \).

Theorem 13.26 (Renyi Theorem). Let \( \{X_n\}_{n=1}^{\infty} \) be i.i.d. and assume that \( F(x) := P(X_n \leq x) \) is continuous. Then \( \{R_n\}_{n=1}^{\infty} \) is an independent sequence,

\[
P(R_n = k) = \frac{1}{n} \quad \text{for} \quad k = 1, 2, \ldots, n,
\]

and the events, \( A_n = \{X_n \text{ is a record}\} = \{R_n = 1\} \) are independent as \( n \) varies and

\[
P(A_n) = P(R_n = 1) = \frac{1}{n}.
\]

Proof. By Problem 6 on p. 110 of Resnick or by Fubini’s theorem, \((X_1, \ldots, X_n)\) and \((X_{\sigma_1}, \ldots, X_{\sigma_n})\) have the same distribution for any permutation \( \sigma \).

Since \( F \) is continuous, it now follows that up to a set of measure zero,

\[
\Omega = \sum_{\sigma} \{X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}\}
\]

and therefore

\[
1 = P(\Omega) = \sum_{\sigma} P(\{X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}\}).
\]

Since \( P(\{X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}\}) \) is independent of \( \sigma \), we may now conclude that

\[
P(\{X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}\}) = \frac{1}{n!}
\]

for all \( \sigma \). As observed before the statement of the theorem, to each realization \((\varepsilon_1, \ldots, \varepsilon_n)\), (here \( \varepsilon_i \in \mathbb{N} \) with \( \varepsilon_i \leq i \)) of \((R_1, \ldots, R_n)\) there is a uniquely
determined permutation, \( \sigma = \sigma (\varepsilon_1, \ldots, \varepsilon_n) \), such that \( X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n} \). (Notice that there are \( n! \) permutations of \( \{1, 2, \ldots, n\} \) and there are also \( n! \) choices for the \( \{(\varepsilon_1, \ldots, \varepsilon_n) : 1 \leq \varepsilon_i \leq i\} \).) From this it follows that

\[
\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\} = \{X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}\}
\]

and therefore,

\[
P (\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\}) = P (X_{\sigma_1} < X_{\sigma_2} < \cdots < X_{\sigma_n}) = \frac{1}{n!}.
\]

Since

\[
P (\{R_n = \varepsilon_n\}) = \sum_{(\varepsilon_1, \ldots, \varepsilon_{n-1})} P (\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\})
\]

\[
= \sum_{(\varepsilon_1, \ldots, \varepsilon_{n-1})} \frac{1}{n!} = (n-1)! \cdot \frac{1}{n!} = \frac{1}{n}
\]

we have shown that

\[
P (\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\}) = \frac{1}{n!} = \prod_{j=1}^{n-1} \frac{1}{j} = \prod_{j=1}^{n} P (\{R_j = \varepsilon_j\}).
\]

**Remark 13.27.** Renyi’s Theorem is really about random permutations with the uniform distribution. The point is that to each realization, \((X_1, \ldots, X_n)\), there is a permutation \( \sigma \) such that \((X_{\sigma_1}, \ldots, X_{\sigma_n})\) is listed in increasing order. What we showed above is that \( \sigma \) is a random permutation with the uniform distribution. We then defined

\[
R_k (\sigma) := \text{rank of } (k) \text{ in the list } (\sigma (1), \ldots, \sigma (k)).
\]

Thus if \( n = 5 \) and \( (\sigma (1), \ldots, \sigma (5)) = (3, 5, 2, 1, 4) \) 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

\[
R_3 (\sigma) = 2 \implies (\sigma (1), \ldots, \sigma (5)) = (\ast, \ast, \ast, 1, 4)
\]

\[
R_3 (\sigma) = 3 \implies (\sigma (1), \ldots, \sigma (5)) = (\ast, \ast, 2, 1, 4)
\]

\[
R_3 (\sigma) = 1 \implies (\sigma (1), \ldots, \sigma (5)) = (\ast, 5, 2, 1, 4)
\]

\[
R_1 (\sigma) = 1 \implies (\sigma (1), \ldots, \sigma (5)) = (3, 5, 2, 1, 4).
\]

Thus it follows that

\[
P (R_i = \varepsilon_i : 1 \leq i \leq n) = P (\sigma = (\varepsilon_1, \ldots, \varepsilon_n)) = \frac{1}{n!}.
\]

Further observe that

\[
P (R_k = \varepsilon_k) = \frac{1}{n!} \cdot (k-1)! \cdot n \cdot (n-1) \ldots (n-k+1) = \frac{1}{k}.
\]

Therefore we see that

\[
P (R_i = \varepsilon_i : 1 \leq i \leq n) = \prod_{k=1}^{n} P (R_k = \varepsilon_k).
\]

### 13.3 Independence for Gaussian Random Vectors

As you saw in Exercise [13.2](#), uncorrelated random variables are typically not independent. However, if the random variables involved are jointly Gaussian (see Definition [12.40](#)), then independence and uncorrelated are actually the same thing!

**Lemma 13.28.** Suppose that \( Z = (X, Y)^t \) is a Gaussian random vector with \( X \in \mathbb{R}^k \) and \( Y \in \mathbb{R}^l \). Then \( X \) is independent of \( Y \) iff \( \text{Cov} (X_1, Y_1) = 0 \) for all \( 1 \leq i \leq k \) and \( 1 \leq j \leq l \). This lemma also holds more generally. Namely if \( \{X^l\}_{l=1}^{n} \) is a sequence of random vectors such that \((X^1, \ldots, X^n)\) is a Gaussian random vector. Then \( \{X^1\}_{l=1}^{n} \) are independent iff \( \text{Cov} (X^l_1, X^k_1) = 0 \) for all \( l \neq k \) and \( i \) and \( k \).

**Proof.** We know by Exercise [13.2](#) that if \( X_i \) and \( Y_j \) are independent, then \( \text{Cov} (X_i, Y_j) = 0 \). For the converse direction, if \( \text{Cov} (X_i, Y_j) = 0 \) for all \( 1 \leq i \leq k \) and \( 1 \leq j \leq l \) and \( x \in \mathbb{R}^k \) and \( y \in \mathbb{R}^l \), then

\[
\text{Var} (x \cdot X + y \cdot Y) = \text{Var} (x \cdot X) + \text{Var} (y \cdot Y) + 2 \text{Cov} (x \cdot X, y \cdot Y)
\]

\[
= \text{Var} (x \cdot X) + \text{Var} (y \cdot Y).
\]

Therefore using the fact that \( (X, Y) \) is a Gaussian random vector,

\[
\mathbb{E} [e^{ix \cdot X} e^{iy \cdot Y}] = \mathbb{E} [e^{i(x \cdot X + y \cdot Y)}]
\]

\[
= \exp \left( -\frac{1}{2} \text{Var} (x \cdot X + y \cdot Y) + \mathbb{E} (x \cdot X + y \cdot Y) \right)
\]

\[
= \exp \left( -\frac{1}{2} \text{Var} (x \cdot X) + i\mathbb{E} (x \cdot X) - \frac{1}{2} \text{Var} (y \cdot Y) + i\mathbb{E} (y \cdot Y) \right)
\]

\[
= \mathbb{E} [e^{ix \cdot X}] \cdot \mathbb{E} [e^{iy \cdot Y}],
\]

and because \( x \) and \( y \) were arbitrary, we may conclude from Theorem [13.15](#) that \( X \) and \( Y \) are independent. 

\[\blacksquare\]
Corollary 13.29. 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

$$
E \left[ e^{i(x,y)'(X,Y)} \right] = E \left[ e^{i(x,y)'X Y} \right] = E \left[ e^{iX \cdot Y} \right] = E \left[ e^{iX \cdot Y} \right] = \exp \left( -\frac{1}{2} \text{Var} (x \cdot X) + i \text{E} (x \cdot X) \right) \times \exp \left( -\frac{1}{2} \text{Var} (y \cdot Y) + i \text{E} (y \cdot Y) \right)
$$

which shows that $(X, Y)$ is again Gaussian. ■

Notation 13.30 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 $X_1 + X_2 + \ldots + X_n$ for $X_1 + \ldots + X_n$ under the additional assumption that the $\{X_i\}_{i=1}^n$ are independent.

Corollary 13.31. 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:

$$
\text{Var} (S_n) = \sum_{i=1}^n \text{Var} (X_i) \quad \text{and} \quad \mathbb{E} S_n = \sum_{i=1}^n \mathbb{E} X_i ,
$$
i.e.

$$
X_1 + X_2 + \ldots + X_n \overset{d}{=} \mathbb{N} \left( \sum_{i=1}^n \text{Var} (X_i), \sum_{i=1}^n \mathbb{E} X_i \right).
$$

In particular if $\{X_i\}_{i=1}^\infty$ are i.i.d. Gaussian random variables with $\mathbb{E} X_i = \mu$ and $\sigma^2 = \text{Var} (X_i)$, then

$$
\frac{S_n - \mu}{\frac{1}{n} \sum_{i=1}^n \text{Var} (X_i)} \overset{d}{=} \mathbb{N} \left( 0, \frac{\sigma^2}{n} \right) \quad \text{and} \quad \frac{S_n - \mu}{\sqrt{\text{Var} (S_n)}} \overset{d}{=} \mathbb{N} \left( 0, 1 \right).
$$

Equation 13.8 is a very special case of the central limit theorem while Eq. 13.8 leads to a very special case of the strong law of large numbers, see Corollary 13.32.

Proof. The fact that $S_n$, $\frac{S_n}{n} - \mu$, and $\frac{S_n - \mu}{\sigma \sqrt{n}}$ are all Gaussian follows from Corollary 13.31 and Lemma 12.42 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 10.14 states that if $\{A_n\}_{n=1}^\infty$ are measurable sets, then

$$
\sum_{n=1}^\infty P (A_n) < \infty \implies P (\{A_n \text{ i.o.}\}) = 0. \quad (13.10)
$$

Corollary 13.32. Let $\{X_i\}_{i=1}^\infty$ be i.i.d. Gaussian random variables with $\mathbb{E} X_i = \mu$ and $\sigma^2 = \text{Var} (X_i)$. 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

$$
\left| \frac{S_n}{n} - \mu \right| \leq n^{-\alpha} \quad \text{for} \quad n \geq N_\alpha.
$$

In particular, $\lim_{n \to \infty} \frac{S_n}{n} = \mu$ a.s.

Proof. Let $Z \overset{d}{=} N (0, 1)$ so that $\frac{\sigma}{\sqrt{n}} Z \overset{d}{=} N \left( 0, \frac{\alpha^2}{n} \right)$. From the Eq. (13.8) and Eq. (10.50),

$$
P \left( \left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) = P \left( \left| \frac{\sigma}{\sqrt{n}} Z \right| \geq \varepsilon \right) = P \left( \left| Z \right| \geq \frac{\sqrt{n} \varepsilon}{\sigma} \right) \leq \exp \left( -\frac{1}{2} \left( \frac{\sqrt{n} \varepsilon}{\sigma} \right)^2 \right) = \exp \left( -\frac{\varepsilon^2}{2\sigma^2 n} \right).
$$

Taking $\varepsilon = n^{-\alpha}$ with $1 - 2\alpha > 0$, it follows that

$$
\sum_{n=1}^\infty P \left( \left| \frac{S_n}{n} - \mu \right| \geq n^{-\alpha} \right) \leq \sum_{n=1}^\infty \exp \left( -\frac{1}{2} \frac{\varepsilon^2}{n} \right) < \infty \quad \text{and so by the first Borel-Cantelli lemma,}
$$

$$
P \left( \left\{ \left| \frac{S_n}{n} - \mu \right| \geq n^{-\alpha} \text{ i.o.} \right\} \right) = 0.
$$

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. ■

13.4 Summing independent random variables

Exercise 13.4. Suppose that $X \overset{d}{=} N (0, a^2)$ and $Y \overset{d}{=} N (0, b^2)$ and $X$ and $Y$ are independent. Show by direct computation using the formulas for the distributions of $X$ and $Y$ that $X + Y \overset{d}{=} N (0, a^2 + b^2)$. 

Page: 191 | job: prob | macro: svmobob.cls | date/time: 2-Nov-2018/7:31
Exercise 13.33 (Gamma Distribution Sums). We will show here that
\( \Gamma(k, \theta) + \Gamma(l, \theta) = \Gamma(k + l, \theta) \). In Exercise 10.13 you showed if \( k, \theta > 0 \) then
\[
E[e^{tX}] = (1 - \theta t)^{-k} \quad \text{for } t < \theta^{-1}
\]
where \( X \) is a positive random variable with \( X \overset{d}{=} \Gamma(k, \theta) \), i.e.
\[
(X, P) \overset{d}{=} \Gamma(k, \theta)
\]
Therefore it follows from Exercise 10.13 that \( X + Y \overset{d}{=} \Gamma(k + l, \theta) \).

Let us now make the change of variables, \( x = x + y \) so that \( dx dy = dx dz \), to find,
\[
E[f(X + Y)] = \frac{1}{\theta^{k+l} \Gamma(k) \Gamma(l)} \int_{0}^{\infty} f(z) z^{k-1} e^{-z/\theta} dz.
\] (13.11)
To finish the proof we must now do that \( x \) integral and show,
\[
\int_{0}^{\infty} x^{k-1} e^{-x/\theta} \, dx = \frac{\Gamma(k) \Gamma(l)}{\Gamma(k + l)}
\]
(13.12)
Combing these results with Eq. (13.11) then shows,
\[
E[f(X + Y)] = \frac{B(k, l)}{\theta^{k+l} \Gamma(k) \Gamma(l)} \int_{0}^{\infty} f(z) z^{k+l-1} e^{-z/\theta} dz.
\] (13.13)
Since we already know that
\[
\int_{0}^{\infty} z^{k+l-1} e^{-z/\theta} \, dz = \theta^{k+l} \Gamma(k + l)
\]
it follows by taking \( f = 1 \) in Eq. (13.13) that
\[
1 = \frac{B(k, l)}{\theta^{k+l} \Gamma(k) \Gamma(l)} \frac{\Gamma(k) \Gamma(l)}{\Gamma(k + l)}
\]
which implies,
\[
B(k, l) = \frac{\Gamma(k) \Gamma(l)}{\Gamma(k + l)}
\] (13.14)
Therefore, using this back in Eq. (13.13) implies
\[
E[f(X + Y)] = \frac{1}{\theta^{k+l} \Gamma(k + l)} \int_{0}^{\infty} f(z) z^{k+l-1} e^{-z/\theta} \, dz
\]
from which it follows that \( X + Y \overset{d}{=} \Gamma(k + l, \theta) \).
Let us pause to give a direct verification of Eq. (13.14). By definition of the gamma function,
\[
\Gamma(k) \Gamma(l) = \int_{\mathbb{R}^2} x^{k-1}e^{-x}y^{l-1}e^{-y}dxdy = \int_{\mathbb{R}^2} x^{k-1}y^{l-1}e^{-(x+y)}dxdy.
\]

Making the change of variables, \(x = x + y\) it follows,
\[
\Gamma(k) \Gamma(l) = \int_{0 \leq x \leq z < \infty} x^{k-1}(z-x)^{l-1}e^{-z}dxdz.
\]

Now make the change of variables, \(x = zt\) to find,
\[
\Gamma(k) \Gamma(l) = \int_{0}^{\infty} dt \int_{0}^{1} dz (zt)^{k-1}(z-tz)^{l-1}z
\]
\[
= \int_{0}^{\infty} e^{-z}z^{k+l-1}dz \int_{0}^{1} t^{k-1}(1-t)^{l-1}dt
\]
\[
= \Gamma(k+l) B(k,l).
\]

Definition 13.35 (Beta distribution). The \(\beta\) – distribution is
\[
d\mu_{x,y}(t) = \frac{t^{x-1}(1-t)^{y-1}dt}{B(x,y)}.
\]

Observe that
\[
\int_{0}^{1} t^{2}d\mu_{x,y}(t) = B(x+1,y) = \frac{\Gamma(x+1)\Gamma(y)}{\Gamma(x+y+1)} = \frac{x}{x+y}.
\]

Exercise 13.6. Use the following outline to give a proof of Theorem 13.36
1. First show that \(x^p \leq 1 + x^4\) for all \(x \geq 0\) and \(1 \leq p \leq 4\). Use this to conclude:
\[
E |X_n|^p \leq 1 + E |X_n|^4 < \infty \text{ for } 1 \leq p \leq 4.
\]

Thus \(\gamma := E [X_n - \mu]^4\) and the standard deviation \((\sigma^2)\) of \(X_n\) defined by,
\[
\sigma^2 := E[X_n^2] - \mu^2 = E[(X_n - \mu)^2] < \infty,
\]

are finite constants independent of \(n\).
2. Show for all \(n \in \mathbb{N}\) that
\[
E \left[ \left( \frac{S_n}{n} - \mu \right)^4 \right] = \frac{1}{n^4} (n\gamma + 3n(n-1)\sigma^4)
\]
\[
= \frac{1}{n^2} [n^{-1} \gamma + 3 (1 - n^{-1}) \sigma^4].
\]

(Thus \(\frac{S_n}{n} \to \mu\) in \(L^4(P).\))
3. Use item 2. and Chebyshev’s inequality to show
\[
P \left( \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \leq \frac{n^{-1}\gamma + 3 (1 - n^{-1}) \sigma^4}{\varepsilon^4n^2}.
\]

4. Use item 3. and the first Borel Cantelli Lemma 13.14 to conclude \(\lim_{n \to \infty} \frac{S_n}{n} = \mu\) a.s.

13.6 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 13.37. Suppose that \(\{U, X, Y\}\) are independent random variables such that \(E \left[ |U|^3 + |X|^3 + |Y|^3 \right] < \infty\), \(EX = EY\), and \(EX^2 = EY^2\). Then for every function, \(f \in C^3(\mathbb{R})\) with \(M := \sup_{x \in \mathbb{R}} |f^{(3)}(x)| < \infty\),
\[
|E [f(U + X) - f(U + Y)]| \leq \frac{M}{3!} \cdot E \left[ |U|^3 + |X|^3 + |Y|^3 \right].
\]

Proof. By Taylor’s theorem for \(u, \Delta \in \mathbb{R}\) there exists \(c(u, \Delta)\) between \(u\) and \(\Delta\) so that
Then it follows that

\[ |\mathbb{E} f(N) - \mathbb{E} f(\bar{S}_n)| = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty \]  

which is to say, $\bar{S}_n$ is “close” in distribution to $N$, which we abbreviate by $\bar{S}_n \overset{d}{=} N$ for large $n$. (It should be noted that the estimate in Eq. (13.17) 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 \( \{Y_k, N_k\}_{k=1}^n \) be a collection of independent random variables such that

\[ Y_k \overset{d}{=} \bar{X}_k = \frac{X_k - \mathbb{E} X_k}{\sigma(S_n)} \quad \text{and} \quad N_k \overset{d}{=} N(0, \text{Var}(Y_k)) \quad \text{for } 1 \leq k \leq n. \]

Let \( S_n^Y = Y_1 + \cdots + Y_n \overset{d}{=} \bar{S}_n \) and \( T_n := N_1 + \cdots + N_n. \) Since

\[ \sum_{k=1}^n \text{Var}(N_k) = \sum_{k=1}^n \text{Var}(Y_k) = \frac{1}{\sigma(S_n)^2} \sum_{k=1}^n \text{Var}(X_k - \mathbb{E} X_k) = \frac{1}{\sigma(S_n)^2} \sum_{k=1}^n \text{Var}(X_k) = 1, \]

it follows by Corollary 13.31 that \( T_n \overset{d}{=} N(0,1) \).

To compare \( \mathbb{E} f(\bar{S}_n) \) with \( \mathbb{E} f(N) \) we may compare \( \mathbb{E} f(S_n^Y) \) with \( \mathbb{E} f(T_n) \) which we will do by interpolating between \( S_n^Y \) and \( T_n \). To this end, for \( 0 \leq k \leq n \), let

\[ V_k := N_1 + \cdots + N_k + Y_{k+1} + \cdots + Y_n \]

with the convention that \( V_0 = T_n \) and \( V_n = S_n^Y \). Then by a telescoping series argument, it follows that

\[ \mathbb{E} f(T_n) - \mathbb{E} f(S_n^Y) = \mathbb{E} [f(V_n) - f(V_0)] = \sum_{k=1}^n \mathbb{E} [f(V_k) - f(V_{k-1})]. \] 

(13.20)

If we set

\[ U_k := N_1 + \cdots + N_{k-1} + Y_{k+1} + \cdots + Y_n \]

we have \( V_k = U_k + N_k \) and \( V_{k-1} = U_k + Y_k \). Therefore it follows by an application of Lemma 13.37 that

\[ |\mathbb{E} [f(V_k) - f(V_{k-1})]| \leq M |\mathbb{E} |N_k|^3 + \mathbb{E} |Y_k|^3| \]

which combined with Eq. (13.20) shows

\[ \mathbb{E} f(T_n) - \mathbb{E} f(S_n^Y) = \sum_{k=1}^n \mathbb{E} [f(V_k) - f(V_{k-1})] = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty \]
\[
\left| E f (N) - E f (\bar{S}_n) \right| = \left| E f (T_n) - E f (S_n^Y) \right| \\
\leq \sum_{k=1}^{n} \left| E [f (V_k) - f (V_{k-1})] \right| \\
\leq \frac{M}{3^3} \sum_{k=1}^{n} \left( E |N_k|^3 + E |Y_k|^3 \right). \tag{13.21}
\]

We further have the estimates:
\[
E |Y_k|^3 = E \left| \left( \frac{X_k - EX_k}{\sigma (S_n)} \right)^3 \right| \leq \frac{C}{\sigma (S_n)^3} \quad \text{and}
\]
\[
E |N_k|^3 = \sqrt{8/\pi} \cdot \text{Var} (N_k)^{3/2} = \sqrt{8/\pi} \cdot \text{Var} (Y_k)^{3/2} = \frac{\lambda^3}{2} \cdot (E Y_k^2)^{3/2} \\
\leq \sqrt{8/\pi} \cdot E |Y_k|^3 \leq \frac{C}{\sigma (S_n)^3},
\]

wherein we have used Eq. (13.16) and Jensen’s (or Hölder’s) inequality (see Chapter 15 below) in the last two lines, or the last inequality. Combining these estimates with Eq. (13.21) gives the estimate in Eq. (13.17).

**Corollary 13.40.** Suppose that \( \{X_n\}_{n=1}^\infty \) is a sequence of i.i.d. random variables in \( L^3 (\mathbb{P}) \), \( C := \text{E} |X_1 - EX_1|^3 < \infty \), \( S_n := X_1 + \cdots + X_n \), and \( N \overset{d}{=} N (0, 1) \). Then for every function, \( f \in C^3 (\mathbb{R}) \) with \( M := \sup_{x \in \mathbb{R}} |f^{(3)} (x)| < \infty \) we have
\[
|E f (N) - E f (\bar{S}_n)| \leq \frac{M}{3^3} \left( 1 + \sqrt{8/\pi} \right) \frac{C}{\text{Var} (X_1)^{3/2}}. \tag{13.22}
\]
(This is a specialized form of the “Berry–Esseen theorem.”)

**Example 13.41.** Let \( \lambda > 0 \). Recall from Exercise 13.35 if \( \{X_i\}_{i=1}^n \) are i.i.d. with \( X_i \overset{d}{=} \text{Poi} (\lambda / n) \), then \( X = S_n = \sum_{i=1}^{n} X_i \overset{d}{=} \text{Poi} (\lambda) \). From Exercise 10.5 we know that
\[
EX = \lambda = \text{Var} (X)
\]
and so we might expect from Corollary 13.40 that \( N \overset{d}{=} \bar{S}_n \) should be close to the standard normal and in fact by letting \( n \to \infty \) we might errantly conclude that \( N \overset{d}{=} N (0, 1) \). To see what is going on let us consider the estimate in Eq. (13.22), which states,
\[
|E f (N) - E f (X - \lambda / \sqrt{\lambda})| \leq \frac{M}{3^3} \left( 1 + \sqrt{8/\pi} \right) \frac{1}{\text{Var} (X_1)^{3/2}} \\
= \frac{M}{3^3} \left( 1 + \sqrt{8/\pi} \right) \frac{1}{\lambda^{3/2}} \\
= \frac{M}{3 \lambda^{3/2}} \left( 1 + \sqrt{8/\pi} \right) \left( \frac{1}{\lambda} \right)^{3/2} \tag{13.23}
\]
To see what is happening as \( n \to \infty \) we need to first show
\[
EX^3 = \lambda^3 + 3\lambda^2 + \lambda.
\]
To see this recall from Exercise 10.5 if \( X \overset{d}{=} \text{Poi} (\lambda) \) then
\[
EX = \lambda = \text{Var} (X) \quad \text{and} \quad E [z^X] = \sum_{k=0}^{\infty} z^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda (z - 1)}.
\]
Using \( EX^k = \left( \frac{d}{dz} \right)^k \bigg|_{z=1} e^{\lambda (z - 1)} \), we learn
\[
\left( \frac{d}{dz} \right) e^{\lambda (z - 1)} = \lambda z e^{\lambda (z - 1)} \quad \implies \quad EX = \lambda,
\]
\[
\left( \frac{d^2}{dz^2} \right) e^{\lambda (z - 1)} = (\lambda^2 z^2 + \lambda z) e^{\lambda (z - 1)} \quad \implies \quad EX^2 = \lambda^2 + \lambda
\]
\[
\left( \frac{d^3}{dz^3} \right) e^{\lambda (z - 1)} = \left[ (\lambda^2 z^2 + \lambda z) \lambda z + 2\lambda z^2 + \lambda z \right] e^{\lambda (z - 1)} \quad \implies \quad EX^3 = \lambda^3 + 3\lambda^2 + \lambda.
\]
We now see that
\[
|E [X - \lambda]|^3 \leq E [X - \lambda]^3 \leq E (X + \lambda)^3.
\]
Since
\[
E (X + \lambda)^3 = E \left[ X^3 + \left( \frac{3}{2} \right) (\lambda X^2 \pm \lambda^2 X) \right] = \lambda + O (\lambda^2),
\]
we conclude that
\[
E |X - EX|^3 = \lambda + O (\lambda^2) \quad \text{as} \quad \lambda \downarrow 0.
\]
Using this result in Eq. (13.23) then shows with \( K = \frac{M}{3 \lambda^{3/2}} \left( 1 + \sqrt{8/\pi} \right) \) that
Taking Eq. (13.25) with sequence of i.i.d. random variables in \( f \)

\[
\left| \mathbb{E} f (N) - \mathbb{E} f \left( \frac{X - \lambda}{\sqrt{\lambda}} \right) \right| \leq \frac{K}{\lambda^{3/2}} n \cdot \left| \frac{\lambda}{n} + O \left( \frac{\lambda/n^2}{\lambda} \right) \right|.
\]

Letting \( n \to \infty \) implies

\[
\left| \mathbb{E} f (N) - \mathbb{E} f \left( \frac{X - \lambda}{\sqrt{\lambda}} \right) \right| \leq \frac{K}{\sqrt{\lambda}}.
\]

(13.24)

We certainly cannot conclude from this estimate that \( X \) is Gaussian which is good as it is not. On the other hand if \( X_\lambda \overset{d}{=} \text{Poi}(\lambda) \), then Eq. (13.24) 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 13.43 below.

By a slight modification of the proof of Theorem 13.39 we have the following central limit theorem.

**Theorem 13.42 (A CLT proof w/o Fourier).** Suppose that \( \{X_n\}_{n=1}^\infty \) is a sequence of i.i.d. random variables in \( L^2 (P) \), \( S_n := X_1 + \cdots + X_n \), and \( N \overset{d}{=} N (0,1) \). Then for every function, \( f \in C^2 (\mathbb{R}) \) with \( M := \sup_{x \in \mathbb{R}} |f^{(2)}(x)| < \infty \) 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:

\[
f (x + \Delta) - f (x) = f' (x) \Delta + \frac{1}{2} f'' (x) \Delta^2 + r (x, \Delta) \Delta^2
\]

where

\[
r (x, \Delta) = \int_0^1 [f'' (x + t\Delta) - f'' (x)] (1 - t) \, dt.
\]

Taking Eq. (13.25) with \( \Delta \) replaced by \( \delta \) and subtracting the results then implies

\[
f (x + \Delta) - f (x + \delta) = f' (x) (\Delta - \delta) + \frac{1}{2} f'' (x) (\Delta^2 - \delta^2) + \rho (x, \Delta, \delta)
\]

where now,

\[
\rho (x, \Delta, \delta) = r (x, \Delta) \Delta^2 - r (x, \delta) \delta^2.
\]

Since \( f'' \) is uniformly continuous it follows that

\[
\varepsilon (\Delta) := \frac{1}{2} \sup \{|f'' (x + t\Delta) - f (x)| : x \in \mathbb{R} \text{ and } 0 \leq t \leq 1\} \to 0
\]

Thus we may conclude that

\[
|r (x, \Delta)| \leq \int_0^1 |f'' (x + t\Delta) - f'' (x)| (1 - t) \, dt \leq \int_0^1 2 \varepsilon (\Delta) (1 - t) \, dt = \varepsilon (\Delta).
\]

and therefore that

\[
|\rho (x, \Delta, \delta)| \leq \varepsilon (\Delta) \Delta^2 + \varepsilon (\delta) \delta^2.
\]

So working just as in the proof of Theorem 13.39 we may conclude,

\[
|\mathbb{E} f (N) - \mathbb{E} f (\bar{S}_n)| \leq \sum_{k=1}^n \mathbb{E} |R_k|
\]

where now,

\[
|R_k| = \varepsilon (N_k) N^2_k + \varepsilon (Y_k) Y^2_k.
\]

Since the \( \{Y_k\}_{k=1}^n \) and the \( \{N_k\}_{k=1}^n \) are i.i.d. now it follows that

\[
|\mathbb{E} f (N) - \mathbb{E} f (\bar{S}_n)| \leq n \cdot \mathbb{E} \left[ \varepsilon (N_1) N^2_1 + \varepsilon (Y_1) Y^2_1 \right].
\]

Since \( \text{Var} (S_n) = n \cdot \text{Var} (X_1) \), we have \( Y_1 = \frac{X_1 - \mathbb{E} X_1}{\sqrt{\text{Var} (X_1)}} \), \( \text{Var} (N_1) = \text{Var} (Y_1) = \frac{1}{n} \) and therefore \( N_1 = \frac{1}{\sqrt{n}} N \). Combining these observations shows,

\[
|\mathbb{E} f (N) - \mathbb{E} f (\bar{S}_n)| \leq \mathbb{E} \left[ \varepsilon \left( \frac{1}{\sqrt{n}} N \right) N^2 + \varepsilon \left( \frac{X_1 - \mathbb{E} X_1}{\sqrt{\text{Var} (X_1)}} \right) \frac{\left(X_1 - \mathbb{E} X_1\right)^2}{\sigma^2 (X_1)} \right]
\]

which goes to zero as \( n \to \infty \) by the DCT.

**Lemma 13.43.** Suppose that \( \{W\} \cup \{W_n\}_{n=1}^\infty \) is a collection of random variables such that \( \lim_{n \to \infty} \mathbb{E} f (W) = \mathbb{E} f (W) \) for all \( f \in C^\infty (\mathbb{R}) \), 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 24.32 below it suffices to show \( \lim_{n \to \infty} \mathbb{E} f (W_n) = \mathbb{E} f (W) \) for all \( f \in C^\infty (\mathbb{R}) \). For such a function, \( f \in C^\infty (\mathbb{R}) \), we may find \( f_k \in C^\infty (\mathbb{R}) \) 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,

\[
|\mathbb{E} f (W) - \mathbb{E} f (W_n)| \leq |\mathbb{E} f (W) - \mathbb{E} f_k (W)| + |\mathbb{E} f_k (W) - \mathbb{E} f_k (W_n)| + |\mathbb{E} f_k (W_n) - \mathbb{E} f (W_n)|
\]

\[
\leq |\mathbb{E} f (W) - f_k (W)| + |\mathbb{E} f_k (W) - \mathbb{E} f_k (W_n)| + |\mathbb{E} f_k (W_n) - f (W_n)|
\]

\[
\leq 2 \varepsilon_k + |\mathbb{E} f_k (W) - f (W_n)|.
\]

1 We will eventually prove this standard real analysis fact later in the course.
Therefore it follows that
\[
\limsup_{n \to \infty} |\mathbb{E} f(W) - \mathbb{E} f(W_n)| \leq 2\varepsilon_k + \limsup_{n \to \infty} |\mathbb{E} f_k(W) - \mathbb{E} f_k(W_n)|
\]
\[= 2\varepsilon_k \to 0 \text{ as } k \to \infty.
\]

Corollary 13.44. 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 13.39 or Theorem 13.42 we have \(S_n \Rightarrow N(0, 1)\), i.e. \(\lim_{n \to \infty} \mathbb{E} f(S_n) = \mathbb{E} f(N(0, 1))\) 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 25 and 26 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.

13.7 The Second Borel-Cantelli Lemma

Lemma 13.45. If \(0 \leq x \leq \frac{1}{2}\), then
\[
e^{-2x} \leq 1 - x \leq e^{-x}.
\]
(13.26)
Moreover, the upper bound in Eq. (13.26) is valid for all \(x \in \mathbb{R}\).

Proof. The upper bound follows by the convexity of \(e^{-x}\), see Figure 13.1.

Fig. 13.1. A graph of \(1 - x\) and \(e^{-x}\) showing that \(1 - x \leq e^{-x}\) for all \(x\).

For the lower bound we use the convexity of \(\varphi(x) = e^{-2x}\) to conclude that the

\[
\prod_{n=1}^{N} (1 - a_n) := \lim_{N \to \infty} \prod_{n=1}^{N} (1 - a_n).
\]

The limit exists since, \(\prod_{n=1}^{N} (1 - a_n)\) decreases as \(N\) increases.

Corollary 13.46. If \(\{a_n\}_{n=1}^{\infty} \subset [0, 1]\) then
\[
\lim_{N \to \infty} \prod_{n=N}^{\infty} (1 - a_n) = \begin{cases} 1 & \text{if } \sum_{n=1}^{\infty} a_n < \infty \\ 0 & \text{if } \sum_{n=1}^{\infty} a_n = \infty. \end{cases}
\]


Exercise 13.8. 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, \(\iff\), holds even if \(a_n = 1\) is allowed.

Lemma 13.47 (Second Borel-Cantelli Lemma). Suppose that \(\{A_n\}_{n=1}^{\infty}\) are independent sets. If
\[
\sum_{n=1}^{\infty} P(A_n) = \infty,
\]
then
\[
P(\{A_n \ i.o.\}) = 1.
\]
Combining this with the first Borel Cantelli Lemma gives the (Borel) Zero-One law,
\[
P(A_n \ i.o.) = \begin{cases} 
0 & \text{if} \sum_{n=1}^{\infty} P(A_n) < \infty \\
1 & \text{if} \sum_{n=1}^{\infty} P(A_n) = \infty.
\end{cases}
\]

**Proof.** We are going to prove Eq. (13.28) by showing,
\[
0 = P(\{A_n \ i.o.\}^c) = P(\{A_n^c \ \text{a.a.}\}) = P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c).
\]
Since \(\cap_{k \geq n} A_k^c \uparrow \cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c\) as \(n \to \infty\) and \(\cap_{k=1}^{m} A_k^c \downarrow \cap_{n=1}^{\infty} \cup_{k \geq n} A_k\) as \(m \to \infty\),
\[
P(\cap_{n=1}^{\infty} \cap_{k \geq n} A_k^c) = \lim_{n \to \infty} P(\cap_{k \geq n} A_k^c) = \lim_{n \to \infty} \lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c).
\]
Making use of the independence of \(\{A_k\}_{k=1}^{\infty}\) and hence the independence of \(\{A_k^c\}_{k=1}^{\infty}\), we have
\[
P(\cap_{m \geq k \geq n} A_k^c) = \prod_{m \geq k \geq n} P(A_k^c) = \prod_{m \geq k \geq n} (1 - P(A_k)).
\]
Using the upper estimate in Eq. (13.26) along with Eq. (13.29) shows
\[
P(\cap_{m \geq k \geq n} A_k^c) \leq \prod_{m \geq k \geq n} e^{-P(A_k)} = \exp\left(-\sum_{k=n}^{m} P(A_k)\right).
\]
Using Eq. (13.27), we find from the above inequality that
\[
\lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = 0
\]
and hence
\[
P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c) = \lim_{n \to \infty} \lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = \lim_{n \to \infty} 0 = 0
\]
Note: we could also appeal to Exercise 13.8 above to give a proof of the Borel Zero-One law without appealing to the first Borel Cantelli Lemma. \(\blacksquare\)

**Proposition 13.49 (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
\[
\sum_{n=1}^{\infty} P(X_1 > \alpha^{-1} c_n) = \infty
\]
while
\[
\sum_{n=1}^{\infty} P(X_1 > \alpha c_n) < \infty.
\]
Then
\[
\limsup_{n \to \infty} \frac{X_n}{c_n} = 1 \quad \text{a.s.}
\]

**Proof.** By the second Borel-Cantelli Lemma, Eq. (13.30) implies
\[
P(X_n \leq \alpha c_n \ i.o. \ n) = 1
\]
from which it follows that
\[
\limsup_{n \to \infty} \frac{X_n}{c_n} \geq \alpha^{-1} \quad \text{a.s.}
\]
Taking \(\alpha = \alpha_k = 1 + 1/k\), we find
\[
P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} \geq 1\right) = P\left(\cap_{k=1}^{\infty} \left(\limsup_{n \to \infty} \frac{X_n}{c_n} \geq \frac{1}{\alpha_k}\right)\right) = 1.
\]
Similarly, by the first Borel-Cantelli lemma, Eq. (13.31) implies
\[
P(X_n \leq \alpha c_n \ i.o. \ n) = 0
\]
or equivalently,
\[
P(X_n \leq \alpha c_n \ a.a. \ n) = 1.
\]
That is to say,
\[
\limsup_{n \to \infty} \frac{X_n}{c_n} \leq \alpha \quad \text{a.s.}
\]
and hence working as above,
\[
P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} \leq 1\right) = P\left(\cap_{k=1}^{\infty} \left(\limsup_{n \to \infty} \frac{X_n}{c_n} \leq \alpha_k\right)\right) = 1.
\]
Hence,
\[
P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} = 1\right) = P\left(\left(\limsup_{n \to \infty} \frac{X_n}{c_n} \geq 1\right) \cap \left(\limsup_{n \to \infty} \frac{X_n}{c_n} \leq 1\right)\right) = 1.
\] \(\blacksquare\)
Example 13.50. Let \( \{X_n\}_{n=1}^{\infty} \) be i.i.d. standard normal random variables. Then by Mills’ ratio (see Lemma 10.71),

\[
P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi \alpha c_n}} e^{-\alpha^2 c_n^2/2}.
\]

Now, suppose that we take \( c_n \) so that

\[
e^{-\alpha^2 c_n^2/2} = \frac{1}{n} \implies c_n = \sqrt{2\ln(n)}.
\]

It then follows that

\[
P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi \alpha \sqrt{2\ln(n)}}} e^{-\alpha^2 \ln(n)} = \frac{1}{2\alpha \sqrt{\pi \ln(n)}} n^{-\alpha^2}
\]

and therefore

\[
\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) = \infty \text{ if } \alpha < 1
\]

and

\[
\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) < \infty \text{ if } \alpha > 1.
\]

Hence an application of Proposition 13.49 shows

\[
\limsup_{n \to \infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \text{ a.s.}
\]

Example 13.51. Let \( \{E_n\}_{n=1}^{\infty} \) be a sequence of i.i.d. random variables with exponential distributions determined by

\[
P(E_n > x) = e^{-(\alpha x)} \text{ or } P(E_n \leq x) = 1 - e^{-(\alpha x)}.
\]

(Observe that \( P(E_n \leq 0) = 0 \)) so that \( E_n > 0 \) a.s. Then for \( c_n > 0 \) and \( \alpha > 0 \), we have

\[
\sum_{n=1}^{\infty} P(E_n > \alpha c_n) = \sum_{n=1}^{\infty} e^{-\alpha c_n} = \sum_{n=1}^{\infty} (e^{-\alpha c_n})^\alpha.
\]

Hence if we choose \( c_n = \ln n \) so that \( e^{-\alpha c_n} = 1/n \), 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 if \( \alpha > 1 \). So by Proposition 13.49 it follows that

\[
\limsup_{n \to \infty} \frac{E_n}{\ln n} = 1 \text{ a.s.}
\]

Example 13.52. * Suppose now that \( \{X_n\}_{n=1}^{\infty} \) are i.i.d. distributed by the Poisson distribution with intensity, \( \lambda \), i.e.

\[
P(X_1 = k) = \frac{\lambda^k e^{-\lambda}}{k!}.
\]

In this case we have

\[
P(X_1 \geq n) = e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} \geq \frac{\lambda^n}{n!} e^{-\lambda}
\]

and

\[
\sum_{k=n}^{\infty} \frac{\lambda^k}{k!} = \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{n!}{k!} \lambda^{k-n} = \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k = \frac{\lambda^n}{n!}.
\]

Thus we have shown that

\[
\frac{\lambda^n}{n!} e^{-\lambda} \leq P(X_1 \geq n) \leq \frac{\lambda^n}{n!}.
\]

Thus in terms of convergence issues, we may assume that

\[
P(X_1 \geq x) \sim \frac{\lambda^x}{x!} \sim \frac{\lambda^x}{\sqrt{2\pi \lambda x}} x^x
\]

wherein we have used Stirling’s formula,

\[
x! \sim \sqrt{2\pi \lambda} x^{x-1/2}.
\]

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

\[
x = \ln n
\]

and upon iteration we find,

\[
\frac{\ln n}{\ln \left( \frac{\ln n}{x} \right)} = \frac{\ln n}{\ell_2(n) - \ell_2(x)} = \frac{\ln n}{\ell_2(n) - \ell_2 \left( \frac{\ln n}{x} \right)} = \frac{\ln n}{\ell_2(n) - \ell_3(n) + \ell_3(x)}.
\]
where \( \ell_k = \ln \circ \ln \circ \cdots \circ \ln \). Since, \( x \leq \ln \left( n \right) \), it follows that \( \ell_3 \left( x \right) \leq \ell_3 \left( n \right) \) and hence
\[
x = \frac{\ln \left( n \right)}{\ell_2 \left( n \right) + O \left( \ell_3 \left( n \right) \right)} = \frac{\ln \left( n \right)}{\ell_2 \left( n \right)} \left( 1 + O \left( \frac{\ell_3 \left( n \right)}{\ell_2 \left( n \right)} \right) \right).
\]
Thus we are led to take \( c_n := \frac{\ln \left( n \right)}{\ell_2 \left( n \right)} \). We then have, for \( \alpha \in (0, \infty) \) that
\[
\left( \alpha c_n \right)^{\alpha c_n} = \exp \left( \alpha c_n \left( \ln \alpha + \ln c_n \right) \right)
\]
\[
= \exp \left( \frac{\ln \left( n \right)}{\ell_2 \left( n \right)} \left( \ln \alpha + \ell_2 \left( n \right) - \ell_3 \left( n \right) \right) \right)
\]
\[
= \exp \left( \frac{\ln \alpha - \ell_3 \left( n \right)}{\ell_2 \left( n \right)} + 1 \right) \ln \left( n \right)
\]
\[
= n^{\alpha \left( 1 + \frac{1}{\ell_2 \left( n \right)} \right)}
\]
where
\[
\varepsilon_n \left( \alpha \right) := \frac{\ln \alpha - \ell_3 \left( n \right)}{\ell_2 \left( n \right)}.
\]
Hence we have
\[
P \left( X_1 \geq \alpha c_n \right) \sim \frac{\lambda^{\alpha c_n}}{\sqrt{2 \pi \alpha c_n} e^{-\alpha c_n}} \left( \alpha c_n \right)^{\alpha c_n} \sim \frac{1}{\sqrt{2 \pi c_n} n^{\alpha \left( 1 + \varepsilon_n \left( \alpha \right) \right)}}.
\]
Since
\[
\ln \left( \lambda / e \right)^{\alpha c_n} = \alpha c_n \ln \left( \lambda / e \right) = \alpha \frac{\ln n}{\ell_2 \left( n \right)} \ln \left( \lambda / e \right) = \ln n^{\alpha \ln \left( \lambda / e \right)}
\]
it follows that
\[
\left( \lambda / e \right)^{\alpha c_n} = n^{\alpha \ln \left( \lambda / e \right)}
\]
Therefore,
\[
P \left( X_1 \geq \alpha c_n \right) \sim \frac{n^{\alpha \ln \left( \lambda / e \right)}}{\sqrt{\ln \left( n \right)}} \frac{1}{n^{\alpha \left( 1 + \varepsilon_n \left( \alpha \right) \right)}} = \sqrt{\frac{\ell_2 \left( n \right)}{\ln \left( n \right)}} \frac{1}{n^{\alpha \left( 1 + \delta_n \left( \alpha \right) \right)}}
\]
where \( \delta_n \left( \alpha \right) \to 0 \) as \( n \to \infty \). From this observation, we may show,
\[
\sum_{n=1}^{\infty} P \left( X_1 \geq \alpha c_n \right) < \infty \text{ if } \alpha > 1 \text{ and }
\sum_{n=1}^{\infty} P \left( X_1 \geq \alpha c_n \right) = \infty \text{ if } \alpha < 1
\]
and so by Proposition 13.49 we may conclude that
\[
\limsup_{n \to \infty} \frac{X_n}{\ln \left( n \right) / \ell_2 \left( n \right)} = 1 \text{ a.s.}
\]

13.8 Kolmogorov and Hewitt-Savage Zero-One Laws

Let \( \left\{ X_n \right\}_{n=1}^{\infty} \) be a sequence of random variables on a measurable space, \( \left( \Omega, B \right) \). Let \( B_n := \sigma \left( X_1, \ldots, X_n \right), B_\infty := \sigma \left( X_1, X_2, \ldots \right), \) and \( T_n := \sigma \left( X_{n+1}, X_{n+2}, \ldots \right) \), and \( \mathcal{T} := \bigcap_{n=1}^{\infty} \mathcal{T}_n \subset B_\infty \). We call \( \mathcal{T} \) the tail \( \sigma \) - field and events, \( A \in \mathcal{T} \), are called tail events.

**Example 13.53.** Let \( S_n := X_1 + \cdots + X_n \) and \( \left\{ b_n \right\}_{n=1}^{\infty} \subset (0, \infty) \) such that \( b_n \uparrow \infty \). Here are some example of tail events and tail measurable random variables:

1. \( \left\{ \sum_{n=1}^{\infty} X_k \text{ converges} \right\} \in \mathcal{T} \). Indeed,
   \[
   \left\{ \sum_{k=1}^{\infty} X_k \text{ converges} \right\} \subset \mathcal{T}
   \]
   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. \( \left\{ \lim_{n \to \infty} X_n \text{ exists} \in \mathbb{R} \right\} = \left\{ \limsup_{n \to \infty} X_n = \liminf_{n \to \infty} X_n \in \mathcal{T} \right\} \) and similarly,
   \[
   \left\{ \lim_{n \to \infty} \frac{S_n}{b_n} \text{ exists in } \mathbb{R} \right\} \subset \mathcal{T}
   \]
   and
   \[
   \left\{ \lim_{n \to \infty} \frac{S_n}{b_n} \text{ exists in } \mathbb{R} \right\} = \left\{ -\infty < \limsup_{n \to \infty} \frac{S_n}{b_n} = \liminf_{n \to \infty} \frac{S_n}{b_n} < \infty \right\} \in \mathcal{T}.
   \]

4. \( \left\{ \lim_{n \to \infty} \frac{S_n}{b_n} = 0 \right\} \in \mathcal{T} \). Indeed, for any \( k \in \mathbb{N} \),
   \[
   \lim_{n \to \infty} \frac{S_n}{b_n} = \lim_{n \to \infty} \frac{X_{k+1} + \cdots + X_n}{b_n}
   \]
   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 13.54.** Let \( \left( \Omega, B, P \right) \) be a probability space. A \( \sigma \) - field, \( \mathcal{F} \subset B \) is **almost trivial** iff \( P \left( \mathcal{F} \right) = \left\{ 0, 1 \right\} \), i.e. \( P \left( A \right) \in \left\{ 0, 1 \right\} \) for all \( A \in \mathcal{F} \).

The following conditions on a sub-\( \sigma \)-algebra, \( \mathcal{F} \subset B \) are equivalent; 1) \( \mathcal{F} \) is almost trivial, 2) \( P \left( A \right) = P \left( A \right)^2 \) for all \( A \in \mathcal{F} \), and 3) \( \mathcal{F} \) is independent of itself. For example if \( \mathcal{F} \) is independent of itself, then \( P \left( A \right) = P \left( A \right)^2 = P \left( A \right) P \left( A \right) \) for all \( A \in \mathcal{F} \) which implies \( P \left( A \right) = 0 \) or 1. If \( \mathcal{F} \) is almost trivial and \( A, B \in \mathcal{F} \), then \( P \left( A \cap B \right) = P \left( A \right) P \left( B \right) \) if \( P \left( A \right) = P \left( B \right) = 1 \) and \( P \left( A \cap B \right) = 0 = P \left( A \right) P \left( B \right) \) if either \( P \left( A \right) = 0 \) or \( P \left( B \right) = 0 \). Therefore \( \mathcal{F} \) is independent of itself.
Lemma 13.55. Suppose that $X : \Omega \rightarrow \mathbb{R}$ is a random variable which is $\mathcal{F}$ measurable, where $\mathcal{F} \subset \mathcal{B}$ is almost trivial. Then there exists $c \in \mathbb{R}$ 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 $x \in \mathbb{R}$, $\{X \leq x\} \in \mathcal{F}$ and therefore, $P(X \leq x)$ is either 0 or 1. Since the function, $F(x) := P(X \leq x) \in (0, 1)$ 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 $X : \Omega \rightarrow \mathbb{R}$ is an integrable random variable, we know that $X$ is independent of itself and therefore $X^2$ is integrable and $\mathbb{E}X^2 = (\mathbb{E}X)^2 = c$. Thus it follows that $\mathbb{E}(X - c)^2 = 0$, i.e. $X = c$ a.s.

For general $X : \Omega \rightarrow \mathbb{R}$, let $X_M := (M \land X) \lor (-M)$, 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, $X = \lim_{M \rightarrow \infty} X_M \overset{a.s.}{=} \lim_{M \rightarrow \infty} c = c$.

Proposition 13.56 (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 13.55 have probability either 0 or 1.

Proof. For each $n \in \mathbb{N}$, $\mathcal{T} \subset \sigma(X_{n+1}, X_{n+2}, \ldots)$ which is independent of $\mathcal{B}_n := \sigma(X_1, \ldots, X_n)$. Therefore $\mathcal{T}$ is independent of $\cup \mathcal{B}_n$ which is a multiplicative system. Therefore $\mathcal{T}$ 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 13.57. Keeping the assumptions in Proposition 13.56 and let $(b_n)_{n=1}^\infty \subset (0, \infty)$ such that $b_n \uparrow \infty$. Then $\limsup_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} X_n \leq \limsup_{n \rightarrow \infty} \frac{b_n}{c}$, and $\liminf_{n \rightarrow \infty} \frac{b_n}{c}$ are all constant almost surely. In particular, either $P\left(\left\{\lim_{n \rightarrow \infty} \frac{b_n}{c} \text{ exists}\right\}\right) = 0$ or $P\left(\left\{\lim_{n \rightarrow \infty} \frac{b_n}{c} \text{ exists}\right\}\right) = 1$ and in the latter case $\lim_{n \rightarrow \infty} \frac{b_n}{c} = c$ a.s.

Example 13.58. Suppose that $\{A_n\}_{n=1}^\infty$ are independent sets and let $X_n := 1_{A_n}$ for all $n$ and $\mathcal{T} := \cap_{n \geq 1} \sigma(X_n, X_{n+1}, \ldots)$. 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 13.47) tells when $P(\{A_n \text{ i.o.}\}) = 0$ and when it is 1 depending on whether $\sum_{n=1}^\infty P(A_n)$ is finite or infinite respectively.

13.8 Kolmogorov and Hewitt-Savage Zero-One Laws

13.8.1 Hewitt-Savage Zero-One Law

In this subsection, let $\Omega := \mathbb{R}^\infty = \mathbb{R}^n$ and $X_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, and $\mathcal{B} := \sigma(X_1, X_2, \ldots)$ be the product $\sigma$ - algebra on $\Omega$. We say a permutation (i.e. a bijective map on $\mathbb{N}$), $\pi : \mathbb{N} \rightarrow \mathbb{N}$ is finite if $\pi(n) = n$ for a.a. $n$. Define $T_\pi : \Omega \rightarrow \Omega$ by $T_\pi(\omega) = (\omega_{\pi_1}, \omega_{\pi_2}, \ldots)$. Since $X_1 \circ T_\pi(\omega) = \omega_{\pi_1} = X_{\pi_1}(\omega)$ 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 $(\mathbb{R}, \mathcal{B})$ and let $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $(\Omega = \mathbb{R}^n, \mathcal{B})$. Then $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with $\text{Law}_{P}(X_n) = \mu$ for all $n$. If $\pi : \mathbb{N} \rightarrow \mathbb{N}$ is a finite permutation and $A_1 \in \mathcal{B}_R$ for all $i$, then

$$T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \ldots) = A_{\pi^{-1}1} \times A_{\pi^{-1}2} \times \ldots$$

Since sets of the form, $A_1 \times A_2 \times A_3 \times \ldots$, form a $\pi$ – system generating $\mathcal{B}$ and

$$P \circ T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \ldots) = \prod_{i=1}^\infty \mu(A_{\pi^{-1}i}) = \prod_{i=1}^\infty \mu(A_1) = P(A_1 \times A_2 \times A_3 \times \ldots),$$

we may conclude that $P \circ T_\pi^{-1} = P$.

Definition 13.59. 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 13.60 (Hewitt-Savage Zero-One Law). Let $\mu$ be a probability measure on $(\mathbb{R}, \mathcal{B})$ and $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $(\Omega = \mathbb{R}^n, \mathcal{B})$ so that $\{X_n\}_{n=1}^\infty$ (recall that $X_n(\omega) = \omega_n$) is an i.i.d. sequence with $\text{Law}_{P}(X_n) = \mu$ for all $n$. Then $\mathcal{S}$ is $P$ - almost trivial.

Proof. Let $B \in \mathcal{S}$, $f = 1_B$, and $g = G(X_1, X_2, \ldots, X_n)$ - measurable function such that $\sup_{\omega \in \Omega} |g(\omega)| \leq 1$. Further let $\pi$ be a finite permutation such that $\{\pi_1, \ldots, \pi_n\} \cap \{1, 2, \ldots, n\} = \emptyset$ - for example we could take $\pi(j) = j + n$, $\pi(j + n) = j$ for $j = 1, 2, \ldots, n$, and $\pi(j + 2n) = j + 2n$ for all $j \in \mathbb{N}$. Then $g \circ T_\pi = G(X_{\pi_1}, X_{\pi_2}, X_{\pi_3}, \ldots)$ is independent of $g$ and therefore

$$(Eg)^2 = Eg \cdot E[g \circ T_\pi] = E[g \cdot g \circ T_\pi].$$

Since $f \circ T_\pi = 1_{T_\pi^{-1}(B)} = 1_B = f$, it follows that $Ef = Ef^2 = Ef \cdot f \circ T_\pi$ and therefore,
That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.

That is $0$.
Example 13.63. Continue the notation of Example 13.62 and further assume that $\Lambda \subset \mathbb{R}$ and let $X_i : \Omega \to \mathbb{R}$ 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 13.62.

Alternatively, from Exercise 13.62, 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 $f_i : A \to \mathbb{R}$. Taking $A_i \subset A$ and $f_i := 1_{A_i}$ in the above identity shows that
\[
P(X_1 \in A_1, \ldots , X_n \in A_n) = \mathbb{E}_P \left[ \prod_{i=1}^n 1_{A_i}(X_i) \right] = \prod_{i=1}^n \mathbb{E}_P [1_{A_i}(X_i)]
\]
\[
= \prod_{i=1}^n P(X_i \in A_i)
\]
as desired.

Theorem 13.64 (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 $\Lambda = \{0, 1, 2, \ldots, n\}$ on $((0, 1), \mathcal{B}, m)$ such that
\[
m(\{X_1 = i_1, \ldots , X_k = i_k\}) = q_{i_1} \cdots q_{i_k}
\]
for all $i_1, i_2, \ldots , i_k \in \{0, 1, 2, \ldots, n\}$ and all $k \in \mathbb{N}$. (See Example 13.19 above and Theorem 13.68 below for the general case of this theorem.)

Proof. For $i = 0, 1, \ldots, n$, let $\sigma_{-1} = 0$ and $\sigma_j := \sum_{i=0}^j q_i$ and for any interval, $(a, b]$, let
\[
T_i((a, b]) := (a + \sigma_{i-1} (b-a), a + \sigma_j (b-a)].
\]
Given $i_1, i_2, \ldots , i_k \in \{0, 1, 2, \ldots, n\}$, let
\[
J_{i_1, i_2, \ldots , i_k} := T_{i_k} (T_{i_{k-1}} (\ldots T_{i_1} ((0, 1])))
\]
and define $\{X_k\}_{k=1}^\infty$ on $(0, 1]$ by
\[
X_k := \sum_{i_1, i_2, \ldots , i_k \in \{0, 1, 2, \ldots, n\}} i_k 1_{J_{i_1, i_2, \ldots , i_k}}
\]
see Figure 13.3. Repeated applications of Corollary 9.27 shows the functions, $X_k : (0, 1] \to \mathbb{R}$ are measurable.

Observe that
\[
m(T_i((a, b])) = q_i (b-a) = q_i m((a, b]),
\]
and so by induction,
\[
m(J_{i_1, i_2, \ldots , i_k}) = q_{i_1} q_{i_{k-1}} \cdots q_{i_1}.
\]
The reader should convince herself/himself that $\{X_1 = i_1, \ldots , X_k = i_k\} = J_{i_1, i_2, \ldots , i_k}$ and therefore, we have
\[
m(\{X_1 = i_1, \ldots , X_k = i_k\}) = m(J_{i_1, i_2, \ldots , i_k}) = q_{i_1} q_{i_{k-1}} \cdots q_{i_1}
\]
as desired.

Corollary 13.65 (Independent variables on product spaces). Suppose $\Lambda = \{0, 1, 2, \ldots, n\}$, $q_i > 0$ with $\sum_{i=0}^n q_i = 1$, $\Omega = \Lambda^\infty = \Lambda^N$, 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
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 13.13 to show that random variables constructed in Example 6.26 or Theorem 13.64 fit the bill.

Proposition 13.67. 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 $P(U \leq x) = (0 \vee x) \wedge 1$, 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

$$\left\{ U < \frac{1}{2} \right\} = \{X_1 = 0\},$$

$$\left\{ U < \frac{1}{4} \right\} = \{X_1 = 0, X_2 = 0\} \quad \text{and}$$

$$\left\{ \frac{1}{2} \leq U < \frac{3}{4} \right\} = \{X_1 = 1, X_2 = 0\}$$

and hence that

$$\left\{ U < \frac{3}{4} \right\} = \left\{ U < \frac{1}{2} \right\} \cup \left\{ \frac{1}{2} \leq U < \frac{3}{4} \right\} = \{X_1 = 0\} \cup \{X_1 = 1, X_2 = 0\}.$$  

From these identities, it follows that

$$P(U < 0) = 0, \quad P\left( U < \frac{1}{4} \right) = \frac{1}{4}, \quad P\left( U < \frac{1}{2} \right) = \frac{1}{2}, \quad \text{and} \quad P\left( U < \frac{3}{4} \right) = \frac{3}{4}.$$  

More generally, we claim that if $x = \sum_{j=1}^{n} \varepsilon_j 2^{-j}$ with $\varepsilon_j \in \{0,1\}$, then

$$P(U < x) = x.$$  

The proof is by induction on $n$. Indeed, we have already verified (13.35) when $n = 1,2$. Suppose we have verified (13.35) up to some $n \in \mathbb{N}$ and let $x = \sum_{j=1}^{n} \varepsilon_j 2^{-j}$ and consider

$$P\left( U < x + 2^{-(n+1)} \right) = P(U < x) + P\left( x \leq U < x + 2^{-(n+1)} \right) = x + P\left( x \leq U < x + 2^{-(n+1)} \right).$$

Since

$$\left\{ x \leq U < x + 2^{-(n+1)} \right\} = \left[ \cap_{j=1}^{n} \{X_j = \varepsilon_j\} \right] \cap \{X_{n+1} = 0\}$$

we see that

$$P\left( x \leq U < x + 2^{-(n+1)} \right) = 2^{-(n+1)}$$

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 $P(U \leq x) = x$ for all $x \in (0,1)$. Hence we may conclude that

$$P(U \leq x) = (0 \vee x) \wedge 1.$$  

We may now show the existence of independent random variables with arbitrary distributions.

Theorem 13.68. 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 13.66, 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 13.67, \( U_i \) is uniformly distributed on \([0, 1]\).

Moreover by the grouping Lemma 13.20, \( \sigma \left( \{X_{i,j}\}_{j=1}^\infty \right) \) 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 \( G_i(y) = \inf \{x : F_i(x) \geq y\} \). Then according to Theorem 9.48, \( Y_i := G_i(U_i) \) 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. \( \blacksquare \)
The Standard Poisson Process

14.1 Poisson Random Variables

Recall from Exercise 10.5 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 $X \overset{d}{=} \text{Poi}(a)$. Let us also recall that

$$E [z^X] = \sum_{k=0}^{\infty} z^k \frac{a^k}{k!} e^{-a} = e^{az} e^{-a} = e^{a(z-1)}$$

and as in Exercise 10.5 we have $E X = a = \text{Var}(X)$.

**Lemma 14.1.** If $X \overset{d}{=} \text{Poi}(a)$ and $Y \overset{d}{=} \text{Poi}(b)$ and $X$ and $Y$ are independent, then $X + Y \overset{d}{=} \text{Poi}(a+b)$.

**Proof.** For $k \in \mathbb{N}_0$,

$$P (X + Y = k) = \sum_{l=0}^{k} P (X = l, Y = k-l) = \sum_{l=0}^{k} P (X = l) P (Y = k-l)$$

$$= \sum_{l=0}^{k} e^{-a} \frac{a^l}{l!} e^{-b} \frac{b^{k-l}}{(k-l)!} = e^{-(a+b)} \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} a^l b^{k-l}$$

$$= e^{-(a+b)} \frac{k!}{k!} (a+b)^k.$$ 

**Alternative Proof.** Notice that

$$E [z^{X+Y}] = E [z^X] E [z^Y] = e^{a(z-1)} e^{b(z-1)} = \exp ((a+b)(z-1)).$$

This suffices to complete the proof.  

**Lemma 14.2.** 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 14.1 for each $n \in \mathbb{N}$, $\sum_{i=1}^{n} N_i \overset{d}{=} \text{Pois} (\sum_{i=1}^{n} \lambda_i)$. Since for each $k \in \mathbb{N}_0$, $\{\sum_{i=1}^{n} N_i = k\} \downarrow \{N = k\}$ as $n \uparrow \infty$ we have

$$P (N = k) = \lim_{n \to \infty} P \left( \sum_{i=1}^{n} N_i = k \right) = \lim_{n \to \infty} \frac{\left( \sum_{i=1}^{n} \lambda_i \right)^k}{k!} e^{-\sum_{i=1}^{n} \lambda_i}$$

$$= \frac{\lambda^k}{k!} e^{-\lambda},$$

which shows $N \overset{d}{=} \text{Pois} (\lambda)$.

**Lemma 14.3.** 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 14.1 we see that $1 - e^{-\lambda} \geq \frac{1}{2} (1 \wedge \lambda)$ for all $\lambda \geq 0$. Therefore,

$$\sum_{i=1}^{\infty} P (N_i \geq 1) = \sum_{i=1}^{\infty} (1 - P (N_i = 0)) = \sum_{i=1}^{\infty} (1 - e^{-\lambda_i}) \geq \frac{1}{2} \sum_{i=1}^{\infty} \lambda_i \wedge 1 = \infty.$$
and so by the second Borel Cantelli Lemma, $P\left(\{N_i \geq 1 \text{ i.o.}\}\right) = 1$. From this it certainly follows that $\sum_{i=1}^{\infty} N_i = \infty$ a.s.

**Alternatively**, let $A_n = \lambda_1 + \cdots + \lambda_n$, then

$$P\left(\sum_{i=1}^{\infty} N_i \geq k\right) \geq P\left(\sum_{i=1}^{n} N_i \geq k\right) = 1 - e^{-A_n} \sum_{i=0}^{k-1} \frac{A_n^i}{i!} \to 1 \text{ as } n \to \infty.$$ 

Therefore $P\left(\sum_{i=1}^{\infty} N_i \geq k\right) = 1$ for all $k \in \mathbb{N}$ and hence,

$$P\left(\sum_{i=1}^{\infty} N_i \geq \infty\right) = P\left(\cap_{k=1}^{\infty} \left\{\sum_{i=1}^{\infty} N_i \geq k\right\}\right) = 1.$$

### 14.2 Exponential Random Variables

Recall from Definition 10.67 that $T \overset{d}{=} E(\lambda)$ 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

$$E[e^{aT}] = \frac{1}{1 - a\lambda} \text{ for } a < \lambda. \quad (14.1)$$

$ET = \lambda^{-1}$ and $\text{Var}(T) = \lambda^{-2}$, and (see Theorem 10.68) 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 14.4.** Let $\{T_j\}_{j=1}^{\infty}$ be independent random variables such that $T_j \overset{d}{=} E(\lambda_j)$ with $0 < \lambda_j < \infty$ for all $j$. Then:

1. If $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ then $P\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 0$ (i.e. $P\left(\sum_{n=1}^{\infty} T_n < \infty\right) = 1$).
2. If $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $P\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 1$.

(By Kolmogorov’s zero-one law (see Proposition 13.56) it follows that $P\left(\sum_{n=1}^{\infty} T_n = \infty\right)$ is always either 0 or 1. We are showing here that $P\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 1$ iff $E\left[\sum_{n=1}^{\infty} T_n\right] = \infty$.)

**Proof.** 1. Since

$$E\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} E[T_n] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} T_n < \infty$ a.s., i.e. $P\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 0$.

2. By the DCT, independence, and Eq. (14.1) with $a = -1$,

$$E\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = \lim_{N \to \infty} E\left[e^{-\sum_{n=1}^{N} T_n}\right] = \lim_{N \to \infty} \prod_{n=1}^{N} E\left[e^{-T_n}\right]$$

$$= \lim_{N \to \infty} \prod_{n=1}^{N} \left(\frac{1}{1 + \lambda_n}\right) = \prod_{n=1}^{\infty} (1 - a_n)$$

where

$$a_n = 1 - \frac{1}{1 + \lambda_n^{-1}} = \frac{1}{1 + \lambda_n^{-1}}.$$

Hence by Exercise 13.8 $E\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ 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

$$E\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0 \text{ iff } e^{-\sum_{n=1}^{\infty} T_n} = 0 \text{ a.s.} \text{ or equivalently } \sum_{n=1}^{\infty} T_n = \infty \text{ a.s.}$$

### 14.2.1 Appendix: More properties of Exponential random Variables*

**Theorem 14.5.** 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 $q_k := \sum_{j \in I} q_k \in (0, \infty)$. Let $T := \inf_{k \in I} T_k$ 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 $\Lambda_n \subset I$ such that $\Lambda_n \uparrow I \setminus \{k\}$, then

$$P(K = k, T > t) = P(\cap_{j \neq k} \{T_j > T_k\}, T > t) = \lim_{n \to \infty} P(\cap_{j \in \Lambda_n} \{T_j > T_k\}, T > t)$$

$$= \lim_{n \to \infty} \int_{(0, \infty)^{\Lambda_n \cup \{k\}}} \prod_{j \in \Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n$$

$$\left(\{t_j\}_{j \in \Lambda_n}\right) q_k e^{-q_k t_k} dt_k$$

where $\mu_n$ is the joint distribution of $\{T_j\}_{j \in \Lambda_n}$. So by Fubini’s theorem,
\[ P(K = k; T > t) = \lim_{n \to \infty} \int_{t}^{\infty} q_k e^{-q_k t} dt_k \int_{0, \infty}^{A_n} \prod_{j \in A_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n \left( \{t_j\}_{j \in A_n} \right) \]
\[
\begin{align*}
  &= \lim_{n \to \infty} \int_{t}^{\infty} P(\cap_{j \in A_n} \{T_j > t_k\}) q_k e^{-q_k t} dt_k \\
  &= \int_{t}^{\infty} P(\cap_{j \neq k} \{T_j > \tau\}) q_k e^{-q_k \tau} d\tau \\
  &= \int_{t}^{\infty} \prod_{j \neq k} e^{-q_j \tau} q_k e^{-q_k \tau} d\tau = \int_{t}^{\infty} \prod_{j \in k} e^{-q_j \tau} q_k d\tau \\
  &= \int_{t}^{\infty} e^{-\sum_{j=1}^{\infty} q_j \tau} q_k d\tau = \int_{t}^{\infty} e^{-q \tau} q_k d\tau = \frac{q_k}{q} e^{-q t}. \quad (14.2)
\end{align*}
\]

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. (14.2) 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 14.6.** 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
\[
\begin{align*}
\mu P(S \leq t < S + R) &= \mu \int_{0}^{t} e^{-\lambda s} P(t < s + R) ds \\
&= \mu \int_{0}^{t} e^{-\lambda s} e^{-\mu(t-s)} ds \\
&= \mu e^{-\mu t} \int_{0}^{t} e^{-(\lambda-\mu)s} ds = \mu e^{-\mu t} \cdot \frac{1 - e^{-(\lambda-\mu)t}}{\lambda - \mu} \\
&= \mu \lambda \cdot \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu}
\end{align*}
\]
which is symmetric in the interchanged of \( \mu \) and \( \lambda \). Alternatively:

\[
\begin{align*}
P(S \leq t < S + R) &= \lambda \mu \int_{0}^{t} e^{-\lambda s} e^{-\mu(t-s)} ds \\
&= \lambda \mu \int_{0}^{t} ds \int_{t-s}^{\infty} e^{-\lambda s} e^{-\mu r} dr \\
&= \lambda e^{-\mu t} \int_{0}^{t} ds \int_{0}^{\infty} e^{-(\lambda-\mu)s} ds \\
&= \lambda e^{-\mu t} \frac{1 - e^{-(\lambda-\mu)t}}{\lambda - \mu}.
\end{align*}
\]

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
\[
\lambda P(R \leq t < R + S) = \mu \lambda \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu}.
\]

**Example 14.7.** 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 \( F_T(t) := P(T \leq t) = 1 - e^{-at} \). 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,
\[
E[e^{i\alpha T}] = \int_{0}^{\infty} ae^{-at} e^{i\lambda t} dt = \frac{a}{a - i\lambda} = \hat{\mu}(\lambda).
\]

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
\[
ET = \hat{\mu}'(0)i = a^{-1} \text{ and } E\tau^2 = \hat{\mu}''(0) \frac{i^2}{a^2} = \frac{2}{a^2},
\]
and hence \( \text{Var}(T) = \frac{2}{a^2} - (\frac{1}{a})^2 = a^{-2} \).
14.3 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 \( n \)th event to occur. Because of Theorem 14.4 we know that \( \lim_{n \to \infty} W_n = \infty \) a.s.

Definition 14.8 (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

\[
\Delta_n(T) := \{(w_1, \ldots, w_n) \in \mathbb{R}_n^n : 0 < w_1 < w_2 < \cdots < w_n < T\}
\]

and let

\[
\Delta_n := \cup_{T > 0} \Delta_n(T) = \{(w_1, \ldots, w_n) \in \mathbb{R}_n^n : 0 < w_1 < w_2 < \cdots < w_n < \infty\}.
\]

(We equip each of these spaces with their Borel \( \sigma \)-algebras.)

Exercise 14.1. Show \( m_n(\Delta_n(T)) = T^n/n! \) where \( m_n \) is Lebesgue measure on \( \mathbb{R}_n^n \).

Exercise 14.2. If \( n \in \mathbb{N} \) and \( g : \Delta_n \to \mathbb{R} \) bounded (non-negative) measurable, then

\[
\mathbb{E}[g(W_1, \ldots, W_n)] = \int_{\Delta_n} g(w_1, w_2, \ldots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \cdots dw_n.
\]

As a simple corollary we have the following direct proof of Example 13.34

Corollary 14.9. If \( n \in \mathbb{N} \), then \( W_n \overset{d}{=} \text{Gamma}(n, \lambda^{-1}) \).

Proof. Taking \( g(w_1, w_2, \ldots, w_n) = f(w_n) \) in Eq. (14.5) we find with the aid of Exercise 14.1 that

\[
\mathbb{E}[f(W_n)] = \int_{\Delta_n} f(w_n) \lambda^n e^{-\lambda w_n} dw_1 \cdots dw_n
= \int_0^\infty f(w) \lambda^n \frac{w^{n-1}}{(n-1)!} e^{-\lambda w} dw
\]

which shows that \( W_n \overset{d}{=} \text{Gamma}(n, \lambda^{-1}) \).

Corollary 14.10. If \( t \in \mathbb{R}_+ \) and \( f : \Delta_n(t) \to \mathbb{R} \) is a bounded (or non-negative) measurable function, then

\[
\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n]
= \lambda^n e^{-\lambda t} \int_{\Delta_n(t)} f(w_1, w_2, \ldots, w_n) dw_1 \cdots dw_n.
\]

Proof. Making use of the observation that \( \{N_t = n\} = \{W_n \leq t < W_{n+1}\} \), we may apply Eq. (14.3) at level \( n + 1 \) with

\[
g(w_1, w_2, \ldots, w_{n+1}) = f(w_1, w_2, \ldots, w_n) 1_{W_n \leq t < W_{n+1}}
\]

to learn

\[
\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n]
= \int_{0 < w_1 < \cdots < w_n < t < W_{n+1}} f(w_1, w_2, \ldots, w_n) \lambda^{n+1} e^{-\lambda w_n} dw_1 \cdots dw_n
\]

Exercise 14.3. Show \( N_t \overset{d}{=} \text{Poi}(\lambda t) \) for all \( t > 0 \).

Definition 14.11 (Order Statistics). Suppose that \( X_1, \ldots, X_n \) are non-negative random variables such that \( P(X_i = X_j) = 0 \) for all \( i \neq j \). The order statistics of \( X_1, \ldots, X_n \) are the random variables, \( \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n \) defined by

\[
\tilde{X}_k = \min \max \{X_i : i \in A\}
\]

where \( A \) always denotes a subset of \( \{1, 2, \ldots, n\} \) in Eq. (14.5).

The reader should verify that \( \tilde{X}_1 \leq \tilde{X}_2 \leq \cdots \leq \tilde{X}_n \), \( \{X_1, \ldots, X_n\} = \{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\} \) with repetitions, and that \( \tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n \) if \( \tilde{X}_i \neq \tilde{X}_j \) 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 14.4. Suppose that \( X_1, \ldots, X_n \) are non-negative random variables such that \( P(X_i = X_j) = 0 \) for all \( i \neq j \). Show;

\footnote{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

\[
\mathbb{E} \left[ f \left( \tilde{X}_1, \ldots, \tilde{X}_n \right) \right] = \sum_{\sigma \in S_n} \mathbb{E} \left[ f \left( X_{\sigma 1}, \ldots, X_{\sigma n} \right) : X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n} \right].
\]

where \( S_n \) is the permutation group on \( \{1, 2, \ldots, n\} \).

2. If we further assume that \( X_1, \ldots, X_n \) are i.i.d. random variables, then

\[
\mathbb{E} \left[ f \left( \tilde{X}_1, \ldots, \tilde{X}_n \right) \right] = n! \cdot \mathbb{E} \left[ f \left( X_1, \ldots, X_n \right) : X_1 < X_2 < \cdots < X_n \right].
\]

(14.6)

(It is important that \( f \left( \tilde{X}_1, \ldots, \tilde{X}_n \right) \) is not defined on the null set, \( \cup_{i \neq j} \{ X_i = X_j \} \).)

3. If \( f : \mathbb{R}^n_+ \to \mathbb{R} \) is a bounded (non-negative) measurable symmetric function (i.e., \( f \left( w_{\sigma 1}, \ldots, w_{\sigma n} \right) = f \left( w_1, \ldots, w_n \right) \) for all \( \sigma \in S_n \) and \( (w_1, \ldots, w_n) \in \mathbb{R}^n_+ \)) then

\[
\mathbb{E} \left[ f \left( \tilde{X}_1, \ldots, \tilde{X}_n \right) \right] = \mathbb{E} \left[ f \left( X_1, \ldots, X_n \right) \right].
\]

4. Suppose that \( Y_1, \ldots, Y_n \) is another collection of non-negative random variables such that \( P \left( Y_i = Y_j \right) = 0 \) for all \( i \neq j \) such that

\[
\mathbb{E} \left[ f \left( X_1, \ldots, X_n \right) \right] = \mathbb{E} \left[ f \left( Y_1, \ldots, Y_n \right) \right]
\]

for all bounded (non-negative) measurable symmetric functions from \( \mathbb{R}^n_+ \to \mathbb{R} \). Show that \( \left( \tilde{X}_1, \ldots, \tilde{X}_n \right) \overset{d}{=} \left( \tilde{Y}_1, \ldots, \tilde{Y}_n \right) \).

Hint: if \( g : \Delta_n \to \mathbb{R} \) is a bounded measurable function, define \( f : \mathbb{R}^n_+ \to \mathbb{R} \) by;

\[
f \left( y_1, \ldots, y_n \right) = \sum_{\sigma \in S_n} 1_{y_{\sigma 1} < y_{\sigma 2} < \cdots < y_{\sigma n}} g \left( y_{\sigma 1}, y_{\sigma 2}, \ldots, y_{\sigma n} \right)
\]

and then show \( f \) is symmetric.

**Exercise 14.5.** 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, \( \left( \tilde{U}_1, \ldots, \tilde{U}_n \right) \), of \( \left( U_1, \ldots, U_n \right) \) has the same distribution as \( (W_1, \ldots, W_n) \) given \( N_t = n \). (Thus, given \( N_t = n \), the collection of points, \( \{ W_1, \ldots, W_n \} \), has the same distribution as the collection of points, \( \{ U_1, \ldots, U_n \} \), in \( [0, t] \).)

**Theorem 14.12 (Joint Distributions).** If \( \{ A_i \}_{i=1}^k \subset \mathcal{B}_{[0,t]} \) is a partition of \( [0, t] \), then \( \{ N \left( A_i \right) \}_{i=1}^k \) are independent random variables and \( N \left( A \right) \overset{d}{=} \text{Poi} \left( \lambda t \left( A \right) \right) \) for all \( A \subset \mathcal{B}_{[0,t]} \) with \( m(A) < \infty \). In particular, if \( 0 < t_1 < t_2 < \cdots < t_n \), then \( \{ N_{t_i} - N_{t_{i-1}} \}_{i=1}^n \) are independent random variables and \( N_t - N_s \overset{d}{=} \text{Poi} \left( \lambda \left( t - s \right) \right) \) for all \( 0 \leq s < t < \infty \). (We say that \( \{ N_t \}_{t \geq 0} \) is a stochastic process with independent increments.)

**Proof.** If \( z \in \mathbb{C} \) and \( A \subset \mathcal{B}_{[0,t]} \), then

\[
z^{N(A)} = z \sum_{i=1}^n 1_A (W_i) \text{ on } \{ N_t = n \}.
\]

Let \( n \in \mathbb{N} \), \( z_t \in \mathbb{C} \), and define

\[
f \left( w_1, \ldots, w_n \right) = \sum_{i=1}^n 1_{A_1} (w_i) \cdots \sum_{k=1}^n 1_{A_k} (w_i)
\]

which is a symmetric function. On \( N_t = n \) we have,

\[
z_1^{N(A_1)} \cdots z_k^{N(A_k)} = f \left( W_1, \ldots, W_n \right)
\]

and therefore,

\[
\mathbb{E} \left[ z_1^{N(A_1)} \cdots z_k^{N(A_k)} | N_t = n \right] = \mathbb{E} \left[ f \left( W_1, \ldots, W_n \right) | N_t = n \right]
\]

\[
= \mathbb{E} \left[ f \left( U_1, \ldots, U_n \right) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i=1}^n 1_{A_1} (U_i) \cdots \sum_{k=1}^n 1_{A_k} (U_i) \right]
\]

\[
= \prod_{i=1}^n \mathbb{E} \left[ \left( z_{1}^{A_1} (U_i) \cdots z_k^{A_k} (U_i) \right) \right]
\]

\[
= \left( \mathbb{E} \left[ z_1^{A_1} (U_i) \cdots z_k^{A_k} (U_i) \right] \right)^n
\]

\[
= \left( \frac{1}{t} \sum_{i=1}^k m \left( A_i \right) \cdot z_t \right)^n
\]

wherein we have made use of the fact that \( \{ A_1 \}_{i=1}^n \) is a partition of \( [0, t] \) so that

\[
z_1^{A_1} (U_i) \cdots z_k^{A_k} (U_i) = \sum_{i=1}^k z_{1}^{A_i} (U_i).
\]

Thus it follows that
From this result it follows that
\[ P \left( N(A) = n \right) = P \left( \sum_{l=0}^{n} 1_{A_l} = n \right) = P \left( \sum_{l=0}^{n} 1_{A_l} = n | N_t = n \right) \cdot P(N_t = n) \]
\[ = \sum_{n=0}^{\infty} \left( \frac{1}{n!} \prod_{i=1}^{k} \left( \frac{m(A_i)}{a_i} \right)^{a_i} \right) \cdot \frac{(\lambda t)^n}{n!} e^{-\lambda t} \]
\[ = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \lambda \sum_{i=1}^{k} m(A_i) \cdot z_i \right) \cdot e^{-\lambda t} \]
\[ = \exp \left( \lambda \sum_{i=1}^{k} m(A_i) (z_i - t) \right) \]
\[ = \exp \left( \lambda \sum_{i=1}^{k} m(A_i) (z_i - 1) \right). \]

From this result it follows that \( \{ N(A_i) \}_{i=1}^{k} \) are independent random variables and \( N(A) = \text{Poi}(\lambda m(A)) \) for all \( A \in \mathcal{B}_R \) with \( m(A) < \infty \).

Alternatively; suppose that \( a_i \in \mathbb{N}_0 \) and \( n := a_1 + \cdots + a_k \), then
\[ P \left( N(A) = a_1, \ldots, N(A_k) = a_k | N_t = n \right) = P \left( \sum_{i=1}^{n} 1_{A_i} = a_i \text{ for } 1 \leq l \leq k \right) \]
\[ = \frac{n!}{a_1! \cdots a_k!} \prod_{i=1}^{k} \left[ \frac{m(A_i)}{a_i} \right]^{a_i} \]
\[ = \frac{n! \prod_{i=1}^{k} \left[ \frac{m(A_i)}{a_i} \right]^{a_i}}{a_1! \cdots a_k!} \]
and therefore,
\[ P \left( N(A) = a_1, \ldots, N(A_k) = a_k \right) = P \left( N(A) = a_1, \ldots, N(A_k) = a_k | N_t = n \right) \cdot P(N_t = n) \]
\[ = \frac{n! \prod_{i=1}^{k} \left[ \frac{m(A_i)}{a_i} \right]^{a_i}}{a_1! \cdots a_k!} \cdot \frac{(\lambda t)^n}{n!} e^{-\lambda t} \]
\[ = \prod_{i=1}^{k} \left[ \frac{m(A_i)}{a_i} \right]^{a_i} \cdot e^{-\lambda t} \lambda^n \]
\[ = \prod_{i=1}^{k} \frac{m(A_i)}{a_i} \cdot e^{-\lambda t} \lambda^n \]
which shows that \( \{ N(A_i) \}_{i=1}^{k} \) are independent and that \( N(A_i) \overset{d}{=} \text{Poi}(\lambda m(A_i)) \) for each \( i \).
1. Assume that \( \{ A_k \}_{k=1}^n \subset B_S \) is a measurable partition of \( S \) and show \( \{ N(A_k) \}_{k=1}^n \) are i.i.d. with \( N(A_k) = \text{Poi}(\mu(A_k)) \) for each \( k \).

2. Model your proof of item 1. on either of the proofs of Theorem 14.12.

14.4 Poisson Process Extras*

(This subsection still needs work!) In Definition 14.8 we really gave a construction of a Poisson process as defined in Definition 14.14. The goal of this section is to show that the Poisson process, \( \{ N_t \}_{t \geq 0} \), as defined in Definition 14.14 is uniquely determined and is essentially equivalent to what we have already done above.

Definition 14.14 (Poisson Process II). Let \( (\Omega, B, P) \) be a probability space and \( N_t : \Omega \rightarrow \mathbb{N}_0 \) be a random variable for each \( t \geq 0 \). We say that \( \{ N_t \}_{t \geq 0} \) is a Poisson process with intensity \( \lambda \) if: 1) \( N_0 = 0 \), 2) \( N_t - N_s \overset{d}{=} \text{Poi}(\lambda(t - s)) \) for all \( 0 \leq s < t < \infty \), 3) \( \{ N_t \}_{t \geq 0} \) has independent increments, and 4) \( t \rightarrow N_t(\omega) \) is right continuous and non-decreasing for all \( \omega \in \Omega \).

Let \( N_\infty(\omega) := \lim_{t \to \infty} N_t(\omega) \) and observe that \( N_\infty = \sum_{k=0}^\infty (N_k - N_{k-1}) = \infty \) a.s. by Lemma 14.3. Therefore, we may and do assume that \( N_\infty(\omega) = \infty \) for all \( \omega \in \Omega \).

Lemma 14.15. There is zero probability that \( \{ N_t \}_{t \geq 0} \) 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 \( t \rightarrow N_t(\omega) \) makes a jump of 2 or more for \( t \in [0, T] \). Then for all \( n \in \mathbb{N} \) we must have \( \omega \in \bigcup_{k=1}^n \{ N_k T - N_{k-1} T \geq 2 \} \). Therefore, \( P^*(\{ \omega : [0, T] \ni t \rightarrow N_t(\omega) \text{ has jump } \geq 2 \}) \)

\[
\leq \sum_{k=1}^n P\left( N_{kT} - N_{(k-1)T} \geq 2 \right) = \sum_{k=1}^n O(T^2/n^2) = O(1/n) \to 0
\]
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 14.16. Suppose that \( \{ N_t \}_{t \geq 0} \) is a Poisson process with intensity \( \lambda \) as in Definition 14.14.

\( W_n := \inf \{ t : N_t = n \} \) 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 \( \{ N_t \}_{t \geq 0} \).

Then \( \{ T_n := W_n - W_{n-1} \}_{n=1}^\infty \) are i.i.d. \( E(\lambda) \) - random variables. Thus the two descriptions of a Poisson process given in Definitions 14.8 and 14.14 are equivalent.

Proof. Suppose that \( J_i = (a_i, b_i] \) with \( b_i - a_{i+1} < \infty \) for all \( i \). We will begin by showing

\[
P(\bigcap_{i=1}^n \{ W_i \leq J_i \}) = \lambda^n \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} e^{-\lambda w_n} \, dw_n \quad (14.8)
\]

\[
= \lambda^n \int_{J_1 \times J_2 \times \cdots \times J_n} e^{-\lambda w_1} \, \cdots \, e^{-\lambda w_n} \, dw_n. \quad (14.9)
\]

To show this let \( K_i := (b_{i-1}, a_i] \) where \( b_0 = 0 \). Then \( \bigcap_{i=1}^n \{ W_i \leq J_i \} = \bigcap_{i=1}^n \{ N(K_i) = 0 \} \cap \bigcap_{i=1}^{n-1} \{ N(J_i) = 0 \} \cap \{ N(J_n) \geq 2 \} \) and therefore,

\[
P(\bigcap_{i=1}^n \{ W_i \leq J_i \}) = \prod_{i=1}^n e^{-\lambda m(K_i)} \cdot \prod_{i=1}^{n-1} e^{-\lambda m(J_i)} \lambda m(J_i) \cdot \left( 1 - e^{-\lambda m(J_n)} \right)
\]

\[
= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot e^{-\lambda m(J_n)} \int_{J_n} \lambda e^{-\lambda w_n} \, dw_n.
\]

We may now apply a \( \pi - \lambda \) - argument, using \( \sigma(\{ J_1 \times \cdots \times J_n \}) = B_{\Delta_n} \), to show
\[ E[g(W_1, \ldots, W_n)] = \int_{\Delta_n} g(w_1, \ldots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \ldots dw_n \]

holds for all bounded \( B_{\Delta_n}/B_\mathbb{R} \) measurable functions, \( g : \Delta_n \rightarrow \mathbb{R} \). Undoing the change of variables you made in Exercise 14.2 allows us to conclude that \( \{T_n\}_{n=1}^\infty \) are i.i.d. \( E(\lambda) \)-distributed random variables.
\(L^p - \text{spaces}\)

Let \((\Omega, \mathcal{B}, \mu)\) be a measure space and for \(0 < p < \infty\) and a measurable function \(f : \Omega \to \mathbb{C}\) let

\[
\|f\|_p := \left(\int_{\Omega} |f|^p \, d\mu\right)^{1/p}
\]

and when \(p = \infty\), let

\[
\|f\|_{\infty} = \inf \{a \geq 0 : \mu(|f| > a) = 0\}
\]

(15.1)

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 15.1.** 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:

\[
\|f\|_{\infty} = \inf \{a \geq 0 : \mu(|f(\omega)| \leq a)\}.
\]

(15.2)

### 15.1 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 15.2.** 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 \(L^p\) if \(f \in L^p\) and \(f_n \in L^p\) for all \(n\), and \(\lim_{n \to \infty} \|f_n - f\|_p = 0\).

**Definition 15.3.** 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 15.22 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 15.8 below.

**Lemma 15.4 (\(L^p\) - 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 (10.2),

\[
\mu(|f| \geq \varepsilon) = \mu(|f|^p \geq \varepsilon^p) \leq \frac{1}{\varepsilon^p} \int_{\Omega} |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_{p}^p
\]

and therefore if \(\{f_n\}\) is \(L^p\) - Cauchy, then

\[
\mu(|f_n - f_m| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \|f_n - f_m\|_p \to 0 \text{ as } m, n \to \infty
\]

showing \(\{f_n\}\) is \(L^0\) - Cauchy. A similar argument holds for the \(L^p\) - convergent case.

**Example 15.5.** 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, \((\mathbb{R}_+, \mathcal{B} = \mathcal{B}_{\mathbb{R}_+}, m)\).

1. Let \(f_n = \frac{1}{n} 1_{[0,n]}\) as in Figure 15.1. In this case \(f_n \not\to 0\) in \(L^1\) but \(f_n \to 0\) a.e., \(f_n \to 0\) in \(L^p\) for all \(p > 1\) and \(f_n \xrightarrow{\mu} 0\).
2. Let \( f_n = 1_{[n-1,n]} \) as in the figure below. Then \( f_n \to 0 \) a.e., yet \( f_n \not\to 0 \) in any \( L^p \) space or in measure.

3. Now suppose that \( f_n = n \cdot 1_{[0,1/n]} \) as in Figure 15.2. In this case \( f_n \to 0 \) a.e., \( f_n \not\to 0 \) but \( f_n \to 0 \) in \( L^1 \) or in any \( L^p \) for \( p \geq 1 \). Observe that \( \|f_n\|_p = n^{1-1/p} \) for all \( p \geq 1 \).

4. For \( n \in \mathbb{N} \) and \( 1 \leq k \leq n \), let \( g_{n,k} := 1_{[k/n-1/n]} \). Then define \( \{f_n\} \) as

\[
(f_1, f_2, f_3, \ldots) = (g_{1,1}, g_{2,1}, g_{2,2}, g_{3,1}, g_{3,2}, g_{3,3}, g_{4,1}, g_{4,2}, g_{4,3}, g_{4,4}, \ldots)
\]

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 \( f_n \not\to 0 \) but \( f_n \to 0 \) a.e. and \( f_n \not\to 0 \) in \( L^\infty \). In this case, \( \|g_{n,k}\|_p = (\frac{1}{n})^{1/p} \) for \( 1 \leq p < \infty \) while \( \|g_{n,k}\|_\infty = 1 \) for all \( n, k \).

15.2 Almost Everywhere and Measure Convergence

**Theorem 15.6 (Egorov: a.s. \( \Rightarrow \) 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 \( f_n \overset{\mu}{\to} f \) as \( n \to \infty \).

**Proof.** Let \( f_n \to f \) a.e. Then for all \( \varepsilon > 0 \),

\[
0 = \mu(\{|f_n - f| > \varepsilon \ \text{i.o.} \ n\}) = \lim_{N \to \infty} \mu \left( \bigcup_{n \geq N} \{|f_n - f| > \varepsilon\} \right)
\]

\[
\geq \limsup_{N \to \infty} \mu \left( \bigcup_{n \geq N} \{|f_n - f| > \varepsilon\} \right)
\]

from which it follows that \( f_n \overset{\mu}{\to} f \) as \( n \to \infty \).

We now prove that the convergence is uniform off a small exceptional set.

By Eq. (15.3), there exists an increasing sequence \( \{N_k\}_{k=1}^\infty \) such that \( \mu(E_k) < \varepsilon 2^{-k} \), where
Lemma 15.7. 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
\[
|a_m - a_n| = \left| \sum_{k=n+1}^{m-1} (a_{k+1} - a_k) \right| \leq \sum_{k=n}^{m-1} |a_{k+1} - a_k| \leq \sum_{k=n}^{\infty} \varepsilon_k := \delta_n.
\] So \( |a_m - a_n| \leq \delta_{\text{min}(m,n)} \to 0 \) as \( m, n \to \infty \), i.e. \( \{a_n\} \) is Cauchy. Let \( m \to \infty \) in (15.4) to find \( |a - a_n| \leq \delta_n \).

Theorem 15.8. Let \((\Omega, B, \mu)\) 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 \( f_n \overset{\mu}{\to} f \) and \( f_n \overset{\mu}{\to} g \) then \( f = g \) a.e.
2. If \( f_n \overset{\mu}{\to} f \) and \( g_n \overset{\mu}{\to} g \) then \( \lambda f_n \to \lambda f \) for all \( \lambda \in \mathbb{C} \) and \( f_n + g_n \overset{\mu}{\to} f + g \).
3. If \( f_n \overset{\mu}{\to} f \) 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 \( f_n \overset{\mu}{\to} f \).

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 \( f_n \overset{\mu}{\to} g \) and \( f_n \overset{\mu}{\to} f \) 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
\[
\{|f - g| > \varepsilon\} \subseteq \{|f - f_n| > \varepsilon/2\} \cup \{|g - f_n| > \varepsilon/2\},
\]
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
\[
\mu(|f - g| > 0) = \mu \left( \bigcup_{n=1}^{\infty} \left\{ |f - g| > \frac{1}{n} \right\} \right) \leq \sum_{n=1}^{\infty} \mu \left( |f - g| > \frac{1}{n} \right) = 0,
\]
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 \( f_n \overset{\mu}{\to} f, \varepsilon > 0 \) and \( m, n \in \mathbb{N} \), then \( |f_n - f_m| \leq |f - f_m| + |f_m - f| \).
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 \( \varepsilon_n = 2^{-n} \) 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
\[
F_N := \bigcup_{j \geq N} \{|g_{j+1} - g_j| > \varepsilon_j\} \text{ and } E := \cap_{N=1}^{\infty} F_N = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}.
\]
Since
\[
\mu(F_N) \leq \delta_N < \infty
\]
and \( F_N \downarrow E \) it follows 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 15.7 \( f(\omega) := \lim_{j \to \infty} g_j(\omega) \) exists. For \( \omega \in E \) we may define \( f(\omega) = 0 \).
5. Next we will show \( g_N \overset{\mu}{\to} f \) as \( N \to \infty \) where \( f \) and \( g_N \) are as above. If
\[
\omega \in F_N = \cap_{j \geq N} \{|g_{j+1} - g_j| \leq \varepsilon_j\},
\]
then
\[
|g_{j+1}(\omega) - g_j(\omega)| \leq \varepsilon_j \text{ for all } j \geq N.
\]
Another application of Lemma 15.7 shows \( |f(\omega) - g_j(\omega)| \leq \delta_j \) for all \( j \geq N \), i.e.
\[
1 \text{ Alternatively, } \mu(E) = 0 \text{ by the first Borel Cantelli lemma and the fact that } \sum_{j=1}^{\infty} \mu(|g_{j+1} - g_j| > \varepsilon_j) \leq \sum_{j=1}^{\infty} \varepsilon_j < \infty.
\]
that

\[ F_N^c \subset \cap_{j \geq N} \{ |f - g_j| \leq \delta_j \} \subset \{ |f - g_N| \leq \delta_N \}. \]

Therefore, by taking complements of this equation, \(|f - g_N| > \delta_N \} \subset F_N^c\) and hence

\[ \mu(|f - g_N| > \delta_N) \leq \mu(F_N^c) \leq \delta_N \to 0 \text{ as } N \to \infty \]

and in particular, \(g_N \xrightarrow{\mu} f\) as \(N \to \infty\).

With this in hand, it is straightforward to show \(f_n \xrightarrow{\mu} f\). 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 \(g_j \xrightarrow{\mu} f\).

**Corollary 15.9 (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

\[ |f_n| \leq g \text{ a.e., } f_n \xrightarrow{\mu} f, \; g_n \xrightarrow{\mu} g, \; \text{and} \; \int g_n \to \int g \text{ as } n \to \infty. \]

Then \(f \in L^1\) and \(\lim_{n \to \infty} \|f_n - f\|_1 = 0\), 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 15.8 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) \(\lim_{n \to \infty} \|f_n - f\|_1 \neq 0\) there exists \(\varepsilon > 0\) and a subsequence \(\{f_{n_k}\}\) of \(\{f_n\}\) such that

\[ \int |f_n - f| \geq \varepsilon \text{ for all } k. \]  \hspace{1cm} (15.5)

Using item 4. of Theorem 15.8 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 10.27 implies \(\lim_{k \to \infty} \int |f_n - f_{n_k}| = 0\) which contradicts Eq. (15.5).

**Exercise 15.1 (Fatou’s Lemma).** Let \((\Omega, \mathcal{B}, \mu)\) be a measure space. If \(f_n \geq 0\) and \(f_n \xrightarrow{\mu} f\), then \(\int f \, d\mu \leq \liminf_{n \to \infty} \int f_n \, d\mu\).

**Lemma 15.10.** Suppose \(1 \leq p \leq \infty\), \(\{f_n\}_{n=1}^{\infty} \subset L^p(\mu)\), and \(f_n \xrightarrow{\mu} f\), 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 \(\lim_{n \to \infty} \|f_n\|_p = \|f\|_p\) and \(f_n \xrightarrow{\mu} f\).

**Proof.** Choose a subsequence, \(g_k = f_{n_k}\), such that \(\lim_{n \to \infty} \|f_n\|_p = \lim_{k \to \infty} \|g_k\|_p\). By passing to a further subsequence if necessary, we may further assume that \(g_k \to f\) a.e. Therefore, by Fatou’s lemma,

\[ \|f\|_p = \int_{\Omega} |f| \, d\mu = \int_{\Omega} \liminf_{k \to \infty} |g_k| \, d\mu \leq \liminf_{k \to \infty} \int_{\Omega} |g_k| \, d\mu = \lim_{n \to \infty} \|f_n\|_p \]

which proves the first assertion.

If \(\|f - f_n\|_p \to 0\) as \(n \to \infty\), then by the triangle inequality, \(\|f\|_p = \lim_{n \to \infty} \|f_n\|_p\) which shows \(\int |f| \, d\mu = \int |f_n| \, d\mu\) if \(f_n \to f\) in \(L^p\). Chebyshev’s inequality implies \(f_n \xrightarrow{\mu} f\) if \(f_n \to f\) in \(L^p\).

Conversely if \(\lim_{n \to \infty} \|f_n\|_p = \|f\|_p < \infty\) and \(f_n \xrightarrow{\mu} f\), let \(F_n := |f - f_n|^p\) and \(G_n := 2^{-p} (|f|^p + |f_n|^p)\). Then \(f_\mu \xrightarrow{\mu} f\) if \(F_n \leq G_n \in L^1\), and \(f \xrightarrow{\mu} f\) if \(G_n \to G\) where \(G := 2^p |f|^p \in L^1\). Therefore, by Corollary 15.9 \(\|f - f_n\|^p = \int F_n \to \int 0 = 0\).

**Exercise 15.2.** 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 \(n, f_n \xrightarrow{\mu} f, \) and \(\int f_n \, d\mu \to \int f \, d\mu\). 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 15.11.** Suppose \((\Omega, \mathcal{B}, \mu)\) is a probability space and \(\{f_n\}_{n=1}^{\infty}\) is a sequence of measurable functions on \(\Omega\). Then \(\{f_n\}_{n=1}^{\infty}\) converges to \(f\) in probability iff every subsequence, \(\{f_{n_k}\}_{k=1}^{\infty}\) of \(\{f_n\}_{n=1}^{\infty}\) has a further subsequence, \(\{f_{n_{k_l}}\}_{l=1}^{\infty}\), which is almost surely convergent to \(f\).

**Proof.** If \(\{f_n\}_{n=1}^{\infty}\) is convergent and hence Cauchy in probability then any subsequence, \(\{f_{n_k}\}_{k=1}^{\infty}\) is also Cauchy in probability. Hence by item 4. of Theorem 15.8 there is a further subsequence, \(\{f_{n_{k_l}}\}_{l=1}^{\infty}\) of \(\{f_{n_k}\}_{k=1}^{\infty}\) which is convergent almost surely.

Conversely if \(\{f_{n_k}\}_{k=1}^{\infty}\) does not converge to \(f\) in probability, then there exists an \(\varepsilon > 0\) and a subsequence, \(\{n_k\}\) such that \(\inf\, \mu(\{f - f_{n_k}\geq \varepsilon\}) > 0\). Any subsequence of \(\{f_{n_k}\}\) would have the same property and hence can not be almost surely convergent because of Egorov’s Theorem 15.6.

\[2\] This is because \(|F_n| \geq \varepsilon\) iff \(|f - f_n| \geq \varepsilon^{1/p}\).
Corollary 15.12. 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 [15.8]

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

$$\lim_{k \to \infty} \psi(f_{n_k''}, g_{n_k''}) = \psi(f, g) \text{ a.s.}$$

which completes the proof. ■

**Example 15.13.** It is not possible to drop the assumption that $\mu(\Omega) < \infty$ in Corollary [15.12]. For example, let $\Omega = \mathbb{R}$, $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$, $\mu = m$ be Lebesgue measure, $f_n(x) = \frac{1}{2^n}$ and $g_n(x) = x^n = g(x)$ . Then $f_n \mu(\Omega) = 0$, $g_n \mu(\Omega)$ 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 $f_n \mu(\Omega)$ does not go to $0$ in measure as $n \to \infty$.

15.3 Jensen’s, Hölder’s and Minkowski’s Inequalities

**Theorem 15.14 (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 $\beta \in \mathbb{R}$ ($\beta = \varphi'(t)$ when $\varphi'(t)$ exists), be such that $\varphi(s) - \varphi(t) \geq \beta(s - t)$ for all $s \in (a, b)$. (See Lemma [15.60] and Figure [15.5] when $\varphi$ is $C^1$ and Theorem [15.63] 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

$$0 \leq \int_{\Omega} \varphi(f) d\mu - \varphi(t) = \int_{\Omega} \varphi(f) d\mu - \varphi(t)$$

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 15.15.** 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

$$\exp\left(\int_{\Omega} f d\mu\right) \leq \int_{\Omega} e^{t} d\mu, \quad (15.6)$$

and for $p \geq 1$,

$$\left(\int_{\Omega} |f| d\mu\right)^p \leq \left(\int_{\Omega} |f|^p d\mu\right).$$

**Example 15.16.** As a special case of Eq. (15.6), if $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i} = \sum_{i=1}^n \frac{s_i}{p_i}. \quad (15.7)$$

Indeed, we have applied Eq. (15.6) with $\Omega = \{1, 2, \ldots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i$. As a special case of Eq. (15.7), 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

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q. \quad (15.8)$$

(When $p = q = 1/2$, the inequality in Eq. 15.8 follows from the inequality, $0 \leq (s - t)^2$.)

As another special case of Eq. (15.7), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \ldots a_n} \leq \prod_{i=1}^n a_i. \quad (15.9)$$
Example 15.17. 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,

\[
\left( \int_{\Omega} |f|^p \, d\mu \right)^{\frac{q}{p}} \leq \int_{\Omega} (|f|^p)^{\frac{q}{p}} \, d\mu = \int_{\Omega} |f|^q \, d\mu
\]

from which it follows that \(|f|_p \leq |f|_q\). In particular, \(L^p(\mu) \subset L^q(\mu)\) for all \(0 < p < q < \infty\). See Corollary 15.31 for an alternative proof.

**Theorem 15.18 (Hölder’s inequality).** Suppose that \(1 \leq p \leq \infty\) and \(q := \frac{p}{p - 1}\), or equivalently \(p^{-1} + q^{-1} = 1\). If \(f\) and \(g\) are measurable functions then

\[
|f|_p \cdot |g|_q \leq |fg|_1 \leq \|f\|_p \cdot \|g\|_q.
\]

Assuming \(p \in (1, \infty)\) and \(|f|_p \cdot |g|_q < \infty\), equality holds in Eq. (15.10) iff \(|f|^p\) and \(|g|^q\) are linearly dependent as elements of \(L^1\) which happens iff

\[
|g|^q \|f\|_p^p = \|g\|_q \|f\|_p^p \ a.e.
\]

(15.11)

**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 \(|g|_p = 0\) or \(|f|_p \cdot |g|_q < \infty\), Eq. (15.10) 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. (15.8) gives,

\[
\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} |f|^p + \frac{1}{q} |g|^q
\]

with equality iff \(|g|/\|g\|_q = |f|^{-p} / \|f\|_p^{-1} = |f|^{p/q} / \|f\|_p^{p/q}\), i.e. \(|g|^q \|f\|_p^p = \|g\|^q \|f\|^p\). Integrating Eq. (15.12) implies

\[
\frac{|fg|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1
\]

with equality if Eq. (15.11) holds. The proof is finished since it is easily checked that equality holds in Eq. (15.10) when \(|f|^p = c \cdot |g|^q\) of \(|g|^q = c \cdot |f|^p\) for some constant \(c\).

Example 15.19. Suppose that \(a_k \in \mathbb{C}\) for \(k = 1, 2, \ldots, n\) and \(p \in [1, \infty)\), then

\[
\left| \sum_{k=1}^{n} a_k \right|^p \leq n^{p-1} \sum_{k=1}^{n} |a_k|^p.
\]

Indeed, by Hölder’s inequality applied using the measure space, \{1, 2, …, n\} equipped with counting measure, we have

\[
\left( \sum_{k=1}^{n} |a_k|^p \right)^{\frac{q}{p}} \leq \sum_{k=1}^{n} |a_k|^q \leq \left( \sum_{k=1}^{n} a_k \right)^{p/q} \left( \sum_{k=1}^{n} a_k \right)^{q/p}
\]

where \(q = p - 1/p\). Taking the \(p\)th – power of this inequality then gives, Eq. (15.14).

**Theorem 15.20 (Generalized Hölder’s inequality).** Suppose that \(f_i : \Omega \to \mathbb{C}\) are measurable functions for \(i = 1, \ldots, n\) and \(p_1, \ldots, p_n\) and \(r\) are positive numbers such that \(\sum_{i=1}^{n} p_i^{-1} = r^{-1}\); then

\[
\left\| \sum_{i=1}^{n} f_i \right\|_r \leq \prod_{i=1}^{n} \|f_i\|_{p_i}.
\]

(15.14)

**Proof.** One may prove this theorem by induction based on Hölder’s Theorem 15.18 above. Alternatively we may give a proof along the lines of the proof of Theorem 15.18 which is what we will do here.

Since Eq. (15.14) 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, \(\sum_{i=1}^{n} p_i = 1\), hence we may replace \(s_i\) by \(s_i^{p_i/r}\) and \(p_i\) by \(p_i/r\) for each \(i\) in Eq. (15.7) to find

\[
\prod_{i=1}^{n} \left( \sum_{i=1}^{n} (s_i^{p_i/r}) \right)^{p_i/r} = \prod_{i=1}^{n} \left( \sum_{i=1}^{n} s_i^{p_i} \right).
\]

Now replace \(s_i\) by \(|f_i| / \|f_i\|_{p_i}\) in the previous inequality and integrate the result to find

\[
\frac{1}{n} \left\| \sum_{i=1}^{n} f_i \right\|_r \leq \prod_{i=1}^{n} \left( \frac{1}{p_i} \|f_i\|_{p_i} \right)^{p_i/r} \int_{\Omega} |f_i|^p \, d\mu = \sum_{i=1}^{n} \frac{r}{p_i} = 1.
\]

**Definition 15.21.** 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) \(|f + g| \leq \|f\| + \|g\|\) 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 15.22 (Minkowski’s Inequality).** If \(1 \leq p \leq \infty\) and \(f, g \in L^p(\mu)\) then

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p.
\]

(15.15)

In particular, \(L^p(\mu), \|\cdot\|_p\) is a normed vector space for all \(1 \leq p \leq \infty\).
Proof. When \( p = \infty \), \(|f| \leq \|f\|_\infty \) a.e. and \(|g| \leq \|g\|_\infty \) a.e. so that \(|f + g| \leq \|f\|_\infty + \|g\|_\infty \) a.e. and therefore
\[
\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.
\]
When \( p < \infty \),
\[
|f + g|^p \leq (2 \max (|f|, |g|))^p = 2^p \max (|f|^p, |g|^p) \leq 2^p (|f|^p + |g|^p),
\]
which implies \( f + g \in L^p \) since
\[
\|f + g\|^p \leq 2^p (\|f\|^p + \|g\|^p) < \infty.
\]
Furthermore, when \( p = 1 \) we have
\[
\|f + g\|_1 = \int \Omega |f + g|d\mu \leq \int \Omega |f|d\mu + \int \Omega |g|d\mu = \|f\|_1 + \|g\|_1.
\]
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
\[
|f + g|^p = |f + g||f + g|^{p-1} \leq (|f| + |g|)|f + g|^{p-1}
\]
and then applying Hölder’s inequality with \( q = p/(p-1) \) gives
\[
\int \Omega |f + g|^pd\mu \leq \int \Omega |f| |f + g|^{p-1}d\mu + \int \Omega |g| |f + g|^{p-1}d\mu
\]
\[
\leq (\|f\|_p + \|g\|_p) \|f + g\|^{p-1}_q,
\]
where
\[
\|f + g\|^{p-1}_q = \int \Omega (|f + g|^{p-1})^q d\mu = \int \Omega |f + g|^q d\mu = \|f + g\|^q.
\]
Combining Eqs. (15.16) and (15.17) implies
\[
\|f + g\|_p \leq \|f\|_p \|f + g\|^{p/q}_q + \|g\|_p \|f + g\|^{p/q}_p
\]
Solving this inequality for \( \|f + g\|_p \) gives Eq. (15.15). \( \Box \)

15.4 Completeness of \( L^p \) – spaces

Definition 15.23 (Banach space). A normed vector space \((Z, \|\cdot\|)\) is a
Banach space if it 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,
\[
\lim_{n\to\infty} \|x_n - x_m\| = 0,
\]
then there exists an \( x \in Z \) such that \( \lim_{n\to\infty} \|x_n - x\| = 0 \).

Theorem 15.24. Let \( \|\cdot\|_\infty \) be as defined in Eq. (15.2), then
\( L^\infty (\Omega, B, \mu) \) is a Banach space. A sequence \( \{f_n\}_{n=1}^\infty \subset L^\infty \) converges to \( f \in L^\infty \) iff there exists \( E \in B \) 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 \(15.22 \), \( \|\cdot\|_\infty \) 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
\[
\mu (\|f - f_n\| > k^{-1}) = 0 \quad \text{for all } n \geq N_k.
\]
Let
\[
E = \bigcup_{k=1}^\infty \bigcup_{n \geq N_k} \{\|f - f_n\| > k^{-1}\}.
\]
Then \( \mu(E) = 0 \) and for \( x \in E^c \), \( |f(x) - f_n(x)| \leq k^{-1} \) for all \( n \geq N_k \). This shows that \( f_n \to f \) uniformly on \( E^c \). Conversely, if there exists \( E \in B \) such that \( \mu(E) = 0 \) and \( f_n \to f \) uniformly on \( E^c \), then for any \( \varepsilon > 0 \),
\[
\mu (\|f - f_n\| \geq \varepsilon) = \mu (\|f - f_n - \varepsilon\| \cap E^c) = 0
\]
for all \( n \) sufficiently large. That is to say \( \lim sup \|f - f_n\|_\infty \leq \varepsilon \) for all \( \varepsilon > 0 \).

The density of simple functions follows from the approximation Theorem \(9.39 \).

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 \( E_{m,n} = \{\|f_m - f_n\| > \varepsilon_{m,n}\} \) and \( E := \cup_{E_{m,n}} \), then \( \mu(E) = 0 \) and
\[
\sup_{x \in E^c} |f_m(x) - f_n(x)| \leq \varepsilon_{m,n} \to 0 \quad \text{as } m, n \to \infty.
\]
Therefore, \( f := \lim_{n \to \infty} f_n \) exists on \( E^c \) and the limit is uniform on \( E^c \). Letting \( f = \lim_{n \to \infty} f_n \) , then it follows that \( \lim_{n \to \infty} \|f_n - f\|_\infty = 0 \).

Theorem 15.25 (Completeness of \( L^p (\mu) \)). For \( 1 \leq p < \infty \), \( L^p (\mu) \) equipped with the \( L^p \) – norm, \( \|\cdot\|_p \) (see Eq. (15.14)), is a Banach space.
Proof. By Minkowski’s Theorem, $\|\cdot\|_p$ 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 $1 \leq p < \infty$, the case $p = \infty$ being done in Theorem 15.24.

Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev’s inequality (Lemma 15.4), $\{f_n\}$ is $L^p$-Cauchy (i.e. Cauchy in measure) and by Theorem 15.8 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou’s Lemma,

$$
\|g_j - f\|_p = \int \liminf_{k \to \infty} \|g_j - g_k\|_p d\mu \leq \liminf_{k \to \infty} \int |g_j - g_k|^p d\mu
$$

$$= \lim_{k \to \infty} \|g_j - g_k\|_p \to 0 \text{ as } j \to \infty.
$$

In particular, $\|f\|_p \leq \|g_j - f\|_p + \|g_j\|_p < \infty$ so the $f \in L^p$ and $g_j \overset{L^p}{\to} f$. The proof is finished because,

$$
\|f_n - f\|_p \leq \|f_n - g\|_p + \|g\|_p \to 0 \text{ as } j, n \to \infty.
$$

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 $\sum_{j=1}^\infty \|g_j - g_{j+1}\|_p < \infty$ and let $U := \sum_{j=0}^\infty |g_j - g_{j+1}|$ where $g_0 \equiv 0$. Then

$$
\|U\|_p \overset{\text{MCT}}{=} \lim_{N \to \infty} \left\| \sum_{j=0}^N g_{j+1} - g_j \right\|_p
$$

$$\leq \lim_{N \to \infty} \sum_{j=0}^N \|g_{j+1} - g_j\|_p = \sum_{j=0}^N \|g_{j+1} - g_j\|_p < \infty
$$

and therefore $U < \infty$ a.e. But on the set $\{U < \infty\}$ of full measure, the sum $\sum_{j=0}^\infty (g_j - g_{j+1})$ 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 17.2 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 15.26. Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{B}_0$ be a sub – $\sigma$ – algebra of $\mathcal{B}$. Then for $1 \leq p < \infty$, 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,

$$[f]_0 = \{g \in L^p(\Omega, \mathcal{B}_0, \mu) : g = f \text{ a.e.}\} \text{ and } [f] = \{g \in L^p(\Omega, \mathcal{B}, \mu) : g = f \text{ a.e.}\}.$$ 

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 $L^p(\Omega, \mathcal{B}, \mu)$. Because, $i$, is an isometry it follows that $\{(f_n)\}_{n=1}^\infty$ is a Cauchy and hence convergent sequence in $L^p(\Omega, \mathcal{B}_0, \mu)$. Letting $f \in L^p(\Omega, \mathcal{B}_0, \mu)$ such that $\|f - f_n\|_{L^p(\mu)} \to 0$, we will have, since $i$ is isometric, that $[f_n] \to [i] = \{f\}_0 \in i(L^p(\Omega, \mathcal{B}_0, \mu))$ as desired.

Exercise 15.3. 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 $1 \leq p < \infty$ that $i(L^p(\Omega, \mathcal{B}_0, \mu)) = L^p(\Omega, \mathcal{B}, \mu)$ i.e. to each $f \in L^p(\Omega, \mathcal{B}, \mu)$ there exists a $g \in L^p(\Omega, \mathcal{B}_0, \mu)$ such that $f = g$ a.e.

Hints: 1. verify the last assertion for simple functions in $L^p(\Omega, \mathcal{B}_0, \mu)$ . 2. then make use of Theorem 9.33 and Exercise 9.5.

Exercise 15.4. Suppose that $1 \leq p < \infty$, $(\Omega, \mathcal{B}, \mu)$ 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 $L^p(\Omega, \mathcal{B}, \mu)$.

15.5 Density Results

Theorem 15.27 (Density Theorem). Let $p \in [1, \infty)$, $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathcal{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that

1. $\mathcal{M} \subset L^p(\mu, \mathbb{R})$ and $\sigma(\mathcal{M}) = \mathcal{B}$.
2. There exists $\psi_k \in \mathcal{M}$ such that $\psi_k \to 1$ boundedly.

Then to every function $f \in L^p(\mu, \mathbb{R})$, there exist $\varphi_n \in \mathcal{M}$ such that $\lim_{n \to \infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. $\mathcal{M}$ 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 $\{\varphi_n\}_{n=1}^\infty \subset \mathcal{M}$ 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 $\Omega$, $1 \in \mathbb{H}$, $\mathcal{M} \subset \mathbb{H}$ 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, \( \lim_{n \to \infty} \| \psi_k (f - f_n) \|_{L^p(\mu)} = 0 \). (Take the dominating function to be \( g = [2C |f|]^p \) 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 \( \| \varphi_n - \psi_k f_n \|_{L^p(\mu)} \leq \frac{1}{n} \) then

\[
\lim_{n \to \infty} \sup_{n} \| \psi_k f - \varphi_n \|_{L^p(\mu)} \leq \lim_{n \to \infty} \sup_{n} \| \psi_k (f - f_n) \|_{L^p(\mu)} + \lim_{n \to \infty} \| \psi_k f_n - \varphi_n \|_{L^p(\mu)} = 0 \tag{15.19}
\]

which implies \( f \in \mathbb{H} \).

An application of Dynkin’s Multiplicatory System Theorem \[11.20\] now shows \( \mathbb{H} \) contains all bounded measurable functions on \( \Omega \). Let \( f \in L^p(\mu) \) be given. The dominated convergence theorem implies \( \lim_{k \to \infty} \| \psi_k 1_{[|f| \leq k]} f - f \|_{L^p(\mu)} = 0 \). (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

\[
\| \psi_k 1_{[|f| \leq k]} f - \varphi_k \|_{L^p(\mu)} \leq \frac{1}{k}.
\]

The same line of reasoning used in Eq. \[15.19\] now implies \( \lim_{k \to \infty} \| f - \varphi_k \|_{L^p(\mu)} = 0 \).

**Example 15.28.** 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 \( 1 \leq p < \infty \). To see this, apply Theorem \[15.27\] 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 15.29.** 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 \( \mathcal{S}(\mathcal{A}, \mu) \) denote the measurable simple functions, \( \varphi : \Omega \to \mathbb{R} \) such \( \{\varphi = y\} \in \mathcal{A} \) for all \( y \in \mathbb{R} \) and \( \mu(\{ \varphi \neq 0 \}) < \infty \). Then \( \mathcal{S}(\mathcal{A}, \mu) \) is dense subspace of \( L^p(\mu) \).

**Proof.** Let \( \mathbb{M} := \mathcal{S}(\mathcal{A}, \mu) \). By assumption there exists \( \Omega_k \subset \mathcal{A} \) 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 \( \sigma(\mathcal{M}) \) – measurable for every \( A \in \mathcal{A} \). So we have shown that \( \mathcal{A} \subset \sigma(\mathcal{M}) \subset \mathcal{B} \) and therefore \( \mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathcal{M}) \subset \mathcal{B} \), i.e. \( \sigma(\mathcal{M}) = \mathcal{B} \). The theorem now follows from Theorem \[15.27\] after observing \( \psi_k := 1_{\Omega_k} \in \mathbb{M} \) and \( \psi_k \to \psi \) boundedly.\footnote{It is at this point that the proof would break down if \( p = \infty \).}

**Theorem 15.30 (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

\[
\mathcal{D} = \{ \sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q}, A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty \}
\]

is a countable dense subset.

**Proof.** It is left to reader to check \( \mathcal{D} \) is dense in \( \mathcal{S}(\mathcal{A}, \mu) \) relative to the \( L^p(\mu) \) – norm. Once this is done, the proof is then complete since \( \mathcal{S}(\mathcal{A}, \mu) \) is a dense subspace of \( L^p(\mu) \) by Theorem \[15.29\].

### 15.6 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 \( q \geq p \) as is shown in Corollary \[15.31\] 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 \( q \geq p \), see Corollary \[15.36\] below.

**Corollary 15.31 (Example \[15.17\] revisited).** If \( \mu(\Omega) < \infty \) and \( 0 < p \leq q \leq \infty \), then \( L^q(\mu) \subset L^p(\mu) \), the inclusion map is bounded and in fact

\[
\| f \|_p \leq [\mu(\Omega)]^{(\frac{1}{q} - \frac{1}{p})} \| f \|_q .
\]

**Proof.** Take \( a \in [1, \infty) \) such that \( \frac{1}{p} = \frac{1}{a} + \frac{1}{q} \), i.e. \( a = \frac{pq}{q - p} \). Then by Theorem \[15.20\]

\[
\| f \|_p = \| f \cdot 1 \|_p \leq \| f \|_q \cdot \| 1 \|_a = \mu(\Omega)^{1/a} \| f \|_q = \mu(\Omega)^{\left(\frac{1}{q} - \frac{1}{p}\right)} \| f \|_q.
\]

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 15.32 (Power Inequalities). Let $a := (a_1, \ldots, a_n)$ with $a_i > 0$ for $i = 1, 2, \ldots, n$ 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 15.31 $p \to \|a\|_p$ is increasing in $p$ for $p > 0$. For $p = -q < 0$, we have

$$\|a\|_p := \left(\frac{1}{n} \sum_{i=1}^{n} a_i^{-q} \right)^{-1/q} = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{a_i}\right)^q \right)^{1/q} = \frac{1}{\|a\|_q}^{-1},$$

where $\frac{1}{q} := (1/a_1, \ldots, 1/a_n)$. So for $p < 0$, as $p$ increases, $q = -p$ decreases, so that $\|a\|_q$ is decreasing and hence $\|a\|_q^{-1}$ is increasing. Hence we have shown that $p \to \|a\|_p$ is increasing for $p \in \mathbb{R} \setminus \{0\}$.

We now claim that $\lim_{p \to 0} \|a\|_p = \sqrt[1/q]{a_1 \ldots a_n}$. 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,

$$\frac{1}{n} \sum_{i=1}^{n} a_i^p = 1 + \frac{1}{n} \sum_{i=1}^{n} p \ln a_i + O(p^2).$$

Hence it follows that

$$\lim_{p \to 0} \|a\|_p = \lim_{p \to 0} \left(\frac{1}{n} \sum_{i=1}^{n} a_i^p \right)^{1/p} = \lim_{p \to 0} \left(1 + \frac{1}{n} \sum_{i=1}^{n} p \ln a_i + O(p^2) \right)^{1/p} = e^{\frac{1}{q} \sum_{i=1}^{n} \ln a_i} = \sqrt[1/q]{a_1 \ldots a_n}.$$

So if we now define $\|a\|_0 := \sqrt[1/q]{a_1 \ldots a_n}$, 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 $p > 0$,

$$\frac{1}{n} M^p \leq \frac{1}{n} \sum_{i=1}^{n} a_i^p \leq M^p$$

and therefore,

$$\left(\frac{1}{n}\right)^{1/p} M \leq \|a\|_p \leq M.$$

Since $(\frac{1}{n})^{1/p} \to 1$ as $p \to \infty$, it follows that $\lim_{p \to \infty} \|a\|_p = M$. For $p = -q < 0$, we have

$$\lim_{p \to -\infty} \|a\|_p = \lim_{q \to -\infty} \left(\frac{1}{\|a\|_q} \right)^{1/q} = \frac{1}{\max_i (1/a_i)} = \frac{1}{m} = m = \min_i a_i.$$

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 $\mathbb{R} \ni p \to \|a\|_p \in (0, \infty)$ is a continuous non-decreasing function of $p$.

Proposition 15.33. Suppose that $0 < p_0 < p_1 \leq \infty$, $\lambda \in (0, 1)$ and $p_\lambda \in (p_0, p_1)$ be defined by

$$\frac{1}{p_\lambda} = \frac{1 - \lambda}{p_0} + \frac{\lambda}{p_1},$$

with the interpretation that $\lambda/p_1 = 0$ if $p_1 = \infty$. 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 $1 \leq p_0 < p_1 \leq \infty$ and $f \in L^{p_0} + L^{p_1}$ let

$$\|f\| := \inf \left\{ \|g\|_{p_0} + \|h\|_{p_1} : f = g + h \right\}.$$

Then $(L^{p_0} + L^{p_1}, \|\|)$ is a Banach space and the inclusion map from $L^{p_\lambda}$ to $L^{p_0} + L^{p_1}$ is bounded; in fact $\|f\| \leq 2 \|f\|_{p_\lambda}$ 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 $g = f|_E$ and $h = f|_{E^c}$ 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

$$\|g\|_{p_0} = \int_E |f|^{p_0} 1_{|f| > M} = M^{p_0} \int_E \frac{|f|^{p_0}}{M} 1_{|f| > M} \leq M^{p_0} \int_E \frac{|f|^{p_0}}{M} 1_{|f| > M} \leq M^{p_0} \|f\|_{p_0}^{p_0} < \infty$$

and

$$\|h\|_{p_1} = \int_E |f|^{p_1} 1_{|f| \leq M} = \int_E \frac{|f|^{p_1}}{M} 1_{|f| \leq M} \leq M^{p_1} \int_E \frac{|f|^{p_1}}{M} 1_{|f| \leq M} \leq M^{p_1} \|f\|_{p_1}^{p_1} < \infty.$$

A little algebra shows that $\lambda$ may be computed in terms of $p_0, p_\lambda$ and $p_1$ by

$$\lambda = \frac{p_0}{p_\lambda}, \frac{p_1 - p_\lambda}{p_1 - p_0}.$$
Moreover this shows
\[ \|f\| \leq M^{1-p_1/p_0} \|f\|_{p_0}^{p_1/p_0} + M^{1-p_1/p_2} \|f\|_{p_2}^{p_1/p_2}. \]
Taking \( M = \lambda \|f\|_{p_0} \) then gives
\[ \|f\| \leq \left( \lambda^{1-p_1/p_0} + \lambda^{1-p_1/p_2} \right) \|f\|_{p_0}, \]
and then taking \( \lambda = 1 \) shows \( \|f\| \leq 2 \|f\|_{p_0} \). The proof that \((L^{p_0} + L^{p_1}, \|\cdot\|)\) is a Banach space is left as Exercise 15.11 to the reader.

**Corollary 15.34 (Interpolation of \( L^p \)-norms).** Suppose that \( 0 < p_0 < p_1 \leq \infty \), \( \lambda \in (0, 1) \) and \( p\lambda \in (p_0, p_1) \) be defined as in Eq. (15.20), then \( L^{p_0} \cap L^{p_1} \subset L^{p\lambda} \) and
\[ \|f\|_{p\lambda} \leq \left( \lambda^{1-p_1/p_0} + \lambda^{1-p_1/p_2} \right) \|f\|_{p_0}. \tag{15.21} \]
Further assume \( 1 \leq p_0 < p\lambda < p_1 \leq \infty \), 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
\[ \|f\|_{p\lambda} \leq \max \left( \lambda^{-1}, (1-\lambda)^{-1} \right) \left( \|f\|_{p_0} + \|f\|_{p_1} \right). \tag{15.22} \]

The heuristic expression 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 15.20
\[ \|f\|_{p\lambda} = \|f\|_{p_1} \|f\|_{p_0} \leq \|f\|_{p_0} \|f\|_{p_1} \leq \|f\|_{p_0} \|f\|_{p_1} \leq \|f\|_{p_0} \|f\|_{p_1}. \]

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 \( \lim_{n \to \infty} \|g - f_n\|_{p_1} = 0 \). By Chebyshev's inequality (Lemma 15.4) \( f_n \to f \) and \( f_n \to g \) in measure and therefore by Theorem 15.8 \( f = g \) a.e. It now is clear that \( \lim_{n \to \infty} \|f - f_n\| = 0 \). The estimate in Eq. (15.22) is left as Exercise 15.10 to the reader.

**Remark 15.35.** Combining Proposition 15.33 and Corollary 15.34 gives
\[ L^{p_0} \cap L^{p_1} \subset L^{p\lambda} \subset L^{p_0} + L^{p_1} \]
for \( 0 < p_0 < p_1 \leq \infty \), \( \lambda \in (0, 1) \) and \( p\lambda \in (p_0, p_1) \) as in Eq. (15.20).

**Corollary 15.36.** Suppose now that \( \mu \) is counting measure on \( \Omega \). Then \( L^p(\mu) \subset L^q(\mu) \) for all \( 0 < p < q \leq \infty \) and \( \|f\|_q \leq \|f\|_p \).

**Proof.** Suppose that \( 0 < p < q < \infty \), then
\[ \|f\|_q = \sup \{|f(x)|^q : x \in \Omega\} \leq \sum_{x \in \Omega} |f(x)|^p = \|f\|_p, \]
i.e. \( \|f\|_q \leq \|f\|_p \) for all \( 0 < p < \infty \). For \( 0 < p \leq q \leq \infty \), apply Corollary 15.34 with \( p_0 = p \) and \( p_1 = \infty \) to find
\[ \|f\|_q \leq \|f\|_p^{p,q} \|f\|_\infty^{1-p,q} \leq \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} = \|f\|_p. \]

**15.6.1 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 \( p \leq q \), then \( \ell^p \subset \ell^q \) and \( \|f\|_q \leq \|f\|_p \).
3. Since \( \mu(|f| > \varepsilon) \leq \varepsilon^{-p} \|f\|_p^p \), \( L^p \)-convergence implies \( L^q \) convergence.
4. \( L^q \) - convergence implies almost everywhere convergence for some subsequence.
5. If \( \mu(\Omega) < \infty \) then almost everywhere convergence implies uniform convergence on certain sets of small measure and in particular we have \( L^0 \) - convergence.
6. If \( \mu(\Omega) < \infty \), then \( L^0 \subset L^p \) for all \( p \leq q \) and \( L^q \) - convergence implies \( L^p \) - convergence.

**15.7 Uniform Integrability**

Let \((\Omega, B, \mu)\) be a probability space, \( 1 \leq p < \infty \), 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 15.37** For \( f \in L^1(\mu) \) and \( E \in B \), let
\[ \mu(f : E) := \int_E f d\mu. \]
and more generally if \( A, B \in B \) 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 15.38. A collection of functions, \( A \subset L^1(\mu) \) is said to be uniformly integrable if,
\[
\lim_{a \to \infty} \sup_{f \in A} \mu(|f| : |f| \geq a) = 0. \tag{15.23}
\]

In words, \( A \subset L^1(\mu) \) is uniformly integrable if “tail expectations” can be made uniformly small.

Example 15.39. If \( A \subset L^1(\mu) \) and there exists a dominating function, \( g \in L^1(\mu) \), such that \( |f| \leq g \) for all \( f \in A \), then \( A \) is uniformly integrable. Indeed,
\[
\sup_{f \in A} \mu(|f| : |f| \geq a) \leq \mu(g : g \geq a) \to 0 \text{ as } a \to \infty
\]
by the dominated convergence theorem.

Example 15.40. Suppose that \( A := \{X_n\}_{n=1}^{\infty} \subset L^1(\mu) \) is a sequence of random variables which all have the same law, then \( A \) is uniformly integrable. This is because, \( \mu(|X_n| : |X_n| \geq a) = \mu(|X_1| : |X_1| \geq a) \), and so
\[
\sup_n \mu(|X_n| : |X_n| \geq a) = \mu(|X_1| : |X_1| \geq a) \to 0 \text{ as } a \to \infty.
\]
This example illustrates the fact that uniform integrability is really a condition on the collection of measures, \( \{\text{Law}_\mu(X) : X \in A\} \), on \( (\mathcal{C}, \mathcal{B}_\mathcal{C}) \).

Example 15.41. Suppose that \( A := \{X_n\}_{n=1}^{\infty} \subset L^1(\mu) \) 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 \( \mu\left(\{|X_n| \geq a \text{ i.o.}\}\right) = 1 \) 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 15.39 to show \( A \) is uniformly integrable. Of course we do know by Example 15.40 that \( A \) is uniformly integrable.

Exercise 15.5. Suppose \( A \) is an index set, \( \{f_\alpha\}_{\alpha \in A} \) and \( \{g_\alpha\}_{\alpha \in A} \) are two collections of random variables and \( C \subset (0, \infty) \). 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 15.42. If \( A \subset L^1(\mu) \) is uniformly integrable, then \( \sup_{f \in A} \|f\|_1 < \infty \).

Proof. Choose \( a \in (0, \infty) \) sufficiently large so that \( \sup_{f \in A} \mu(|f| : |f| \geq a) \leq 1 \). Then for \( f \in A \),
\[
\|f\|_1 = \mu(|f| : |f| \geq a) + \mu(|f| : |f| < a) \leq 1 + a \mu(\Omega).
\]

Lemma 15.43. If \( \{f_n\}_{n=1}^{\infty} \) is a collection of random variables such that \( f_n \overset{\mu}{\to} 0 \). Then for every \( a \in (0, \infty) \), \( g_n := f_n1_{|f_n| \leq a} \to 0 \) in \( L^p(\mu) \) for all \( 1 \leq p < \infty \).

Proof. As \( |g_n| \leq |f_n| \) it follows that \( g_n \overset{\mu}{\to} 0 \) as \( n \to \infty \). Since \( |g_n| \leq a \), we now apply the dominated convergence theorem (Corollary 15.9) to conclude \( \lim_{n \to \infty} \|g_n\|_p = 0 \).

Proposition 15.44. If \( \{f_n\}_{n=1}^{\infty} \) is a sequence of random variables, then \( \lim_{n \to \infty} \|f_n\|_1 = 0 \) iff \( f_n \overset{\mu}{\to} 0 \) and \( \{f_n\}_{n=1}^{\infty} \) is uniformly integrable.

Proof. (\( \Rightarrow \)) Suppose that \( \lim_{n \to \infty} \|f_n\|_1 = 0 \). By Chebyshev’s inequality,
\[
\mu(|f_n| \geq \varepsilon) \leq \frac{1}{\varepsilon} \|f_n\|_1 \to 0 \text{ as } n \to \infty,
\]
i.e. \( f_n \overset{\mu}{\to} 0 \) as \( n \to \infty \). If \( N \in \mathbb{N} \), then using Example 15.39 it follows that
\[
\sup_n \mu(|f_n| : |f_n| \geq a) \leq \sup_{n<N} \mu(|f_n| : |f_n| \geq a) \vee \sup_{n \geq N} \|f_n\|_1 \to \sup_{n \geq N} \|f_n\|_1 \text{ as } a \to \infty
\]
and so
\[
\lim_{a \to \infty} \left[ \sup_n \mu(|f_n| : |f_n| \geq a) \right] \leq \sup_{n \geq N} \|f_n\|_1 \to 0 \text{ as } N \to \infty.
\]

This shows \( \{f_n\}_{n=1}^{\infty} \) is uniformly integrable.

(\( \Leftarrow \)) Now suppose that \( f_n \overset{\mu}{\to} 0 \) and \( \{f_n\}_{n=1}^{\infty} \) is uniformly integrable. Then given \( a \in (0, \infty) \),
\[
\|f_n\|_1 = \mu(|f_n|) = \mu(|f_n|1_{|f_n| \leq a}) + \mu(|f_n| : |f_n| \geq a) \leq \mu(|f_n|1_{|f_n| \leq a}) + \sup_k \mu(|f_k| : |f_k| \geq a).
\]

\footnote{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, \( \{f_n := 1_{[-n,n]}\}_{n=1}^{\infty} \) are uniformly integrable but not bounded in \( L^1(\mu) \).}
Remark 15.47 (Optional). $\epsilon > 0$ are bounded in $A$ and $\mu(E) < \delta$ is uniformly absolutely continuous.

Definition 15.45. A collection of functions, $A \subset L^1(\mu)$, is said to be uniformly absolutely continuous if for all $\epsilon > 0$ there exists $\delta > 0$ such that
\[
\sup_{f \in A} \mu(\{ |f| : E \}) < \epsilon \quad \text{whenever} \quad \mu(E) < \delta.
\] (15.24)

Equivalently put,
\[
\limsup_{\delta \downarrow 0} \sup_{f \in A} \{ |f| : E \} = f \in A \quad \text{and} \quad \mu(E) < \delta = 0.
\] (15.25)

Example 15.46 (Optional). I claim that $E = \{ f \}$ with $f \in L^1(\mu)$ is uniformly absolutely continuous. If not there would exist and $\epsilon > 0$ and $E_n \in B$ such that $\mu(\{ |f| : E \}) \geq \epsilon$ while $\lim_{n \to \infty} \mu(E_n) = 0$. But this is not possible since $1_{E_n} f \not \to 0$ as $n \to \infty$ and $|1_{E_n} f| \leq |f| \in L^1(\mu)$ and so by dominated convergence theorem (Corollary 15.19),
\[
\epsilon \leq \lim_{n \to \infty} \mu(\{ |f| : E_n \}) = \lim_{n \to \infty} \mu(\{ E_n f \}) = 0.
\]

Remark 15.47 (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 $\Omega = \{ * \} \cup \{ * \} = 1$. 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 $\{f_n\}$ 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 $\{E_i\}_{i=1}^k$ with $E_i \in A$ and $\mu(E_i) < \delta$, see Lemma 15.55 below. If Eq. (15.24) holds with $\epsilon = 1$, then
\[
\mu(\{ |f_n| : E \}) \leq k
\]
showing that $\mu(\{ |f_n| \}) \leq k$ for all $n$.

Proposition 15.48. A subset $A \subset L^1(\mu)$ is uniformly integrable iff $A \subset L^1(\mu)$ is bounded and uniformly absolutely continuous.

Proof. ($\implies$) We have already seen that uniformly integrable subsets, $A$, are bounded in $L^1(\mu)$. Moreover, for $f \in A$, and $E \in B$,
\[
\mu(\{ f : E \}) = \mu(\{ |f| \} = M, E) + \mu(\{ |f| : |f| < M, E \}) \leq \mu(\{ |f| \} = M) + M \mu(E).
\]

Therefore,
\[
\limsup_{\delta \downarrow 0} \sup_{f \in A} \{ |f| : E \} = f \in A \quad \text{and} \quad \mu(E) \leq \mu(\{ |f| \} = M) \to 0 \text{ as } M \to \infty
\]
which verifies that $A$ is uniformly absolutely continuous.

($\impliedby$) Let $K := \sup_{f \in A} \|f\|_1 < \infty$. Then for $f \in A$, we have
\[
\mu(\{ |f| \} = M) \leq \|f\|_1 / a \leq K/a \text{ for all } a > 0.
\]
Hence given $\epsilon > 0$ and $\delta > 0$ as in the definition of uniform absolute continuity, we may choose $a = K/\delta$ in which case
\[
\sup_{f \in A} \mu(\{ |f| : |f| \geq M \}) < \epsilon.
\]
Since $\epsilon > 0$ was arbitrary, it follows that $\lim_{n \to \infty} \sup_{f \in A} \mu(\{ |f| : |f| \geq a \}) = 0$ as desired.

Corollary 15.49. Suppose $\{f_n\}_{\alpha \in A}$ and $\{g_\alpha\}_{\alpha \in A}$ are two uniformly integrable collections of functions, then $\{f_n + g_\alpha\}_{\alpha \in A}$ is also uniformly integrable.

Proof. By Proposition 15.48 $\{f_n\}_{\alpha \in A}$ and $\{g_\alpha\}_{\alpha \in A}$ are both bounded in $L^1(\mu)$ and are both uniformly absolutely continuous. Since $\|f_n + g_\alpha\|_1 \leq \|f_n\|_1 + \|g_\alpha\|_1$ it follows that $\{f_n + g_\alpha\}_{\alpha \in A}$ is bounded in $L^1(\mu)$ as well. Moreover, for $\epsilon > 0$ we may choose $\delta > 0$ such that $\mu(\{ |f_n| : |f_n| \leq \epsilon \}$ and $\mu(\{ |g_\alpha| : |g_\alpha| < \epsilon \}$ whenever $\mu(E) < \delta$. For this choice of $\epsilon$ and $\delta$, we then have
\[
\mu(\{ f_n + g_\alpha : \} \leq \mu(\{ |f_n| + |g_\alpha| : E \}) < 2\epsilon \text{ whenever } \mu(E) < \delta,
\]
showing $\{f_n + g_\alpha\}_{\alpha \in A}$ uniformly absolutely continuous. Another application of Proposition 15.48 completes the proof.

Corollary 15.50. If $1 \leq p < \infty$ and $\{f_n\}_{n=1}^\infty \cup \{0\} \subset L^p(\mu)$, then $\{f_n^p\}_{n=1}^\infty$ is uniformly integrable iff $\{f_n^p\}_{n=1}^\infty$ is uniformly integrable.

Proof. ($\implies$) By Hölder’s inequality,
\[
|f_n - f|^p \leq 2^\frac{p}{p} |f_n|^p + |f|^p
\]
and so by Corollary 15.49 and Exercise 15.5 it follows that $\{f_n^p\}_{n=1}^\infty$ is uniformly integrable.

($\impliedby$) The proof here is similar but now based on
\[
|f_n|^p \leq |f_n - f| + |f|^p \leq 2^\frac{p}{p} |f_n - f|^p + |f|^p.
\]
Exercise 15.6 (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 \( \{ \frac{X_n}{n} \}_{n=1}^{\infty} \) is uniformly integrable. [Actually there is no reason to assume independence here – identically distributed is enough.]

Theorem 15.51 (Vitali Convergence Theorem). Let \( 1 \leq p < \infty \), \((\Omega, \mathcal{B}, \mu)\) be a finite measure space, \( A := \{f_n\}_{n=1}^{\infty} \) be a sequence of functions in \( L^p(\mu) \), and \( f : \Omega \rightarrow \mathbb{C} \) be a measurable function. Then \( f \in L^p(\mu) \) and \( \|f - f_n\|_p \rightarrow 0 \) as \( n \rightarrow \infty \) iff \( f_n \xrightarrow{\mu} f \) and \( \{\|f_n\|_p\}_{n=1}^{\infty} \) is uniformly integrable.

Proof. (\( \implies \)) If \( f_n \xrightarrow{\mu} f \) then by Chebyshev’s inequality \( f_n \xrightarrow{\mu} f \). By assumption \( |f_n|_p \xrightarrow{\mu} 0 \) in \( L^1(\mu) \) and so, by Proposition 15.48 \( \{\|f_n\|_p\}_{n=1}^{\infty} \) is uniformly integrable.

Therefore, \( f \) is uniformly integrable. \( \Box \)

\( \{\|f_n\|_p\}_{n=1}^{\infty} \) is uniformly integrable.

\( \iff \) Assume \( f_n \xrightarrow{\mu} f \) as \( n \rightarrow \infty \) and \( \{\|f_n\|_p\}_{n=1}^{\infty} \) is uniformly integrable.

Then by Fatou’s Lemma (Exercise 15.1) and Lemma 15.42

\[
\int_{\Omega} |f|^p \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} |f_n|^p \, d\mu \leq \sup_n \int_{\Omega} |f_n|^p \, d\mu < \infty,
\]

i.e. \( f \in L^p(\mu) \). From Corollary 15.50 we now know that \( \{\|f_n\|_p\}_{n=1}^{\infty} \) is uniformly integrable. As \( f_n \xrightarrow{\mu} f \) is equivalent to \( |f_n - f|_p \xrightarrow{\mu} 0 \) we may apply Proposition 15.48 to show \( |f - f_n|_p \rightarrow 0 \) in \( L^1(\mu) \), i.e. \( f_n \xrightarrow{\mu} f \) in \( L^p(\mu) \).

Example 15.52. Let \( \Omega = [0,1] \), \( \mathcal{B} = \mathcal{B}_{[0,1]} \) and \( P = m \) be Lebesgue measure on \( \mathcal{B} \). Then the collection of functions, \( f_\varepsilon := \mathbb{1}_{[\varepsilon,1]} \) is bounded in \( L^1(P) \), \( f_\varepsilon \rightarrow 0 \) a.e. as \( \varepsilon \downarrow 0 \) but

\[
0 = \int_{\Omega} \lim_{\varepsilon \downarrow 0} f_\varepsilon \, dP \neq \lim_{\varepsilon \downarrow 0} \int_{\Omega} f_\varepsilon \, dP = 1.
\]

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,

\[
\sup_{\varepsilon \in (0,1)} m(\{f_\varepsilon : |f_\varepsilon| \geq a\}) = 1 \quad \text{for all } a > 0.
\]

Example 15.53. 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 \( f_\varepsilon := a_\varepsilon \mathbb{1}_{[\varepsilon,1]} \). Then \( \mathbb{E}[f_\varepsilon] = \varepsilon a_\varepsilon \) and so \( \sup_{\varepsilon \in (0,1)} \|f_\varepsilon\|_1 =: K < \infty \) iff \( \varepsilon a_\varepsilon \leq K \) for all \( \varepsilon \). Since

\[
\sup_{\varepsilon} \mathbb{E}[f_\varepsilon \geq M] = \sup_{\varepsilon} [\varepsilon a_\varepsilon \cdot 1_{\varepsilon a_\varepsilon \geq M}],
\]

we are actually reproving Lemma 15.10 here.

\[
\lim_{\varepsilon \downarrow 0} \sup_{f \in A} \mu(\{f \geq a\}) \leq \mu(\{f \geq a\}) \leq K \varepsilon a,
\]

and hence

\[
\lim_{a \rightarrow \infty} \sup_{f \in A} \mu(\{f \geq a\}) \leq \lim_{a \rightarrow \infty} K \varepsilon a = 0.
\]

2. *(Skip this if you like.) By assumption, \( \varepsilon a := \sup_{f \in A} \mu(\{f \geq a\}) \rightarrow 0 \) as \( a \rightarrow \infty \). 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
Exercise 15.7. Show directly that if \( \mu(\Omega) < \infty \), \( \varphi \) is as in Lemma \ref{lem:5.54} and \( \{f_n\} \subset L^1(\Omega) \) such that \( f_n \overset{L^1}{\to} f \) and \( K := \sup_n E[\varphi(|f_n|)] < \infty \), then \( ||f - f_n||_1 \to 0 \) as \( n \to \infty \).

15.7.1 Atoms

Lemma 15.55 (Saks’ Lemma \[10\], Lemma 7 on p. 308). Suppose that \((\Omega, B, P)\) is a probability space such that \( P \) has no atoms. (Recall that \( A \in B \) is an atom if \( P(A) > 0 \) and for any \( B \subset A \) with \( B \in B \) we have either \( P(B) = 0 \) or \( P(B) = P(A) \).) Then for every \( \delta > 0 \) there exists a partition \( \{E_k\}_{k=1}^n \) of \( \Omega \) with \( \mu(E_k) < \delta \). (For related results along this line also see \[9, 13, 20, 24\] to name a few.)

Proof. For any \( A \in B \) let

\[
\beta(A) := \sup \{P(B) : B \subset A \text{ and } P(B) \leq \delta \}.
\]

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 \( A_n \subset A \setminus \bigcup_{k=1}^{n-1} A_k \) and

\[
0 < P(A_n) < P(A \setminus \bigcup_{k=1}^{n-1} A_k).
\]

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 \( \Omega \)

\[
\delta \geq P(A_k) \geq \frac{1}{2} \beta(\Omega \setminus [A_1 \cup \cdots \cup A_{k-1}] \\
\text{with either } P(\Omega \setminus [A_1 \cup \cdots \cup A_{n-1}]) > 0 \text{ if it happens that } P(\Omega \setminus [A_1 \cup \cdots \cup A_n]) = 0 \text{ it is easy to see we are done. So we may assume that process can be carried on indefinitely. We then let } F := \Omega \setminus \bigcup_{k=1}^\infty A_k \text{ and observe that}
\]

\[
\beta(F) \leq \beta(\Omega \setminus [A_1 \cup \cdots \cup A_{n-1}]) \leq 2P(A_n) \to 0 \text{ as } n \to \infty
\]

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 \( P(\Omega \setminus \bigcup_{k=1}^n A_k) \leq \delta \). We may then define \( E_k = A_k \) for \( 1 \leq k \leq n \) and \( E_n = \Omega \setminus \bigcup_{k=1}^n A_k \) in order to construct the desired partition.

Corollary 15.56. Suppose that \((\Omega, B, P)\) is a probability space such that \( P \) has no atoms. Then for any \( \alpha \in (0, 1) \) there exists \( A \in B \) with \( P(A) = \alpha \).

Proof. We may assume the \( \alpha \in (0, 1/2) \). By dividing \( \Omega \) into a partition \( \{E_k\}_{k=1}^\infty \) with \( P(E_k) \leq \alpha/2 \) we may let \( A_1 := \bigcup_{k=1}^n E_i \) 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.)
15.8 Exercises

Exercise 15.8. 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 15.34 to show \( \limsup_q \|f\|_q \leq \|f\|_\infty \) and to show \( \liminf_q \|f\|_q \geq \|f\|_\infty \), let \( M < \|f\|_\infty \) and make use of Chebyshev’s inequality.

Exercise 15.9. Let \( \infty > a, b > 1 \) with \( a^{-1} + b^{-1} = 1 \). Give a calculus proof of the inequality
\[
\frac{\text{st}}{a} + \frac{\text{tb}}{b} \geq \frac{x^a}{a} + \frac{x^b}{b} \quad \text{for all } x \geq 0.
\]

Hint: by taking \( s = x^b/a \), show that it suffices to prove
\[
\frac{x^a}{a} + \frac{1}{b} \quad \text{for all } x \geq 0.
\]
and then maximize the function \( f(x) = x - x^a/a \) for \( x \in \mathbb{R} \).

Exercise 15.10. Prove Eq. (15.22) in Corollary 15.34 (Part of Folland 6.3 on p. 186.) Hint: Use the inequality, with \( a, b \geq 1 \) with \( a^{-1} + b^{-1} = 1 \) chosen appropriately,
\[
\text{st} \leq \frac{s^a}{a} + \frac{t^b}{b}
\]

applied to the right side of Eq. (15.21).

Exercise 15.11. Complete the proof of Proposition 15.33 by showing \( (L^p + L^\infty, \|\cdot\|) \) is a Banach space.

Exercise 15.12. Let \((\Omega, \mathcal{B}, \mu)\) be a probability space. Show directly that for any \( g \in L^1(\mu), \mathcal{A} = \{g\} \) is uniformly absolutely continuous. (We already know this is true by combining Example 15.39 with Proposition 15.48.

Exercise 15.13. Suppose that \((\Omega, \mathcal{B}, P)\) is a probability space and \( \{X_n\}_{n=1}^\infty \) is a sequence of uncorrelated (i.e. \( \text{Cov}(X_n, X_m) = 0 \) if \( m \neq n \)) square integrable random variables such that \( \mu = \mathbb{E}X_n \) and \( \sigma^2 = \text{Var}(X_n) \) for all \( n \). Let \( S_n := X_1 + \cdots + X_n \). Show \( \frac{\mathbb{E}(S_n)}{\sqrt{n}} - \mu = \frac{\text{Var}(S_n)}{\sqrt{n}} \to 0 \) as \( n \to \infty \).

Exercise 15.14. 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, \( \frac{\mathbb{E}(S_n)}{n} \to \mu \) in \( L^1(\mu) \) as \( n \to \infty \). (Incidentally, this shows that \( \{S_n/n\}_{n=1}^\infty \) is U.I.

Hint: for \( M \in (0, \infty) \), let \( X^M \) \( X^M := X^M \cdot 1_{|X| \leq M} \) and \( S^M_n := X^M_1 + \cdots + X^M_n \) and use Exercise 15.13 to see that
\[
\frac{S_n^M}{n} \to \mathbb{E}X^M_1 \text{ in } L^2(\mu) \subset L^1(\mu)
\]

for all \( M \).

Using this to show \( \lim_{n \to \infty} \|S_n - \mathbb{E}X^M_1\|_1 = 0 \) by taking good control on \( \|S_n - S^M_n\|_1 \) and \( |\mathbb{E}X_n - \mathbb{E}X^M_n| \).

Exercise 15.15. Suppose \( 1 \leq p < \infty \), \( \{X_n\}_{n=1}^\infty \) 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, \( \frac{\mathbb{E}n}{n} \to \mu \) in \( L^p(\mu) \) as \( n \to \infty \). Hint: show \( \{\frac{S_n}{n}\}_{n=1}^\infty \) is U.I. – this is not meant to be hard!

15.9 Appendix: Convex Functions

Reference; see the appendix (page 500) of Revuz and Yor.

Definition 15.57. Given any function, \( \varphi : (a, b) \to \mathbb{R} \), we say that \( \varphi \) is convex if for all \( a < x_0 \leq x_1 < b \) and \( t \in [0, 1] \),
\[
\varphi(x_t) \leq b_t := (1 - t) \varphi(x_0) + t \varphi(x_1) \quad \text{for all } t \in [0, 1],
\]
where
\[
x_t := x_0 + t(x_1 - x_0) = (1 - t)x_0 + tx_1,
\]
see Figure 15.3.

Fig. 15.3. 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 \( m_1 \leq m_3 \leq m_2 \).
Lemma 15.58. Let \( \varphi : (a, b) \to \mathbb{R} \) be a function and
\[
F(x_0, x_1) := \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0}
\]
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. (15.27), 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,
\[
F(x_0, x_t) = \frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} = \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = F(x_0, x_1)
\]
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 15.59 (A generalized FTC). If \( \varphi \in PC^1((a, b) \to \mathbb{R}) \) then for all \( a < x < y < b \),
\[
\varphi(y) - \varphi(x) = \int_x^y \varphi'(t) \, dt.
\]

Proof. Let \( b_1, \ldots, b_{t-1} \) be the points of non-differentiability of \( \varphi \) in \((x, y)\)
and set \( b_0 = x \) and \( b_t = y \). Then
\[
\varphi(y) - \varphi(x) = \sum_{k=1}^t \left[ \varphi(b_k) - \varphi(b_{k-1}) \right] = \sum_{k=1}^t \int_{b_{k-1}}^{b_k} \varphi'(t) \, dt = \int_x^y \varphi'(t) \, dt.
\]

Figure 15.5 below serves as motivation for the following elementary lemma on convex functions.

Lemma 15.60 (Convex Functions). Let \( \varphi \in PC^1((a, b) \to \mathbb{R}) \) and for \( x \in (a, b) \), let
\[
\varphi'(x+) := \lim_{h \to 0^+} \frac{\varphi(x + h) - \varphi(x)}{h} \quad \text{and} \quad \varphi'(x-) := \lim_{h \to 0^-} \frac{\varphi(x + h) - \varphi(x)}{h}.
\]

(Of course, \( \varphi'(x) = \varphi'(x) \) 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
\[
\varphi'(x_0-0) \leq \varphi'(x_00) \quad \text{and} \quad \varphi'(x_0+0) \quad \text{we have,}
\]
\[
\varphi(x_0) + m \leq \varphi(x) \quad \forall \ x_0, x \in (a, b).
\]

PC^3 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,
\[
\{a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b\},
\]
such that \( \varphi|_{[a_{j-1}, a_j]} \in C^1 \) for all \( j = 1, 2, \ldots, n. \)
2. If \( \varphi \in PC^2((a,b) \rightarrow \mathbb{R}) \) with \( \varphi''(x) \geq 0 \) for almost all \( x \in (a,b) \), then Eq. \( (15.29) \) 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 \( 15.9 \) 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

\[
\frac{\varphi(x_0 - h) - \varphi(x_0)}{-h} = \varphi'(c_-(h)) \leq \frac{\varphi(x_0 + h) - \varphi(x_0)}{h} = \varphi'(c_+(h)).
\]

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,

\[
\varphi'(x) \leq \varphi'(x_0-) \leq \varphi'(x_0+) \leq \varphi'(y).
\]

Now let \( m \in (\varphi'(x_0-),\varphi'(x_0+)) \). By the fundamental theorem of calculus in Lemma \( 15.59 \) and making use of Eq. \( (15.30) \), 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

\[
\varphi(x) - \varphi(x_0) = \int_{x}^{x_0} \varphi'(t) \, dt \leq \int_{x}^{x_0} m \, dt = m(x_0 - x).
\]

These two equations implies Eq. \( (15.29) \) 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. \( 10.63 \) with \( F = \varphi \), \( a = x_0 \), and \( b = x \)) implies

\[
\varphi(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0) + (x - x_0)^2 \int_{0}^{1} \varphi''(x_0 + \tau(x - x_0))(1 - \tau) \, d\tau \geq \varphi(x_0) + \varphi'(x_0)(x - x_0).
\]

3. For any \( \xi \in (a,b) \), let \( h_\xi(x) := \varphi(x_0) + \varphi'(x_0)(x - x_0) \). By Eq. \( (15.29) \) 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. \( (15.27) \) under the hypothesis of item 2. Clearly these proofs may be omitted.

3a. By Lemma \( 15.58 \) below it suffices to show either
\[
\frac{d}{dx} \frac{\varphi(y) - \varphi(x)}{y - x} \geq 0 \quad \text{or} \quad \frac{d}{dy} \frac{\varphi(y) - \varphi(x)}{y - x} \geq 0 \quad \text{for} \quad a < x < y < b.
\]

For the first case,
\[
\frac{d}{dx} \frac{\varphi(y) - \varphi(x)}{y - x} = \frac{\varphi(y) - \varphi(x) - \varphi'(x)(y-x)}{(y-x)^2} = \int_0^1 \varphi''(x + t (y-x)) (1-t) \, dt \geq 0.
\]

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 \( \dot{f}(t) = -(v-u)^2 \varphi''(u + t(v-u)) \leq 0\) 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
\[
\dot{f}(t) = \int_{t_0}^t \dot{f}(\tau) \, d\tau.
\]
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,
\[
h(t) = h(0) + \int_0^t h(\tau) \, d\tau = h(0) + th(t) - \int_0^t h(\tau) \tau d\tau
\]
and
\[
h(1) = h(t) + \int_t^1 h(\tau) \, d(\tau - 1) = h(t) - (t-1) h(t) - \int_t^1 h(\tau) (\tau - 1) \, d\tau.
\]
Thus we have shown,
\[
h(t) = h(0) + th(t) - \int_0^t h(\tau) \tau d\tau \quad \text{and}
\]
\[
h(t) = h(1) + (t-1) h(t) + \int_t^1 h(\tau) (\tau - 1) \, d\tau.
\]
So if we multiply the first equation by \((1-t)\) and add to it the second equation multiplied by \(t\) shows,
\[
h(t) = (1-t) h(0) + th(1) - \int_0^t G(t,\tau) \dot{h}(\tau) \, d\tau,
\]
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. (15.31) 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. (15.31) to learn
\[
\varphi(x_0 + t(x_1 - x_0)) = (1-t) \varphi(x_0) + t \varphi(x_1)
\]
\[
- (x_1 - x_0)^2 \int_0^1 G(t,\tau) \varphi(x_0 + \tau(x_1 - x_0)) \, d\tau
\]
\[
\leq (1-t) \varphi(x_0) + t \varphi(x_1),
\]
because \(\varphi \geq 0\) and \(G(t,\tau) \geq 0\).

**Example 15.61.** The functions \(\exp(x)\) and \(-\log(x)\) are convex and \(|x|^p\) is convex iff \(p \geq 1\) as follows from Lemma 15.60.

**Example 15.62 (Proof of Lemma 13.45).** Taking \(\varphi(x) = e^{-x}\) in Lemma 15.60 Eq. (15.29) with \(x_0 = 0\) implies (see Figure 13.1),
\[
1 - x \leq \varphi(x) = e^{-x} \quad \text{for all } x \in \mathbb{R}.
\]
Taking \(\varphi(x) = e^{-2x}\) in Lemma 15.60 Eq. (15.27) with \(x_0 = 0\) and \(x_1 = 1\) implies, for all \(t \in [0,1]\),
\[ e^{-t} \leq \varphi \left( (1-t)0 + t \frac{1}{2} \right) \]
\[ \leq (1-t) \varphi (0) + t \varphi \left( \frac{1}{2} \right) = 1 - t + te^{-1} \leq 1 - \frac{1}{2} t, \]

wherein the last equality we used \( e^{-1} < \frac{1}{2} \). Taking \( t = 2x \) in this equation then gives (see Figure 13.2)

\[ e^{-2x} \leq 1 - x \text{ for } 0 \leq x \leq \frac{1}{2}. \] (15.32)

**Theorem 15.63.** Suppose that \( \varphi : (a, b) \to \mathbb{R} \) is convex and for \( x, y \in (a, b) \) with \( x < y \), let

\[ F(x, y) := \frac{\varphi(y) - \varphi(x)}{y-x}. \]

Then:

1. \( F(x, y) \) is increasing in each of its arguments.
2. The following limits exist,

\[ \varphi'_+ (x) := F(x, x+) := \lim_{y \to x^+} F(x, y) < \infty \text{ and } \]
\[ \varphi'_- (y) := F(y-, y) := \lim_{x \to y^-} F(x, y) > -\infty. \] (15.33)

3. The functions, \( \varphi'_{\pm} \), are both increasing functions and further satisfy,

\[ -\infty < \varphi'_- (x) \leq \varphi'_+ (x) \leq \varphi'_- (y) < \infty \text{ for all } a < x < y < b. \] (15.34)

4. For any \( t \in [\varphi'_- (x), \varphi'_+ (x)] \),

\[ \varphi(y) \geq \varphi(x) + t(y-x) \text{ for all } x, y \in (a, b). \] (15.36)

5. For \( a < \alpha < \beta < b \), let \( K := \max \{|\varphi'_+ (\alpha)|, |\varphi'_- (\beta)|\} \). 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.

\[^{10}\text{The same formula would define } F(x, y) \text{ for } x \neq y. \text{ However, since } F(x, y) = F(y, x), \text{ we would gain no new information by this extension.}\]

**Proof.** BRUCE: The first two items are a repetition of Lemma 15.58. 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

\[ \frac{h_t - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t}, \]

the convexity of \( \varphi \) is equivalent to

\[ \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \leq \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \text{ for all } x_0 \leq x_t \leq x_1 \]

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

\[ \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \leq \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0}. \]

These inequalities may be written more compactly as,

\[ \frac{\varphi(v) - \varphi(u)}{v-u} \leq \frac{\varphi(w) - \varphi(u)}{w-u} \leq \frac{\varphi(w) - \varphi(v)}{w-v}, \] (15.37)

valid for all \( a < u < v < w < b \), again see Figure 15.4. The first (second) inequality in Eq. (15.37) shows \( F(x, y) \) is increasing \((x, y)\). 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 \( t \in [\varphi'_- (x), \varphi'_+ (x)] \). 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. (15.36) holds for \( y \geq x \). Similarly, for \( y < x \),
Corollary 15.64. If \( \varphi : (a, b) \to \mathbb{R} \) is a convex function and \( D \subset (a, b) \) is a dense set, then
\[
\varphi(y) = \sup_{x \in D} \left[ \varphi(x) + \varphi'_+ (x) (y-x) \right] \quad \text{for all } x, y \in (a, b).
\]

Proof. Let \( \psi_\pm(y) := \sup_{x \in D} \left[ \varphi(x) + \varphi_\pm (x) (y-x) \right] \). According to Eq. \((15.36)\) above, we know that \( \varphi(y) \geq \psi_\pm(y) \) for all \( y \in (a, b) \). Now suppose that \( x \in (a, b) \) and \( x_n \in A \) 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 15.65. Suppose that \( \varphi : (a, b) \to \mathbb{R} \) is a non-decreasing function such that
\[
\varphi \left( \frac{1}{2} (x+y) \right) \leq \frac{1}{2} \left[ \varphi(x) + \varphi(y) \right] \quad \text{for all } x, y \in (a, b),
\]
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 \( D_n = \{ \frac{k}{2^n} : 1 \leq k < 2^n \} \). We are going to being by showing Eq. \((15.39)\) implies
\[
\varphi(x_t) \leq (1-t) \varphi(x_0) + t \varphi(x_1) \quad \text{for all } t \in D := \cup_n D_n,
\]
We will do this by induction on \( n \). For \( n = 1 \), this follows directly from Eq. \((15.39)\). So now suppose that Eq. \((15.40)\) holds for all \( t \in D_n \) and now suppose that \( t = \frac{k+1}{2^n} \in D_{n+1} \). Observing that
\[
x_t = \frac{1}{2} \left[ x_{\frac{k}{2^n}} + x_{\frac{k+1}{2^n}} \right],
\]
we may again use Eq. \((15.39)\) to conclude,
\[
\varphi(x_t) \leq \frac{1}{2} \left( \varphi \left( x_{\frac{k}{2^n}} \right) + \varphi \left( x_{\frac{k+1}{2^n}} \right) \right).
\]
Then use the induction hypothesis to conclude,
\[
\varphi(x_t) \leq \frac{1}{2} \left( (1-\frac{k}{2^n}) \varphi(x_0) + \frac{k}{2^n} \varphi(x_1) \right)
\]
\[
= (1-t) \varphi(x_0) + t \varphi(x_1)
\]
as desired.

For general \( t \in (0,1) \), let \( \tau \in D \) such that \( \tau > t \). Since \( \varphi \) is increasing and by Eq. \((15.40)\) 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.
Hilbert Space Basics

Definition 16.1 (Inner Product Space). Let $H$ be a complex vector space. An inner product on $H$ is a function, $\langle \cdot | \cdot \rangle : H \times H \rightarrow \mathbb{C}$, such that

1. $\langle ax + by | z \rangle = a \langle x | z \rangle + b \langle y | z \rangle$ i.e. $x \rightarrow \langle x | z \rangle$ is linear.
2. $\langle x | y \rangle = \langle y | x \rangle$.
3. $\|x\|^2 := \langle x | x \rangle \geq 0$ with $\|x\|^2 = 0$ iff $x = 0$.

Notice that combining properties (1) and (2) that $x \rightarrow \langle z | x \rangle$ 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,

$$\|x + y\|^2 = \langle x+y | x+y \rangle = \|x\|^2 + \|y\|^2 + \langle x | y \rangle + \langle y | x \rangle$$  

$$= \|x\|^2 + \|y\|^2 + 2\Re\langle x | y \rangle. \quad (16.1)$$

Theorem 16.2 (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

$$\|\langle x | y \rangle\| = |ar{\alpha}| \|y\|^2 = \|x\| \|y\|.$$  

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 16.1.) Then

$$0 \leq \|z\|^2 = \left\| x - \frac{\langle x | y \rangle y}{\|y\|^2} \right\|^2 = \|x\|^2 + \frac{|\langle x | y \rangle|^2}{\|y\|^4} \|y\|^2 - 2\Re\langle x | \frac{\langle x | y \rangle y}{\|y\|^2} \rangle y$$  

$$= \|x\|^2 - \frac{|\langle x | y \rangle|^2}{\|y\|^2} \|y\|^2$$

from which it follows that $0 \leq \|y\|^2 \|x\|^2 - |\langle x | y \rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^2 \langle x | y \rangle y$.

**Corollary 16.3.** 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,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\Re\langle x | y \rangle$$

$$\leq \|x\|^2 + \|y\|^2 + 2\|x\||\|y\|| = (\|x\| + \|y\|)^2.$$  

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

$$\langle x + \Delta x | y + \Delta y \rangle - \langle x | y \rangle = \langle x | \Delta y \rangle + \langle \Delta x | y \rangle + \langle \Delta x | \Delta y \rangle$$

$$\leq \|x\||\Delta y\| + \|\Delta x\||y\| + \|\Delta x\||\Delta y\| 

\rightarrow 0 \text{ as } \Delta x, \Delta y \rightarrow 0,$$

from which it follows that $\langle \cdot | \cdot \rangle$ is continuous.

**Definition 16.4.** 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 16.5.** Let $(H, \langle \cdot | \cdot \rangle)$ be an inner product space then
1. (Parallelogram Law)
\[ \|a + b\|^2 + \|a - b\|^2 = 2\|a\|^2 + 2\|b\|^2 \tag{16.2} \]
for all \( a, b \in H \).

2. (Pythagorean Theorem) If \( S \subset H \) is a finite orthogonal set, then
\[ \left\| \sum_{x \in S} x \right\|^2 = \sum_{x \in S} \|x\|^2. \tag{16.3} \]

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;
\[ \|a + b\|^2 + \|a - b\|^2 = \|a\|^2 + \|b\|^2 + 2\text{Re}\langle a|b \rangle + \|a\|^2 + \|b\|^2 - 2\text{Re}\langle a|b \rangle = 2\|a\|^2 + 2\|b\|^2, \]
and
\[ \left\| \sum_{x \in S} x \right\|^2 = \left( \sum_{x \in S} x \right) \sum_{y \in S} \langle x|y \rangle = \sum_{x,y \in S} \langle x|y \rangle \]
\[ = \sum_{x \in S} \langle x|x \rangle = \sum_{x \in S} \|x\|^2. \]

Item 3. is a consequence of the continuity of \( \langle \cdot|\cdot \rangle \) and the fact that
\[ A^\perp = \cap_{x \in A} \text{Null}(\langle \cdot|x \rangle) \]
where \( \text{Null}(\langle \cdot|x \rangle) = \{ y \in H : \langle y|x \rangle = 0 \} \) - a closed subspace of \( H \). Alternatively, if \( x_n \in A^\perp \) and \( x_n \to x \) in \( H \), then
\[ 0 = \lim_{n \to \infty} 0 = \lim_{n \to \infty} \langle x_n|a \rangle = \left\{ \lim_{n \to \infty} x_n \right\} \langle a_a \rangle = \langle x|a \rangle \ \forall \ a \in A \]
which shows that \( x \in A^\perp \). \( \square \)

**Definition 16.6.** A Hilbert space is an inner product space \( (H, \langle \cdot|\cdot \rangle) \) such that the induced Hilbertian norm is complete.

**Example 16.7.** 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 15.28 for the completeness assertion.

**Definition 16.8.** A subset \( C \) of a vector space \( X \) is said to be convex if for all \( x, y \in C \) the line segment \( [x,y] := \{ tx + (1-t)y : 0 \leq t \leq 1 \} \) joining \( x \) to \( y \) is contained in \( C \) as well. (Notice that any vector subspace of \( X \) is convex.)

**Theorem 16.9 (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. (16.2)) and the fact that \( \frac{y+z}{2} \in M \),
\[ \|z - y\|^2 = \|(z - x) + (x - y)\|^2 \]
\[ = 2\|z - x\|^2 + 2\|x - y\|^2 - 2\|\{z - x\} - (x - y)\|^2 \]
\[ = 2\|z - x\|^2 + 2\|x - y\|^2 - 4\|x - y + z\|^2 \]
\[ \leq 2\|z - x\|^2 + 2\|x - y\|^2 - 4\delta^2, \tag{16.4} \]
see Figure 16.2

![Fig. 16.2.](image)

**Uniqueness.** If \( y, z \in M \) minimize the distance to \( x \), then \( \|y - x\| = \delta = \|z - x\| \) and it follows from Eq. (16.4) 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. (16.4) shows
\[ \|y_n - y_m\|^2 \leq 2\delta_n^2 + 2\delta_m^2 - 4\delta^2 \to 0 \text{ as } m, n \to \infty \]
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
\[
g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t \text{Re}(x - y|w) + t^2 \|w\|^2
\]
has a minimum at $t = 0$ and therefore $0 = g'(0) = -2 \text{Re}(x - y|w)$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$, see Figure 16.3. Finally 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 16.10.** 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,
\[
\|x\|^2 = \|w\|^2 + \sum_{i=1}^{n} \langle x|e_i \rangle \langle e_i|x \rangle = d^2(x, M) + \sum_{i=1}^{n} |\langle x|e_i \rangle|^2
\]
and
\[
d^2(x, M) = \|x^2\| - \sum_{i=1}^{n} |\langle x|e_i \rangle|^2.
\]

Suppose $u \in H \setminus \{0\}$, $M = \text{span}\{u\}$, and $\hat{u} := u/\|u\|$. Then from Eq. (16.6),
\[
0 \leq d^2(x, M) = \|x^2\| - |\langle x|\hat{u} \rangle|^2 = \|x^2\| - |\langle x|u \rangle|^2 / \|u\|^2
\]
from which the Cauchy-Schwarz inequality,
\[
|\langle x|u \rangle|^2 \leq \|x\|^2 \|u^2\|
\]
follows. Moreover the proof shows that equality holds iff $x \in M$, i.e. $x = \lambda u$ for some $\lambda \in \mathbb{C}$.

**Notation 16.11** 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 16.1.** Show that a linear operator, $A : X \to Y$, is a bounded iff it is continuous.

**Definition 16.12.** 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 Propostion 16.17 below.) A bounded operator $A : H \to H$ is **self-adjoint** or **Hermitian** if $A = A^*$.

**Definition 16.13.** $H$ 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
\[
\langle x|m \rangle = \langle P_M(x)|m \rangle \text{ for all } m \in M.
\]

Given a linear transformation $A$, we will let Ran($A$) and Nul($A$) denote the **range** and the **null-space** of $A$ respectively.

**Theorem 16.14 (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 $P_Mx$ 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. Ran$(P_M) = M$ and 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. Let \( x, y \in H \), then \( P_M x + \alpha P_M x = x \) and \( P_M x = x \) for all \( x \in M \).

3. Let \( x, y \in H \), then since \( (x - P_M x) \) and \( (y - P_M y) \) are in \( M^\perp \),
\[
\langle P_M x | y \rangle = \langle P_M x P_M y + y - P_M y \rangle = \langle P_M x | P_M y \rangle
\]
\[
= \langle P_M x + (x - P_M x) | P_M y \rangle = \langle x | P_M y \rangle.
\]

4. We have already seen, \( \text{Ran}(P_M) = M \) and \( P_M x = 0 \) iff \( x = 0 \in M^\perp \), i.e. \( \text{Null}(P_M) = M^\perp \).

5. If \( N \subset M \subset H \) it is clear that \( P_M P_N = P_N \) since \( P_M = \text{Id} \) on \( N = \text{Ran}(P_N) \subset M \). 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 \( h \), \( \langle h | n \rangle = \langle P_N x | n \rangle \) 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 \( n \in N \) 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 \( P_N P_M x = P_N x \).

\[\blacksquare\]

Corollary 16.15. 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 \( M \cap M^\perp = \{0\} \).

\[\blacksquare\]

Exercise 16.2. Suppose \( M \) is a subset of \( H \), then \( M^{\perp \perp} = \text{span}(M) \) where (as usual), \( \text{span}(M) \) denotes all finite linear combinations of elements from \( M \).

Theorem 16.16 (Riesz Theorem). Let \( H^* \) be the dual space of \( H \), i.e. \( f \in H^* \) iff \( f : H \rightarrow \mathbb{F} \) is linear and continuous. The map
\[
z \in H \rightarrow \langle z | z \rangle \in H^*
\]
is a conjugate linear 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 \( M = \text{Null}(f) \) – 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 \( x \in H \),
\[
f(x - f(x) e) = f(x) - f(f(x)) = f(x) - f(x) = 0,
\]
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 \( x, z \in H \),
\[
\|z\| = \|x\| = \|z\| \quad \text{for all } x \in H
\]
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.

\[\blacksquare\]

Proposition 16.17 (Adjoints). Let \( H \) and \( K \) be Hilbert spaces and \( A : H \rightarrow K \) be a bounded operator. Then there exists a unique bounded operator \( A^* : K \rightarrow H \) such that
\[
\langle Ax | y \rangle_K = \langle x | A^* y \rangle_H \quad \text{for all } x \in H \text{ and } y \in K.
\]

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 if \(A^*\) has a bounded inverse and \((A^*)^{-1} = (A^{-1})^*\).

**Proof.** For each \(y \in K\), the map \(x \to \langle Ax|y\rangle_K\) is in \(H^*\) and therefore there exists, by Theorem 16.16, a unique vector \(z \in H\) (we will denote this \(z\) by \(A^*(y)\)) such that

\[
\langle Ax|y\rangle_K = \langle x|z\rangle_H \quad \text{for all} \quad x \in H.
\]

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 \(x \in H\),

\[
\langle Ax|y_1 + \lambda y_2\rangle_K = \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K = \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_K = \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_H
\]

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

\[
\langle A^*y|x\rangle_H = \langle x|A^*y\rangle_H = \langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_K
\]

it follows that \(A^{**} = A\). The assertion that \((A + \lambda B)^* = A^* + \bar{\lambda}B^*\) is Exercise 16.3

**Items 3. and 4.** Making use of Schwarz’s inequality (Theorem 16.2), we have

\[
\|A^*\| = \sup_{k \in K; \|k\| = 1} \|A^*k\| = \sup_{k \in H; \|k\| = 1} \sup_{h \in H; \|h\| = 1} |\langle A^*k|h\rangle| = \sup_{h \in H; \|h\| = 1} \sup_{k \in K; \|k\| = 1} |\langle k|Ah\rangle| = \sup_{h \in H; \|h\| = 1} \|Ah\| = \|A\|
\]

so that \(\|A^*\| = \|A\|\). Since

\[
\|A^*A\| \leq \|A^*\| \|A\| = \|A\|^2
\]

and

\[
\|A^*A\| \leq \|A\|^2
\]

we also have \(\|A^*A\| \leq \|A\|^2 \leq \|A^*A\|^2\) which shows \(\|A\|^2 = \|A^*A\|^2\).

Alternatively, from Eq. 16.10,

\[
\|A\|^2 \leq \|A^*A\| \leq \|A\| \|A^*\|
\]

which then implies \(\|A\| \leq \|A^*\|\). Replacing \(A\) by \(A^*\) in this last inequality shows \(\|A^*\|^2 \leq \|A\|\) and hence that \(\|A^*\| = \|A\|\). Using this identity back in Eq. 16.11 proves \(\|A\|^2 = \|A^*A\|\).

Now suppose that \(K = H\). Then

\[
\langle ABh|k\rangle = \langle Bk|A^*h\rangle = \langle h|B^*A^*k\rangle
\]

which shows \((AB)^* = B^*A^*\). If \(A^{-1}\) exists then

\[
(A^{-1})^* A^* = (AA^{-1})^* = I^* = I
\]

and

\[
A^* (A^{-1})^* = (A^{-1}A)^* = I^* = I.
\]

This shows that \(A^*\) is invertible and \((A^*)^{-1} = (A^{-1})^*\). Similarly if \(A^*\) is invertible then so is \(A = A^{**}\).

**Exercise 16.3.** 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 16.4.** 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 16.18.** Suppose \(A : H \to K\) is a bounded operator, then:

1. \(\text{Nul}(A^*) = \text{Ran}(A)^\perp\).
2. \(\text{Ran}(A) = \text{Nul}(A)^\perp\).
3. If \(K = H\) and \(V \subseteq H\) is an \(A\) - invariant subspace (i.e. \(A(V) \subseteq V\)), then \(V^\perp\) is \(A^*\) - invariant.

**Proof.** An element \(y \in K\) is in \(\text{Nul}(A^*)\) iff \(\langle A^*y|x\rangle = \langle y|Ax\rangle\) for all \(x \in H\) which happens iff \(y \in \text{Ran}(A)^\perp\). Because, by Exercise 16.2, the map \(\text{Ran}(A) = \text{Ran}(A)^\perp\), and so by the first item, \(\text{Ran}(A) = \text{Nul}(A)^\perp\). Now suppose \(A(V) \subseteq V\) and \(y \in V^\perp\), then

\[
\langle A^*y|x\rangle = \langle y|Ax\rangle = 0 \quad \text{for all} \quad x \in V.
\]
Theorem 16.20. The next elementary theorem (referred to as the bounded linear transformation theorem, or B.L.T. theorem for short) is often useful.

**Theorem 16.20 (B. L. T. Theorem).** Suppose that $Z$ is a normed space, $X$ is a Banach space, and $S \subset Z$ is a dense linear subspace of $Z$. If $T : S \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \leq C\|z\|$ for all $z \in S$), then $T$ has a unique extension to an element $\bar{T} : L(S, X)$ and this extension still satisfies $\|\bar{T}z\| \leq C\|z\|$ for all $z \in S$.

**Proof.** Let $z \in Z$ and choose $z_n \in S$ such that $z_n \to z$. Since $\|Tz_m - Tz_n\| \leq C\|z_m - z_n\| \to 0$ 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 $w_n \in S$ is another sequence converging to $z$, then $\|Tz_n - Tw_n\| \leq C\|z_n - w_n\| \to 0$ 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 $\|\bar{T}z\| = \lim_{n \to \infty} \|Tz_n\| \leq \lim_{n \to \infty} C\|z_n\| = C\|z\|$ 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. □

### 16.1 Compactness Results for $L^p$ Spaces*

In this section we are going to identify the sequentially “weak” compact subsets of $L^p(\Omega, B, P)$ for $1 \leq p < \infty$, where $(\Omega, B, P)$ is a probability space. The key to our proofs will be the following Hilbert space compactness result.

**Theorem 16.20.** 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 $y_k \rightharpoonup x$.

**Proof.** Let $H_0 := \text{span}\{x_k : k \in \mathbb{N}\}$. 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 $|\langle x_k, h_n \rangle| \leq \|x_k\|\|h_n\| \leq C\|h_n\| < \infty$, the sequence, $\{\langle x_k, h_n \rangle\}_{k=1}^\infty \subset C$, 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 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

$$|\langle y_k, z \rangle - \langle y_l, z \rangle| = |\langle y_k - y_l, z \rangle| \leq |\langle y_k - y_l, h_n \rangle| + |\langle y_k, z \rangle - \langle y_l, z - h_n \rangle|$$

$$\leq |\langle y_k, h_n \rangle| + 2C\|z - h_n\|.$$  

Letting $k, l \to \infty$ in this estimate then shows

$$\lim_{k, l \to \infty} \sup_{z \in H_0} |\langle y_k, z \rangle - \langle y_l, z \rangle| \leq 2C\|z - h_n\|.$$  

Since we may choose $n \in \mathbb{N}$ such that $\|z - h_n\|$ is as small as we please, we may conclude that $\lim_{k \to \infty} \langle y_k, z \rangle = \sup_{z \in H_0} |\langle y_k, z \rangle - \langle y_l, z \rangle|$, i.e. $\varphi(z) := \lim_{k \to \infty} \langle y_k, z \rangle$ exists.

The function, $\varphi(z) = \lim_{k \to \infty} \langle z, y_k \rangle$ is a bounded linear functional on $H$ because $|\varphi(z)| = \lim_{k \to \infty} \|z - y_k\| \leq C\|z\|$.

Therefore by the Riesz Theorem [16.16] there exists $x \in H_0$ such that $\varphi(z) = \langle z, x \rangle$ for all $z \in H_0$. Thus, for this $x \in H_0$ we have shown

$$\lim_{k \to \infty} \langle y_k, z \rangle = \langle x, z \rangle$$

for all $z \in H_0$. (16.12)

To finish the proof we note only observe that Eq. (16.12) 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

$$\lim_{k \to \infty} \langle y_k, z \rangle = \lim_{k \to \infty} \langle y_k, z_0 \rangle = \langle x, z_0 \rangle = \langle x, z \rangle$$

for all $z \in H$.

Since unbounded subsets of $H$ are clearly not sequentially weakly compact, Theorem [16.20] states that a set is sequentially precompact in $H$ iff it is bounded. Let us now use Theorem [16.20] to identify the sequentially compact subsets of $L^p(\Omega, B, P)$ for all $1 \leq p < \infty$. We begin with the case $p = 1$.

**Theorem 16.21.** If $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, B, P)$, there exists a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ and $X \in L^1(\Omega, B, P)$ such that

$$\lim_{k \to \infty} E|Y_k| = E|X|$$

for all $h \in B_0$. (16.13)

**Proof.** For each $m \in \mathbb{N}$ let $X_m := X_{1, X_1 \leq m}$. The truncated sequence $\{X_m\}_{m=1}^\infty$ is a bounded subset of the Hilbert space, $L^2(\Omega, B, P)$, for all $m \in \mathbb{N}$. Therefore by Theorem [16.20] $\{X_m\}_{m=1}^\infty$ has a weakly convergent sub-sequence
for all $m \in \mathbb{N}$. By Cantor’s diagonalization argument, we can find $Y_k^m := X_{n_k}^m \in L^2(\Omega, \mathcal{B}, P)$ such that $Y_k^m \xrightarrow{\text{a.s.}} X^m$ 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
\[
\mathbb{E}\left[\left(\mathbb{E}\left[\left(Y_k^M - Y_k^m\right)^+ h\right]\right)\right] \leq \liminf_{k \to \infty} \mathbb{E}\left[\left(Y_k^M - Y_k^m\right)^+ |h|\right] \\
\leq \|h\|_{\infty} \cdot \liminf_{k \to \infty} \mathbb{E}\left[|Y_k^m| : M > |Y_k^m| > m\right] \\
\leq \|h\|_{\infty} \cdot \liminf_{k \to \infty} \mathbb{E}\left[|Y_k^m| : |Y_k^m| > m\right].
\]

Taking $h = \text{sgn}(X^M - X^m)$ in this inequality shows
\[
\mathbb{E}\left[\left|X^M - X^m\right|\right] \leq \liminf_{k \to \infty} \mathbb{E}\left[|Y_k^m| : |Y_k^m| > m\right]
\]
with the right member of this inequality going to zero as $m, M \to \infty$ with $M \geq m$ 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. (16.13) is valid. For $h \in \mathcal{B}_b$,
\[
\mathbb{E}[X^m - X^m] \leq \liminf_{k \to \infty} \mathbb{E}[|Y_k^m| : |Y_k^m| > m]
\]
with the right member of this inequality going to zero as $m, M \to \infty$ with $M \geq m$ by the assumed uniform integrability of the $\{X_n\}$. Therefore there exists $X \in L^1(\Omega, \mathcal{B}, P)$ such that $\lim_{m \to \infty} \mathbb{E}|X - X^m| = 0$.

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
\[
0 = \mathbb{E}[X \cdot 1] = \lim_{k \to \infty} \mathbb{E}[Y_k^m \cdot 1] = 1
\]
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 16.21.

When $1 < p < \infty$, the situation is simpler.

**Theorem 16.23.** 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
\[
\lim_{k \to \infty} \mathbb{E}|Y_k h| = \mathbb{E}|X h| \text{ for all } h \in L^q(\Omega, \mathcal{B}, P).
\]

**Proof.** Let $C := \sup_{n \in \mathbb{N}} \|X_n\|_p < \infty$ and recall that Lemma 15.54 guarantees that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$. Therefore by Theorem 16.21, there exists $X \in L^1(\Omega, \mathcal{B}, P)$ and a subsequence, $Y_k := X_{n_k}$, such that Eq. (16.13) holds. We will complete the proof by showing:

a) For $h \in L^q(\Omega, \mathcal{B}, P)$ and b) Eq. (16.14) is valid.

a) For $h \in \mathcal{B}_b$ we have
\[
\mathbb{E}|X h| \leq \liminf_{k \to \infty} \mathbb{E}|Y_k h| \leq \liminf_{k \to \infty} \|Y_k|_p \cdot \|h|_q \leq C \|h\|_q.
\]

For $M < \infty$, taking $h = \text{sgn}(X)|X|^{p-1} 1_{|X| \leq M}$ in the previous inequality shows
\[
\mathbb{E}|X|^p 1_{|X| \leq M} \leq C \left\|\text{sgn}(X)|X|^{p-1} 1_{|X| \leq M}\right\|_q \\
= C \left\|X^{(p-1)q} 1_{|X| \leq M}\right\|_q^{1/q} \leq C \left(\mathbb{E}|X|^p 1_{|X| \leq M}\right)^{1/q}
\]
from which it follows that
\[
\left(\mathbb{E}|X|^p 1_{|X| \leq M}\right)^{1/p} \leq \left(\mathbb{E}|X|^p 1_{|X| \leq M}\right)^{1-1/q} \leq C.
\]

Using the monotone convergence theorem, we may let $M \to \infty$ in this equation to find $\|X\|_p = \left(\mathbb{E}|X|^p\right)^{1/p} \leq C < \infty$.

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 $h^M := h|_{|h| \leq M}$, then
\[
\mathbb{E}|X - Y_k h| \leq \mathbb{E}|(X - Y_k)^+ h^M| + \mathbb{E}|(X - Y_k)h^M h^M| \\
\leq \mathbb{E}|(X - Y_k)^+ h^M| + \|X - Y_k\|_p \|h^M h^M\|_q \\
\leq \mathbb{E}|(X - Y_k)^+ h^M| + 2C \|h^M h^M\|_q.
\]
Since $h^M \in B_b$, we may pass to the limit $k \to \infty$ in the previous inequality to find,
\[
\limsup_{k \to \infty} |E [(X - Y_k) h]| \leq 2C \|h1_{|h|>M}\|_q.
\]
This completes the proof, since $\|h1_{|h|>M}\|_q \to 0$ as $M \to \infty$ by DCT.

16.2 Exercises

Exercise 16.5. 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 16.6 (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 $1 \leq m \leq n < \infty$,
2. $(x_n - x_m) \perp M_m$ for all $n \geq m$,
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 16.24. Let $H = \ell^2 := L^2(\mathbb{N}, \text{counting measure})$,
\[
M_n = \{(a(1), \ldots, a(n), 0, 0, \ldots) : a(i) \in \mathbb{C} \text{ for } 1 \leq i \leq n\},
\]
and $x_n(i) = 1_{i \leq n}$, then $x_n = P_m x_n$ for all $n \geq m$ while $\|x_n\|^2 = n \uparrow \infty$ as $n \to \infty$. Thus, we cannot drop the assumption that $\sup_n \|x_n\| < \infty$ in Exercise 16.6.

The rest of this section may be safely skipped.

Exercise 16.7. *Suppose that $(X, \|\cdot\|)$ is a normed space such that parallelogram law, Eq. (16.2), holds for all $x, y \in X$, then there exists a unique inner product on $(\cdot|\cdot)$ 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 \(f \in \mathcal{G}_0\) iff \(f : \Omega \to \mathbb{C}\) is bounded and \(f\) is \((\mathcal{G}, \mathcal{B}_C)\) - measurable. If \(A \in \mathcal{B}\) and \(P(A) > 0\), we will let

\[
\mathbb{E}[X|A] := \frac{\mathbb{E}[X \cdot 1_A]}{P(A)} \quad \text{and} \quad P(B|A) := \frac{P(A \cap B)}{P(A)}
\]

for all integrable random variables, \(X\), and \(B \in \mathcal{B}\). We will often use the factorization Lemma \([40]\) in this section. Because of this let us repeat it here.

**Lemma 17.1.** Suppose that \((\mathcal{Y}, \mathcal{F})\) is a measurable space and \(Y : \Omega \to \mathcal{Y}\) is a map. Then to every \((\sigma(Y), \mathcal{B}_Y)\) - measurable function, \(H : \Omega \to \mathbb{R}\), there is a \((\mathcal{F}, \mathcal{B}_Y)\) - measurable function \(h : \mathcal{Y} \to \mathbb{R}\) 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 \(A_i = 1_{B_i} \circ Y\) and hence \(H = h \circ Y\) with \(h := \sum a_i 1_{B_i}\) - a simple function on \(\mathbb{R}\).

For a general \((\mathcal{F}, \mathcal{B}_Y)\) - measurable function, \(H\), from \(\Omega \to \mathbb{R}\), choose simple functions \(h_n\) converging to \(H\). Let \(h_n : \mathcal{Y} \to \mathbb{R}\) 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 \(\mathcal{Y}\) to \(\mathbb{R}\). \(\square\)

**Definition 17.2 (L² – 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 17.3 (Basic Properties of \(\mathbb{E}_\mathcal{G}\)).** Some remarks on this definition are in order. Let \(f \in L^2(\Omega, \mathcal{B}, P)\).

1. We are identifying \(L^2(\Omega, \mathcal{G}, P)\) with its image,

\[
M := \left\{g_{\mathcal{G}} : g \in L^2(\Omega, \mathcal{G}, P)\right\},
\]

in \(L^2(\Omega, \mathcal{B}, P)\). From Lemma \([26]\) we know that \(M\) is a closed subspace of \(L^2(\Omega, \mathcal{B}, P)\).

2. Thus given \(f \in L^2(\Omega, \mathcal{B}, P)\), we may compute the orthogonal projection (Theorem \([14]\)) 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 \(f \in L^2(\Omega, \mathcal{G}, P)\) 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 \([14]\) 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) \(|f - F|_2 \leq |f - g|_2\) for all \(g \in L^2(\Omega, \mathcal{G}, P)\) or

   b) \(E[fh] = E[Fh]\) for all \(h \in L^2(\Omega, \mathcal{G}, P)\).

4. \(L^1(\mathbf{P})\) – contractivity: \(E[|\mathbb{E}_\mathcal{G} f|] \leq E[|f|]\) 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 \(h := 1_{F \neq 0} \mathbb{F}\) in item 3b. above to find;

\[
E[|F|] = E \left[ F 1_{F \neq 0} \mathbb{F} \right] = E[fh] \leq E[|fh|] \leq E[|f|].
\]

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 \([14]\)

\[
\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^2(\Omega, \mathcal{B}, P). \quad (17.1)
\]

**Lemma 17.4.** If \(f \in L^2(\Omega, \mathcal{B}, P)\) and \(F \in L^2(\Omega, \mathcal{G}, P)\) then the following are equivalent;

\[
E[f : A] = E[F : A] \text{ for all } A \in \mathcal{G}, \quad (17.2)
\]

\[
E[fh] = E[Fh] \text{ for all } h \in \mathcal{G}_0, \quad (17.3)
\]

\(F = \mathbb{E}_\mathcal{G} f\) a.s.

**Proof.** If Eq. \([17.2]\) holds, then by linearity we have \(E[fh] = E[Fh]\) for all \(\mathcal{G}\) - measurable simple functions, \(h\) and hence by the approximation Theorem \([39]\) and the DCT for all \(h \in \mathcal{G}_0\). Therefore Eq. \([17.2]\) implies Eq. \([17.3]\). If Eq. \([17.3]\) holds and \(h \in L^2(\Omega, \mathcal{G}, P)\), we may use DCT to show
by Condition 3b of Remark 17.3 shows \( F = \mathbb{E}_G f \) a.s. Taking \( h = 1_A \) with \( A \in \mathcal{G} \) Condition 3b, or Remark 17.3 we learn that Eq. (17.2) is satisfied as well.

**Theorem 17.5.** 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 \( \mathbb{E}_G : L^2(\Omega, \mathcal{B}, P) \to L^2(\Omega, \mathcal{G}, P) \) 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 \( f \geq 0 \), \( P \)-a.s. then \( \mathbb{E}_G f \geq 0 \), \( P \)-a.s.
2. Monotonicity. If \( f \geq g \), \( P \)-a.s. then \( \mathbb{E}_G f \geq \mathbb{E}_G g \), \( P \)-a.s.
3. \( L^\infty \) – contraction property. \( |\mathbb{E}_G f| \leq \mathbb{E}_G |f| \), \( P \)-a.s.
4. Averaging Property. If \( f \in L^1(\Omega, \mathcal{B}, P) \) then \( F = \mathbb{E}_G f \) iff \( F \in L^1(\Omega, \mathcal{G}, P) \) and

\[
\mathbb{E}(F h) = \mathbb{E}(f h) \quad \text{for all } h \in \mathcal{G}_b.
\]

5. Pull out property or product rule. If \( g \in \mathcal{G}_b \) and \( f \in L^1(\Omega, \mathcal{B}, P) \), then

\[
\mathbb{E}_G(fg) = \mathbb{E}_G|f| \mathbb{E}_G|g|, \quad P \text{-a.s.}
\]

6. Tower or smoothing property. If \( \mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B} \).

Then

\[
\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \quad a.s. \quad \text{for all } f \in L^1(\Omega, \mathcal{B}, P).
\]

**Proof.** By the definition of orthogonal projection, \( f \in L^2(\Omega, \mathcal{B}, P) \) and \( h \in \mathcal{G}_b \),

\[
\mathbb{E}(f h) = \mathbb{E}(f \cdot \mathbb{E}_G h) = \mathbb{E}(\mathbb{E}_G f \cdot h).
\]

Taking

\[
h = \text{sgn}(\mathbb{E}_G f) \cdot \mathbb{E}_G f \cdot 1_{\mathbb{E}_G f > 0}
\]

in Eq. (17.6) shows

\[
\mathbb{E}(|\mathbb{E}_G f|) = \mathbb{E}(\mathbb{E}_G f \cdot h) = \mathbb{E}(f h) \leq \mathbb{E}(|f h|) = \mathbb{E}(|f|).
\]

It follows from this equation and the BLT (Theorem 16.19) that \( \mathbb{E}_G \) 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. (17.6) remains valid for all \( f \in L^1(\Omega, \mathcal{B}, P) \) and \( h \in \mathcal{G}_b \). Indeed, (without reference to Theorem 16.19) if \( f_n := f 1_{|f| \leq n} \in L^2(\Omega, \mathcal{B}, P) \), then \( f_n \to f \) in \( L^1(\Omega, \mathcal{B}, P) \) and hence

\[
\mathbb{E}(|\mathbb{E}_G f_n - \mathbb{E}_G f_m|) = \mathbb{E}(|\mathbb{E}_G (f_n - f_m)|) \leq \mathbb{E}(|f_n - f_m|) \to 0 \quad \text{as } m, n \to \infty.
\]

By the completeness of \( L^1(\Omega, \mathcal{G}, P) \), \( F := L^1(\Omega, \mathcal{G}, P) \)-\( \lim_{n \to \infty} \mathbb{E}_G f_n \) exists. Moreover the function \( F \) satisfies,

\[
\mathbb{E}(F \cdot h) = \mathbb{E}(\lim_{n \to \infty} \mathbb{E}_G f_n \cdot h) = \lim_{n \to \infty} \mathbb{E}(f_n \cdot h) = \mathbb{E}(f \cdot h)
\]

for all \( h \in \mathcal{G}_b \) and by Proposition 10.22 there is at most one, \( F \in L^1(\Omega, \mathcal{G}, P) \), which satisfies Eq. (17.9). We will again denote \( F \) by \( \mathbb{E}_G f \). This proves the existence and uniqueness of \( F \) satisfying the defining relation in Eq. (17.4) of item 4. The same argument used in Eq. (17.8) again shows \( \mathbb{E}|F| \leq \mathbb{E}|f| \) and therefore that \( \mathbb{E}_G : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P) \) is a contraction.

Items 1 and 2. If \( f \in L^2(\Omega, \mathcal{B}, P) \) with \( f \geq 0 \), then

\[
\mathbb{E}(\mathbb{E}_G f \cdot h) = \mathbb{E}(f h) \geq 0 \quad \text{a.e.} \quad \text{in Eq. (17.10)}
\]

An application of Lemma 10.23 then shows that \( \mathbb{E}_G f \geq 0 \) a.s.\(^1\) The proof of item 2. follows by applying item 1. with \( f \) replaced by \( f - g \geq 0 \).

Item 3. If \( f \) is real, \( |f| \leq \mathbb{E}_G |f| \), \( P \)-a.s. \( \mathbb{E}_G |f| \leq \mathbb{E}_G |f| \), \( P \)-a.s. For complex \( f \), let \( h \geq 0 \) be a bounded and \( \mathcal{G} \)-measurable function. Then

\[
\mathbb{E}(|\mathbb{E}_G f|) = \mathbb{E}[\mathbb{E}_G f \cdot \text{sgn} (\mathbb{E}_G f) h] = \mathbb{E}[(f \cdot \text{sgn} (\mathbb{E}_G f) h] \leq \mathbb{E}|f| h = \mathbb{E}(|f| h).
\]

Since \( h \geq 0 \) is an arbitrary \( \mathcal{G} \)-measurable function, it follows, by Lemma 10.23 that \( \mathbb{E}_G f \leq \mathbb{E}_G |f| \), \( P \)-a.s. Recall the item 4. has already been proved.

Item 5. If \( h \in \mathcal{G}_b \) and \( f \in L^1(\Omega, \mathcal{B}, P) \), then

\[
\mathbb{E}([g \mathbb{E}_G f] h) = \mathbb{E}[\mathbb{E}_G f \cdot h g] = \mathbb{E}[f \cdot h g] = \mathbb{E}g \mathbb{E}_G f h = \mathbb{E} [g \mathbb{E}_G (f) h].
\]

Thus \( \mathbb{E}_G (g f) = g \cdot \mathbb{E}_G f \), \( P \)-a.s.

Item 6., by the item 5. of the projection Theorem 16.14 Eq. (17.5) 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. (17.5) continues to hold on \( L^1(\Omega, \mathcal{B}, P) \).

**Second Proof.** For \( h \in (\mathcal{G}_0)_b \), we have

\[
\mathbb{E}[\mathbb{E}_G \mathbb{E}_G f \cdot h] = \mathbb{E}[\mathbb{E}_G f \cdot h] = \mathbb{E}[f \cdot h] = \mathbb{E}[\mathbb{E}_G f \cdot h]
\]

which shows \( \mathbb{E}_G \mathbb{E}_G f = \mathbb{E}_G f \) a.s. By the product rule in item 5., it also follows that

\[
\mathbb{E}_G [\mathbb{E}_G f] = \mathbb{E}_G [\mathbb{E}_G f \cdot 1] = \mathbb{E}_G f \cdot \mathbb{E}_G [1] = \mathbb{E}_G f \) a.s.

Notice that \( \mathbb{E}_G [\mathbb{E}_G f] \) need only be \( \mathcal{G}_1 \)-measurable. What the statement says there are representatives of \( \mathbb{E}_G [\mathbb{E}_G f] \) which \( \mathcal{G}_0 \)-measurable and any such representative is also a representative of \( \mathbb{E}_G f \).

\(^1\) This can also easily be proved directly here by taking \( h = 1_{\mathbb{E}_G f < 0} \) in Eq. (17.10).
Remark 17.6. There is also another standard construction of $\mathbb{E}_G f$ based on the characterization in Eq. (17.4) and the Radon Nikodym Theorem 18.8 below. It goes as follows, for $0 \leq f \in L^1 (P)$, let $Q$ be the measure defined by $dQ := fdP$. Then $Q|_G \ll P|_G$ and hence there exists $0 \leq g \in L^1 (\Omega, G, P)$ such that $dQ|_G = gdP|_G$. This then implies that

$$\int_A f dP = Q(A) = \int_A gdP$$

for all $A \in G$, i.e. $g = \mathbb{E}_G f$. For general real valued, $f \in L^1 (P)$, define $\mathbb{E}_G f = \mathbb{E}_G f_+ - \mathbb{E}_G f_-$ and then for complex $f \in L^1 (P)$ let $\mathbb{E}_G f = \mathbb{E}_G \text{Re } f + i\mathbb{E}_G \text{Im } f$.

Notation 17.7 In the future, we will often write $\mathbb{E}_G f$ as $\mathbb{E} [f|G]$. Moreover, if $(\mathbb{X}, M)$ 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 $P(A|G) := \mathbb{E} [1_A|G]$ be the conditional probability of $A$ given $G$, and $P(A|X) := P(A|\sigma(X))$ be conditional probability of $A$ given $X$.

Exercise 17.1. Suppose $f \in L^1 (\Omega, \mathcal{B}, P)$ and $f > 0$ a.s. Show $\mathbb{E} [f|G] > 0$ a.s. (i.e. show $g > 0$ a.s. for any version, $g$, of $\mathbb{E} [f|G]$.) Use this result to conclude if $f \in (a, b)$ a.s. for some $a, b$ such that $-\infty \leq a < b \leq \infty$, then $\mathbb{E} [f|G] \in (a, b)$ a.s. More precisely you are to show that any version, $g$, of $\mathbb{E} [f|G]$ satisfies, $g \in (a, b)$ a.s.

17.1 Examples

Example 17.8. Suppose $\mathcal{G}$ is the trivial $\sigma$ – algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$. In this case $\mathbb{E}_G f = \mathbb{E} f$ a.s.

Example 17.9. On the opposite extreme, if $\mathcal{G} = \mathcal{B}$, then $\mathbb{E}_G f = f$ a.s.

Exercise 17.2 (Exercise 5.13 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 $\Omega = \bigcup_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the $\sigma$ – algebra generated by $\mathcal{P}$: Show:

1. $B \in \mathcal{G}$ iff $B = \bigcup_{i \in A} A_i$ for some $A \in \mathcal{N}$.
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

$$\mathbb{E}_G f = \sum_{i=1}^\infty \mathbb{E} [f|A_i] 1_{A_i} \text{ a.s.} \quad (17.11)$$

Example 17.10. 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 17.2 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 17.11. 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 $\mathbb{E}_G [f(X)] = \mathbb{E} [f(X)]$ a.s. Conversely if $\mathbb{E}_G [f(X)] = \mathbb{E} [f(X)]$ 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 $\mathbb{E}_G [f(X)] = \mu = \mathbb{E} [f(X)]$ a.s.

Conversely if $\mathbb{E}_G [f(X)] = \mathbb{E} [f(X)]$ a.s. 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 17.12 (Note well.) According to Lemma 17.11, $\mathbb{E} (f|X) = \tilde{f}(X)$ a.s. for some measurable function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$. So computing $\mathbb{E} (f|X) = f(X)$ is equivalent to finding a function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E} [f \cdot h(X)] = \mathbb{E} \tilde{f}(X) h(X) \quad (17.12)$$

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 $f(x) = \mathbb{E} f(X = x)$. 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 17.10. 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. (17.12) 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 17.13. 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, $h : (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ such that

$$
\mathbb{E}[f(X) u(X \wedge t)] = \mathbb{E}[h(X \wedge t) u(X \wedge t)]
$$

for all bounded measurable functions, $u : (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$. Taking $u = 1_{(t)}$ in Eq. (17.13) 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. (17.16) 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,

$$
\mathbb{E}[f(X) | X \wedge t] = 1_{X < t} f(X) + 1_{X \geq t} \mathbb{E}[f(X) | X \geq t]
$$

$$
= 1_{X < t} f(X) + 1_{X \geq t} \mathbb{E}[f(X) | X \geq t].
$$

Exercise 17.3. Let $(\Omega, \mathcal{B}, P)$ be a probability space, $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces, $X : \Omega \to X$ and $Y : \Omega \to Y$ be measurable functions. Let $(X, \mathcal{M})$, $(Y, \mathcal{N})$ be measurable spaces, $(\Omega, \mathcal{F}, P)$ a probability space, and $X : \Omega \to X$ and $Y : \Omega \to Y$ be measurable functions. Further assume that $\mathcal{G} \subseteq \mathcal{F}$ be a $\sigma$-algebra such that $X$ is $\mathcal{G}/M$ measurable and $Y$ is independent of $\mathcal{G}$. Then for any bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ measurable function $f : X \times Y \to \mathbb{R}$ we have

$$
\mathbb{E}[f(X,Y) | \mathcal{G}] = h_f(X) = \mathbb{E}[f(x,Y)] |_{x=X} \text{ a.s.}
$$

where if $\mu := \text{Law}_P(Y)$,

$$
h_f(x) := \mathbb{E}[f(x,Y)] = \int_Y f(x,y) \, d\mu(y). \quad (17.15)
$$

[This exercise is essentially a special case of Exercise 17.5 below.]

Proposition 17.14. Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $(X, \mathcal{M}, \mu)$ and $(\mathcal{Y}, \mathcal{N}, \nu)$ 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

$$
\tilde{\rho}(x) := \int_Y \rho(x,y) \, d\nu(y)
$$

and $x \in \mathbb{X}$ and $B \in \mathcal{N}$, let

$$
Q(x,B) := \begin{cases} \frac{1}{\rho(x)} \int_B \rho(x,y) \, d\nu(y) & \text{if } \tilde{\rho}(x) \neq 0 \\ \delta_{y_0}(B) & \text{if } \tilde{\rho}(x) = 0 \end{cases}
$$

(17.17)

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

$$
\mathbb{E}[f(X,Y) | X] = Q(x,f(x,:)) = \int_Y f(x,y) Q(X,dy) = g(X) \text{ a.s.} \quad (17.18)
$$

As usual we use the notation,

$$
Q(x,v) := \int_Y v(y) Q(x,dy) = \begin{cases} \frac{1}{\rho(x)} \int_Y v(y) \rho(x,y) \, d\nu(y) & \text{if } \tilde{\rho}(x) \neq 0 \\ \delta_{y_0}(v) = v(y_0) & \text{if } \tilde{\rho}(x) = 0 \end{cases}
$$

(17.19)

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 [17.12], we are searching for a bounded measurable function, $g : \mathbb{X} \to \mathbb{R}$, such that

$$
\mathbb{E}[f(X,Y) h(X)] = \mathbb{E}[g(X) h(X)] \text{ for all } h \in \mathcal{M}_b.
$$

(17.19)

(Throughout this argument we are going to repeatedly use the Tonelli - Fubini theorems.) We now explicitly write out both sides of Eq. (17.19):

$$
\mathbb{E}[f(X,Y) h(X)] = \int_{\mathbb{X} \times \mathbb{Y}} h(x) f(x,y) \rho(x,y) \, d\mu(x) \, d\nu(y)
$$

$$
= \int_{\mathbb{X}} h(x) \left[ \int_{\mathbb{Y}} f(x,y) \rho(x,y) \, d\nu(y) \right] d\mu(x) \quad (17.20)
$$

$$
\mathbb{E}[g(X) h(X)] = \int_{\mathbb{X} \times \mathbb{Y}} h(x) g(x) \rho(x,y) \, d\mu(x) \, d\nu(y)
$$

$$
= \int_{\mathbb{X}} h(x) g(x) \tilde{\rho}(x) \, d\mu(x). \quad (17.21)
$$
Since the right sides of Eqs. \((17.20)\) and \((17.21)\) must be equal for all \(h \in \mathcal{M}_b\), we must demand (see Lemma 10.23 and 10.24) that
\[
\int_X f(x, y) \rho(x, y) \, d\nu(y) = g(x) \bar{\rho}(x) \text{ for } \mu - \text{a.e. } x.
\]
\[\text{(17.22)}\]

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
\[
\int_X \bar{\rho}(x) \, d\mu(x) = \int_X \left[ \int_Y \rho(x, y) \, d\nu(y) \right] \, d\mu(x) = 1,
\]
we know that \(\bar{\rho}(x) < \infty\) for \(\mu - \text{a.e. } x\) and therefore it does not matter how \(g\) is defined on \(\{\bar{\rho} = \infty\}\) as long as it is measurable. If
\[
0 = \bar{\rho}(x) = \int_Y \rho(x, y) \, d\nu(y),
\]
then \(\rho(x, y) = 0\) for \(\nu - \text{a.e. } y\) and therefore,
\[
\int_Y f(x, y) \rho(x, y) \, d\nu(y) = 0.
\]
\[\text{(17.23)}\]

Hence Eq. \((17.22)\) will be valid no matter how we choose \(g(x)\) for \(x \in \{\bar{\rho} = 0\}\). So a valid solution of Eq. \((17.22)\) is
\[
g(x) := \begin{cases}
\frac{1}{\bar{\rho}(x)} \int_Y f(x, y) \rho(x, y) \, d\nu(y) & \text{if } \bar{\rho}(x) \in (0, \infty) \\
f(x, y_0) = \delta(y_0(f(x, \cdot))) & \text{if } \bar{\rho}(x) \in [0, \infty] \\
\end{cases}
\]
and with this choice we will have \(\mathbb{E}[f(X,Y)\mid X] = g(X) = Q(X,f)\) a.s. as desired. (Observe here that when \(\bar{\rho}(x) < \infty\), \(\rho(x, \cdot) \in L^1(\nu)\) and hence
\(\int_Y f(x, y) \rho(x, y) \, d\nu(y)\) is a well defined integral.)

It is comforting to observe that
\[
P(X \in \{\bar{\rho} = 0\}) = P(\bar{\rho}(X) = 0) = \int_X 1_{\bar{\rho} = 0} \, d\mu(x) = 0
\]
and similarly
\[
P(X \in \{\bar{\rho} = \infty\}) = \int_X 1_{\bar{\rho} = \infty} \, d\mu(x) = 0.
\]

Thus it follows that \(P(X \in \{x \in X: \bar{\rho}(x) = 0\} = 0\) while the set \(\{x \in X: \bar{\rho}(x) = 0\}\) is precisely where there is ambiguity in defining \(g(x)\).

Just for added security, let us check directly that \(g(X) = \mathbb{E}[f(X,Y)\mid X]\) a.s. According to Eq. \((17.21)\) we have
\[
\mathbb{E}[g(X) \, h(X)] = \int_X h(x) g(x) \bar{\rho}(x) \, d\mu(x)
\]
\[= \int_{X \cap \{0 < \rho < \infty\}} h(x) \bar{\rho}(x) \left( \frac{1}{\bar{\rho}(x)} \int_Y f(x, y) \rho(x, y) \, d\nu(y) \right) \, d\mu(x)
\]
\[= \int_{X \cap \{0 < \rho < \infty\}} h(x) \left( \int_Y f(x, y) \rho(x, y) \, d\nu(y) \right) \, d\mu(x)
\]
\[= \int_X h(x) \left( \int_Y f(x, y) \rho(x, y) \, d\nu(y) \right) \, d\mu(x)
\]
\[= \mathbb{E}[f(X,Y) h(X)] \quad \text{(by Eq. \((17.20)\))},
\]
wherein we have repeatedly used \(\mu(\bar{\rho} = \infty) = 0\) and Eq. \((17.23)\) holds when \(\bar{\rho}(x) = 0\). This completes the verification that \(g(X) = \mathbb{E}[f(X,Y)\mid X]\) a.s.

Proposition 17.14 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 17.14 also gives an example of regular conditional probabilities which we now define.

### 17.1.1 Conditioning Gaussian Random Vectors

**Theorem 17.15.** Suppose that \(Z = (X,Y)^\text{tr}\) is a mean zero Gaussian random vector with \(X \in \mathbb{R}^k\) and \(Y \in \mathbb{R}^l\). Let \(C = C_X := \mathbb{E}[XX^\text{tr}]\) and then let
\[
W := Y - \mathbb{E}[YY^\text{tr}] \mathbb{E}[XX^\text{tr}]^{-1} X
\]
\[\text{(17.24)}\]
where \(C^{-1} = C_{\text{ran}(C)}^{-1} P\) is as in Example 12.50 below. Then \((X,W)^\text{tr}\) is again a Gaussian random vector and moreover \(W\) is independent of \(X\). The covariance matrix for \(W\) is
\[
C_W := \mathbb{E}[WW^\text{tr}] = \mathbb{E}[YY^\text{tr}] - \mathbb{E}[YY^\text{tr}] C^{-1} \mathbb{E}[XY^\text{tr}].
\]
\[\text{(17.25)}\]

**Proof.** Let \(A\) be any \(k \times l\) matrix and let \(W := Y - AX\). Since
\[
\begin{pmatrix}
X \\
W
\end{pmatrix} = \begin{pmatrix}
I & 0 \\
-A & I
\end{pmatrix}
\begin{pmatrix}
X
\end{pmatrix},
\]
according to Lemma 12.42 \((X,W)^\text{tr}\) is still Gaussian. So according to Lemma 13.28 in order to make \(W\) independent of \(X\) it suffices to choose \(A\) so that \(W\) and \(X\) are uncorrelated, i.e.
In the case \( C := \mathbb{E} [XX^\text{T}] \) is non-degenerate, we see that \( A := \mathbb{E} [YY^\text{T}] C^{-1} \) is the desired solution. In fact this works for general \( C \) where \( C^{-1} \) is defined in Example 12.50. To see this is correct, recall

\[
v \cdot Cv = v \cdot \mathbb{E} [XX^\text{T}v] = \mathbb{E} [(v \cdot X)^2]
\]

from which it follows that

\[
\text{Nul} (C) = \{ v \in \mathbb{R}^k : v \cdot X = 0 \}.
\]

Hence it follows that

\[
\mathbb{E} [YY^\text{T}] v = \mathbb{E} [YY^\text{T}] A \mathbb{E} [XX^\text{T}] v \text{ for all } v \in \text{Nul} (C)
\]

no matter how \( A \) is chosen. On the other hand if \( v \in \text{Ran} (C) = \text{Nul} (C)^\perp \),

\[
\mathbb{E} [YY^\text{T}] v = \mathbb{E} [YY^\text{T}] C^{-1} CV = \mathbb{E} [YY^\text{T}] v
\]

as desired.

To prove Eq. (17.26) let \( B := \mathbb{E} [YY^\text{T}] \) so that

\[
W := Y - BC^{-1}X.
\]

We then have

\[
\mathbb{E} [WW^\text{T}] = \mathbb{E} [(Y - BC^{-1}X) (Y - BC^{-1}X)^\text{T}]
= \mathbb{E} [(Y - BC^{-1}X) (YY^\text{T} - X^\text{T}C^{-1}B^\text{T})]
= \mathbb{E} [YY^\text{T} - YX^\text{T}C^{-1}B^\text{T} + BC^{-1}XY^\text{T} + BC^{-1}XX^\text{T}C^{-1}B^\text{T}]
= \mathbb{E} [YY^\text{T}] - BC^{-1}B^\text{T} - BC^{-1}B^\text{T} + BC^{-1}CC^{-1}B^\text{T}
= \mathbb{E} [YY^\text{T}] - BC^{-1}B^\text{T}
= \mathbb{E} [YY^\text{T}] - \mathbb{E} [XX^\text{T}] C^{-1} \mathbb{E} [XY^\text{T}].
\]

**Corollary 17.16.** Suppose that \( Z = (X, Y)^\text{T} \) is a mean zero Gaussian random vector with \( X \in \mathbb{R}^k \) and \( Y \in \mathbb{R}^l \),

\[
A := \mathbb{E} [YY^\text{T}] C^{-1},
C_W := \mathbb{E} [YY^\text{T}] - \mathbb{E} [YY^\text{T}] \mathbb{E} [XY^\text{T}],
\]

and suppose \( W \overset{d}{=} N (C_W, 0) \). If \( f : \mathbb{R}^k \times \mathbb{R}^l \rightarrow \mathbb{R} \) is a bounded measurable function, then

\[
\mathbb{E} [f (X, Y) | X] = \mathbb{E} [f (x, AX + W) | x = X].
\]

As an important special case, if \( x \in \mathbb{R}^k \) and \( y \in \mathbb{R}^l \), then

\[
\mathbb{E} [e^{i(x \cdot X + y \cdot Y)} | X] = e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} \text{Var} (y \cdot W)} = e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} C_W y \cdot y}.
\]

**Proof.** Using the notation in Theorem 17.15

\[
\mathbb{E} [f (X, Y) | X] = \mathbb{E} [f (X, AX + W) | X]
\]

where \( W \overset{d}{=} N (C_W, 0) \) and \( W \) is independent of \( X \). The result now follows by an application of Exercise 17.6. Let us now specialize to the case where \( f (X, Y) = e^{i(x \cdot X + y \cdot Y)} \) in which case

\[
\mathbb{E} [e^{i(x \cdot X + y \cdot Y)} | X] = \mathbb{E} [e^{i(x' \cdot X + y' \cdot Y)} | X] \bigg|_{x' = X} = e^{i(x \cdot X + y \cdot AX)} \mathbb{E} [e^{iy' W}]
= e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} \text{Var} (y \cdot W)} = e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} C_W y \cdot y}.
\]

**Remark 17.17.** Let us go through the above calculations in the special case where \( X \) is non-degenerate. Notice that \( (YY^\text{T})_{ij} = Y_i Y_j \) so that the the condition that \( W := Y - AX \) is independent of \( X \) may be written as

\[
0 = \mathbb{E} [WX^\text{T}] = \mathbb{E} [YY^\text{T}] - \mathbb{E} [XX^\text{T}]
\]

and hence

\[
A := \mathbb{E} [YY^\text{T}] (\mathbb{E} [XX^\text{T}])^{-1} = \mathbb{E} [YY^\text{T}] C_X^{-1}
\]

and so \( W \) defined in Eq. (17.24) is independent of \( X \). Since \( Z = (X, Y)^\text{T} = (X, AX + W)^\text{T} \) it then follows that

\[
\mathbb{E} [f (Z) | X] = \mathbb{E} [f (X, AX + W)^\text{T}) | X] = \mathbb{E} [f (x, AX + W)^\text{T})] |_{x = X}.
\]

In particular we have,
\[ \mathbb{E} \left[ e^{i(x \cdot X + y \cdot Y)} | X \right] = \mathbb{E} \left[ e^{i(x \cdot X + y \cdot AX + y \cdot W)} | X \right] = e^{i(x \cdot X + y \cdot AX)} \mathbb{E} \left[ e^{iy \cdot W} | X \right] = e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} \text{Var}(y \cdot W)} = e^{i(x \cdot X + y \cdot AX)} e^{-\frac{1}{2} C_{W \cdot W}}. \]

Here we have

\[ C_W = \mathbb{E} \left[ W \cdot W^\text{tr} \right] = \mathbb{E} \left[ (Y - AX) \cdot (Y - AX)^\text{tr} \right] = \mathbb{E} \left[ (Y - AX) \cdot (Y^\text{tr} - X^\text{tr} A^\text{tr}) \right] = \mathbb{E} \left[ Y Y^\text{tr} - Y X^\text{tr} A^\text{tr} - A E \left[ X Y^\text{tr} - X X^\text{tr} A^\text{tr} \right] \right] = C_Y - \mathbb{E} \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] - \mathbb{E} \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] + E \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] \]

\[ = C_Y - \mathbb{E} \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] - \mathbb{E} \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] + E \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] \]

\[ = C_Y - \mathbb{E} \left[ Y X^\text{tr} \right] C_X^{-1} E \left[ X Y^\text{tr} \right] \]

From all of this it now follows that

\[ \mathbb{E} \left[ f \left( Z \right) \mid X \right] = \mathbb{E} \left[ f \left( AX + W \right) \mid X \right] = \mathbb{E} \left[ f \left( Ax + W \right) \mid x = X \right] = \mathbb{E} \left[ f \left( Ax + W \right) \mid x \right] \]

\[ \text{where } W \overset{d}{=} N \left( 0, C_W \right). \]

**Remark 17.18.** Suppose that \( Z : \Omega \rightarrow \mathbb{R}^N \) is a mean zero Gaussian random vector, \( C_Z = \mathbb{E} \left[ Z Z^\text{tr} \right] \) so that

\[ \mathbb{E} \left[ (Z \cdot a) \right] \left( Z \cdot b \right) \] = \[ a \cdot C_Z b \forall a, b \in \mathbb{R}^N. \]

Let us assume that \( C_Z > 0 \), i.e. \( Z \) is non-degenerate. Further suppose that \( S : \mathbb{R}^N \rightarrow \mathbb{R}^k \) and \( T : \mathbb{R}^N \rightarrow \mathbb{R}^l \) are linear maps, \( X := SZ \), and \( Y = TZ \). Then have

\[ \mathbb{E} \left[ X Y^\text{tr} \right] = \mathbb{E} \left[ SZZ^\text{tr} T^\text{tr} \right] = SC_Z T^\text{tr} \]

From which it follows that \( X \) is independent of \( Y \) iff \( SC_Z T^\text{tr} = 0 \). The above result specialize to

\[ C_X = \mathbb{E} \left[ X X^\text{tr} \right] = \mathbb{E} \left[ SZZ^\text{tr} S^\text{tr} \right] = SC_Z S^\text{tr} \] and

\[ C_Y = \mathbb{E} \left[ Y Y^\text{tr} \right] = \mathbb{E} \left[ TZZ^\text{tr} T^\text{tr} \right] = TC_Z T^\text{tr}. \]

Our next goal is to find \( A : \mathbb{R}^k \rightarrow \mathbb{R}^N \) linear so that

\[ W := Z - AX = \left( I - AS \right) Z \]

is independent of \( X = SZ \) which happens iff

\[ 0 = \left( I - AS \right) C_Z S^\text{tr} = C_Z S^\text{tr} - AC_X \iff AC_X = C_Z S^\text{tr}. \]

Let us observe that \( C_X x = 0 \) implies

\[ 0 = C_X x \cdot x = SC_Z S^\text{tr} x \cdot x = C_Z S^\text{tr} x \cdot S^\text{tr} x \]

and since \( C_Z > 0 \) we may conclude that \( \text{Null} \left( C_X \right) = \text{Null} \left( S^\text{tr} \right) = \text{Ran} \left( S \right) \).

Hence if we assume that \( S \) is surjective it will follows that \( C_X \) is invertible and we must take

\[ A = C_Z S^\text{tr} C_X^{-1}. \]

In summary,

\[ Z = AX + W \]

where \( W \perp \perp X \) and

\[ C_W = \left( I - AS \right) C_Z \left( I - AS \right)^\text{tr} = \left( I - AS \right) C_Z \left( I - S^\text{tr} A^\text{tr} \right) = C_Z - ASC_Z - C_Z S^\text{tr} A^\text{tr} + ASC_Z S^\text{tr} A^\text{tr} = C_Z - ASC_Z - C_Z S^\text{tr} A^\text{tr} + AC_Z A^\text{tr} = C_Z - ASC_Z - C_Z S^\text{tr} A^\text{tr} + C_S S^\text{tr} A^\text{tr} = C_Z - ASC_Z = \left( I - AS \right) C_Z = \left( I - C_Z S^\text{tr} C_X^{-1} \right) C_Z = C_Z - C_Z S^\text{tr} C_X^{-1} SC_Z. \]

**Exercise 17.4.** Suppose now that \( (X, Y, Z)^\text{tr} \) 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

\[ \mathbb{E} \left[ \exp \left( i \left( Y \cdot W + z \cdot Z \right) \right) \mid X \right] = \exp \left( - \text{Cov} \left( y \cdot W_1, z \cdot W_2 \right) \right) \cdot \mathbb{E} \left[ \exp \left( i y \cdot Y \right) \right] \cdot \mathbb{E} \left[ \exp \left( i z \cdot Z \right) \right] \cdot \mathbb{E} \left[ f \left( Y \right) \right] \cdot \mathbb{E} \left[ f \left( Z \right) \right] \cdot \mathbb{E} \left[ f \left( X \right) \right]. \]  

(17.28)

In performing these computations please use the following definitions,

\[ C := C_X := \mathbb{E} \left[ X X^\text{tr} \right], \] (17.29)

\[ A := \mathbb{E} \left[ \begin{bmatrix} Y \\ Z \end{bmatrix} \right] X^\text{tr} C^{-1} = \left[ \begin{bmatrix} E \left[ Y X^\text{tr} \right] C^{-1} \\ E \left[ Z X^\text{tr} \right] C^{-1} \end{bmatrix} \right] = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \] (17.30)

and

\[ W := \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \] = \[ \begin{bmatrix} Y \\ Z \end{bmatrix} - AX \] = \[ \begin{bmatrix} Y - A_1 X \\ Z - A_2 X \end{bmatrix}. \] (17.31)
17.1.2 Probability Kernels and Regular Conditional Distributions

**Definition 17.19.** Let \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) be measurable spaces. A function, \(Q : X \times N \rightarrow [0, 1]\) is a **probability kernel** on \(X \times Y\) if
1. \(Q(x, \cdot) : N \rightarrow [0, 1]\) is a probability measure on \((Y, \mathcal{N})\) for each \(x \in X\) and
2. \(Q(\cdot, B) : X \rightarrow [0, 1]\) is \(\mathcal{M}/\mathcal{B}_R\) measurable for all \(B \in \mathcal{N}\).

If \(Q\) is a probability kernel on \(X \times Y\) and \(f : Y \rightarrow \mathbb{R}\) is a bounded measurable function or a positive measurable function, then \(x \mapsto Q(x, f) := \int_Y f(y) Q(x, dy)\) is \(\mathcal{M}/\mathcal{B}_R\) measurable. This is clear for simple functions and then for general functions via simple limiting arguments.

**Definition 17.20.** Let \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) be measurable spaces and \(X : \Omega \rightarrow X\) and \(Y : \Omega \rightarrow Y\) be measurable functions. A probability kernel, \(Q\), on \(X \times Y\) is said to be a **regular conditional distribution of \(Y\) given \(X\)** if \(Q(X, B)\) is a version of \(P (Y \in B | X)\) for each \(B \in \mathcal{N}\). Equivalently, we should have \(Q(X, f) = (f \circ Y | X)\) a.s. for all \(f \in \mathcal{N}_b\).

The probability kernel, \(Q\), defined in Eq. (17.17) is an example of a regular conditional distribution of \(Y\) given \(X\).

**Remark 17.21.** Unfortunately, regular conditional distributions do not always exist, see Doob [8, p. 624]. However, if we require \(Y\) to be a “standard Borel space,” (i.e. \(Y\) is isomorphic to a Borel subset of \(\mathbb{R}\)), then a conditional distribution of \(Y\) given \(X\) will always exists. See Theorem 17.41 in the appendix to this chapter. Moreover, it is known that “reasonable” measure spaces are standard Borel spaces, see Section 12.11 above for more details. So in most instances of interest a regular conditional distribution of \(Y\) given \(X\) will exist.

**Exercise 17.5 (Jazzed up pull out property).** Let \((\Omega, \mathcal{B}, P)\) be a probability space, \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) be measurable spaces, \(X : \Omega \rightarrow X\) and \(Y : \Omega \rightarrow Y\) be measurable functions, and assume there exists a regular conditional distribution, \(Q\), of \(Y\) given \(X\). Show:

1. For all bounded measurable functions, \(f : (X \times Y, \mathcal{M} \otimes \mathcal{N}) \rightarrow \mathbb{R}\), the function \(X \ni x \mapsto Q(x, f(x, \cdot))\) is measurable and
   \[
   \mathbb{E}[f(X, Y) | X](\omega) = Q(X(\omega), f(X(\omega), \cdot)) = \int_Y Q(X(\omega), dy) f(X(\omega), y) \text{ for } P - \text{a.e. } \omega. \tag{17.32}
   \]
   \[\tag{17.33}
   \]

**Hint:** let \(\mathcal{H}\) denote the set of bounded measurable functions, \(f\), on \(X \times Y\) 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
   \[
P ((X, Y) \in A) = \int_X Q(x, 1_A(x, \cdot)) d\mu(x) = \int_X d\mu(x) \int_Y 1_A(x, y) Q(x, dy). \tag{17.34}
   \]

**Note:** If we take \(X = \Omega\), \(X : \Omega \rightarrow X = \Omega\) 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
   \[
P \{ (\omega : (\omega, Y(\omega)) \in A) \} = \int_A dP(\omega) \int_Y 1_A(\omega, y) Q(\omega, dy). \tag{17.35}
   \]

**Exercise 17.6 (Compare with Exercise 17.3).** Let \((\Omega, \mathcal{B}, P)\) be a probability space, \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) be measurable spaces, \(X : \Omega \rightarrow X\) and \(Y : \Omega \rightarrow Y\) be measurable functions. Further assume that \(X\) and \(Y\) are independent. Find a regular conditional distribution of \(Y\) given \(X\) and prove
   \[
   \mathbb{E}[f(X, Y) | X] = h_f(X) \text{ a.s. } \forall \text{ bounded measurable } f : X \times Y \rightarrow \mathbb{R},
   \]
   where
   \[
h_f(x) := \mathbb{E}[f(x, Y) | X] \text{ for all } x \in X,
   \]
i.e.
   \[
   \mathbb{E}[f(X, Y) | X] = \mathbb{E}[f(x, Y) | X = x] \text{ a.s.}
   \]

**Exercise 17.7.** Suppose \((\Omega, \mathcal{B}, P)\) and \((\Omega', \mathcal{B}', P')\) are two probability spaces, \((X, \mathcal{M})\) and \((Y, \mathcal{N})\) are measurable spaces, \(X : \Omega \rightarrow X' : \Omega' \rightarrow X, Y : \Omega \rightarrow Y, Y' : \Omega \rightarrow Y\) are measurable functions such that \(P \circ (X, Y)^{-1} = P' \circ (X', Y')\), i.e. \((X, Y) \overset{d}{=} (X', Y')\). If \(f : (X \times Y, \mathcal{M} \otimes \mathcal{N}) \rightarrow \mathbb{R}\) is a bounded measurable function and \(\bar{f} : (X, \mathcal{M}) \rightarrow \mathbb{R}\) is a measurable function such that \(\bar{f}(X) = \mathbb{E}[f(X, Y) | X] P - \text{a.s. then}
   \[
   \mathbb{E}[\bar{f}(X', Y')] = \bar{f}(X') P' \text{ a.s.}
   \]

**Exercise 17.8.** 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 \(S_n := X_1 + \cdots + X_n\) and \(B_n := \sigma(S_n, S_{n+1}, S_{n+2}, \ldots)\). Show
   \[
   \mathbb{E}[X_1 | B_n] = \mathbb{E}[X_1 | S_n, S_{n+1}, S_{n+2}, \ldots] = \frac{S_n}{n} \text{ a.s.}
   \]
   [This problem will be used in Example 21.82 to give a proof of the strong law of large numbers.]
   **Hint:** Use Exercise 17.7 to show
   \[
   \mathbb{E}[X_j | B_n] = \mathbb{E}[X_1 | B_n] \text{ a.s. for all } j \leq n. \tag{17.35}
   \]
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, $P_{\mathcal{G}}(B) = \mathbb{E}_P[G1_B] = \mathbb{E}_P[1_B]$, for all $B \in \mathcal{B}$. If $B = \bigcup_{n=1}^{\infty} B_n$ with $B_n \in \mathcal{B}$, then $1_B = \bigcup_{n=1}^{\infty} 1_{B_n}$ and this sum converges in $L^1(\mathcal{P})$ (in fact in all $L^p(\mathcal{P})$) by the DCT. Since $P_{\mathcal{G}} : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P)$ is a contraction and therefore continuous it follows that

$$P_{\mathcal{G}}(B) = \mathbb{E}_P[1_B] = \sum_{n=1}^{\infty} P_{\mathcal{G}}(B_n) \quad (17.36)$$

where all equalities are in $L^1(\Omega, \mathcal{G}, P)$. Now suppose that we have chosen a representative, $P_{\mathcal{G}}(B) : \Omega \to [0,1]$, of $P_{\mathcal{G}}(B)$ for each $B \in \mathcal{B}$. From Eq. (17.36) it follows that

$$P_{\mathcal{G}}(B) (\omega) = \sum_{n=1}^{\infty} P_{\mathcal{G}}(B_n) (\omega) \text{ for } P - \text{a.e. } \omega. \quad (17.37)$$

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, $P_{\mathcal{G}}(B) : \Omega \to [0,1]$, such that Eq. (17.37) holds for all $\omega \in \Omega$ and all $B, B_n \in \mathcal{B}$ with $B = \bigcup_{n=1}^{\infty} B_n$.

**Definition 17.22.** 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

$$Q(\cdot, B) = P(B|\mathcal{G})(\cdot) \text{ a.s. for every } B \in \mathcal{B}. \quad (17.38)$$

This corresponds to the $Q$ in Definition 17.20 provided, $(\mathcal{X}, \mathcal{M}) = (\Omega, \mathcal{G})$, $(\mathcal{Y}, \mathcal{N}) = (\Omega, \mathcal{B})$, and $X(\omega) = Y(\omega) = \omega$ for all $\omega \in \Omega$.

### 17.2 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 17.23.** For the reader’s convenience let us recall Lemma 10.24 asserts that if $f$ and $G$ are non-negative extended $\mathcal{G}$ – measurable functions on $\Omega$ such that $E[f|A] \geq E[G|A]$ 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

$$\infty > n \geq E[F1_{A_n}] \geq E[G1_{A_n}]$$

from which it follows $G1_{A_n}$ is integrable. Therefore, we may now conclude that $E[F1_{A_n} - G1_{A_n}] \geq 0 \implies E[G1_{A_n} - F1_{A_n}] \leq 0$. As $G1_{A_n} - F1_{A_n} \geq 0$ 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 17.24 (Extending $\mathbb{E}_G$).** If $f : \Omega \to [0, \infty]$ is $\mathcal{B}$ - measurable, there is a $\mathcal{G}$ – measurable function, $F : \Omega \to [0, \infty]$, satisfying

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}. \quad (17.39)$$

By Lemma 10.24, the function $F$ is uniquely determined up to sets of measure zero and hence we denote any such version of $F$ by $\mathbb{E}_Gf$.

1. **Properties 2., 5. (with $0 \leq g \in \mathcal{G}_0$), and 6. of Theorem 17.3 still hold for any $\mathcal{B}$ – measurable functions such that $0 \leq f \leq g$. Namely:**
   a) **Order Preserving.** $\mathbb{E}_Gf \leq \mathbb{E}_Gg$ a.s. when $0 \leq f \leq g$.
   b) **Pull out Property.** $\mathbb{E}_G[hf] = h\mathbb{E}_G[f]$ 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
      $$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_Gf = \mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}f = \mathbb{E}_{\mathcal{G}_0}f \text{ a.s}. \quad (17.40)$$

2. **Conditional Monotone Convergence (cMCT).** Suppose that, almost surely, $0 \leq f_n \leq f_{n+1}$ for all $n$, then $\lim_{n \to \infty} \mathbb{E}_Gf_n = \mathbb{E}_G[\lim_{n \to \infty} f_n]$ 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

$$\mathbb{E}_G\left[ \liminf_{n \to \infty} f_n \right] \leq \liminf_{n \to \infty} \mathbb{E}_G[f_n] \text{ a.s.} \quad (17.40)$$

**Proof.** Since $f \wedge n \in L^1(\Omega, \mathcal{B}, P)$ and $f \wedge n$ is increasing, it follows that $F := \uparrow \lim_{n \to \infty} \mathbb{E}_G[f \wedge n]$ exists a.s. Moreover, by two applications of the standard MCT, we have for any $A \in \mathcal{G}$, that

$$\mathbb{E}[F : A] = \lim_{n \to \infty} \mathbb{E}[\mathbb{E}_G[f \wedge n] : A] = \lim_{n \to \infty} \mathbb{E}[f \wedge n : A] = \lim_{n \to \infty} \mathbb{E}[f : A].$$

Thus Eq. (17.39) holds and this uniquely determines $F$ follows from Lemma 10.24.

**Item 1. a) if** $0 \leq f \leq g$, then

$$\mathbb{E}_Gf = \lim_{n \to \infty} \mathbb{E}_G[f \wedge n] \leq \lim_{n \to \infty} \mathbb{E}_G[g \wedge n] = \mathbb{E}_Gg \text{ a.s.}$$

and so $\mathbb{E}_G$ still preserves order. We will prove items 1b and 1c at the end of this proof.
Item 2. Suppose that, almost surely, $0 \leq f_n \leq f_{n+1}$ for all $n$, then $E_G f_n$ is a.s. increasing in $n$. Hence, again by two applications of the MCT, for any $A \in \mathcal{G}$, we have
\[
E \left[ \lim_{n \to \infty} E_G f_n : A \right] = \lim_{n \to \infty} E \left[ E_G f_n : A \right] = \lim_{n \to \infty} E [f_n : A] = E \left[ \lim_{n \to \infty} f_n : A \right] = E \left[ E_G \left( \lim_{n \to \infty} f_n \right) : A \right]
\]
which combined with Lemma 10.24 implies that $\lim_{n \to \infty} E_G f_n = E_G \left[ \lim_{n \to \infty} f_n \right]$ a.s.

Item 3. For $0 \leq f_n$, let $g_k := \inf_{n \geq k} f_n$. Then $g_k \leq f_k$ for all $k$ and $g_k \uparrow \lim_{n \to \infty} f_n$ and hence by cMCT and item 1,

$E_G \left[ \lim_{n \to \infty} f_n \right] = \lim_{k \to \infty} E_G g_k \leq \lim \inf E_G f_n$ a.s.

Item 1. b) If $h \geq 0$ is a $\mathcal{G}$-measurable function and $f \geq 0$, then by cMCT,

$E_G [hf] \leq \lim_{n \to \infty} E_G [(h \wedge n)(f \wedge n)] \leq \lim_{n \to \infty} E_G [(h \wedge n)] \leq \lim_{n \to \infty} E_G [hf] \text{ a.s.}$

Item 1. c) Similarly by multiple uses of cMCT,

$E_G f \leq \lim_{n \to \infty} E_G (f \wedge n) \leq \lim_{n \to \infty} E_G f$ a.s.

and

$E_G f \leq \lim_{n \to \infty} E_G (f \wedge n) \leq \lim_{n \to \infty} E_G f$ a.s.

\[\quad\] \[\text{Theorem 17.25 (Conditional Dominated Convergence (cDCT)). If } f_n \to f \text{ and } |f_n| \leq g \in L^1(\Omega, \mathcal{B}, P), \text{ then } E_G f_n \to E_G f \text{ a.s.}\]

**Proof.** From Corollary 15.9 we know that $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$ and therefore $E_G f_n \to E_G f$ in $L^1(\Omega, \mathcal{B}, P)$ as conditional expectation is a contraction on $L^1(\Omega, \mathcal{B}, P)$. 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 $0 \leq g \pm f_n$ a.s. for all $n$. Hence by cFatou,

$E_G (g + f) = E_G \left[ \lim_{n \to \infty} (g + f_n) \right] \leq \lim_{n \to \infty} E_G (g + f_n) = E_G g + \left\{ \lim_{n \to \infty} \inf E_G (f_n) \text{ in } + \text{ case } \right\} - \lim_{n \to \infty} \sup E_G (f_n) \text{ in } - \text{ case,}
\]

where the above equations hold a.s. C cancellation $E_G g$ from both sides of the equation then implies

$\limsup_{n \to \infty} E_G (f_n) \leq E_G f \leq \liminf_{n \to \infty} E_G (f_n) \text{ a.s.}$

\[\text{Remark 17.26. Suppose that } f_n \to f, |f_n| \leq g_n \in L^1(\Omega, \mathcal{B}, P), \text{ and } g_n \to g \in L^1(\Omega, \mathcal{B}, P) \text{, then by the } \text{DCT in Corollary 15.9 we know that } f_n \to f \text{ in } L^1(\Omega, \mathcal{B}, P). \text{ Since } E_G \text{ is a contraction, it follows that } E_G f_n \to E_G f \text{ in } L^1(\Omega, \mathcal{B}, P) \text{ and hence } E_G f_n \to E_G f.
\]

The next result in Lemma 17.29 shows how to localize conditional expectations. In order to state and prove the lemma we need a little background work first.

**Definition 17.27.** 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 $A$ if $\mathcal{F}_A = \mathcal{G}_A$. Recall that $\mathcal{F}_A = \{B \cap A : B \in \mathcal{F}\}$.

**Lemma 17.28.** 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

$\mathcal{F}_A = \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A$. \hspace{1cm} (17.41)

**Proof.** If $A \in \mathcal{F}$ we have $B \in \mathcal{F}_A$ if 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

$\mathcal{F}_A \cap \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A = [\mathcal{F} \cap \mathcal{G}]_A$.

Equation (17.41) now clearly follows from this identity when $\mathcal{F}_A = \mathcal{G}_A$. \hfill \Box

**Lemma 17.29 (Localizing Conditional Expectations).** Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\mathcal{F}$ and $\mathcal{G}$ be sub-$\sigma$-fields 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

$E_{\mathcal{F}} X = E_{\mathcal{F} \cap \mathcal{G}} X = E_{\mathcal{F} \cap \mathcal{G}} Y = E_{\mathcal{G}} Y$ a.s. on $A$. \hspace{1cm} (17.42)
Proof. It suffices to prove, $E_{F}X = E_{F \cap G}Y$ a.s. on $A$ and this is equivalent to $1_{A}E_{F}X = 1_{A}E_{F \cap G}Y$ a.s. Since

$$1_{A}E_{F \cap G}Y = E_{F \cap G}[1_{A}Y] = E_{F \cap G}[1_{A}X] = 1_{A}E_{F \cap G}[X] \text{ a.s.}$$

it suffices to show

$$E_{F}[1_{A}X] = 1_{A}E_{F}X = 1_{A}E_{F \cap G}[X] \text{ a.s.} \quad (17.43)$$

For $B \in F$, we have by Lemma 17.28 that

$$A \cap B \in \mathcal{F}_{A} = [\mathcal{F} \cap \mathcal{G}]_{A} \subset \mathcal{F} \cap \mathcal{G}$$

and therefore,


from which Eq. (17.43) follows.

Example 17.30. Let us use Lemma 17.29 to show $E[f(X)|X \land t] = f(X) = f(X \land t)$ on $\{X < t\}$ — a fact we have already seen to be true in Example 17.13.

Let us begin by observing that $\{X < t\} = \{X \land t < t\} \subset \sigma(X) \cap \sigma(X \land t)$.

Moreover, using $\sigma(X)_A = \sigma(X|A)$ for all $A \in \mathcal{B}^2$ we see that

$$\sigma(X|\{X < t\}) = \sigma(X|\{X \land t < t\}) = \sigma(X \land t|\{X < t\}).$$

Therefore it follows that

$$E[f(X)|X \land t] = E[f(X)|\sigma(X \land t)] = E[f(X)|\sigma(X)] = f(X) \text{ a.s. on } \{X < t\}.$$ 

What goes wrong with the above argument if you replace $\{X < t\}$ by $\{X \leq t\}$ everywhere? (Notice that the same argument shows; if $X = Y$ on $A \in \sigma(X) \cap \sigma(Y)$ then $E[f(X)|Y] = f(X)$ a.s. on $A$.)

Theorem 17.31 (Conditional Jensen’s inequality). Let $(\Omega, \mathcal{B}, P)$ be a probability space, $-\infty \leq a < b \leq \infty$, 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 $\varphi(\mathbb{E}_G f) \in L^1(\Omega, \mathcal{B}, \mathbb{P})$.

Proof. Let $A := \{a < b\}$ — a countable dense subset of $(a, b)$. By Theorem 15.63 (also see Lemma 15.60) and Figure 15.5 when $\varphi$ is $C^1$,

$$\varphi(y) \geq \varphi(x) + \varphi'_-(x)(y - x) \quad \text{for all } x, y \in (a, b),$$

where $\varphi'_-(x)$ is the left hand derivative of $\varphi$ at $x$. Taking $y = f$ and then taking conditional expectations imply,

$$\mathbb{E}_G \varphi(f) \geq \mathbb{E}_G \varphi(x) + \varphi'_-(x)(\mathbb{E}_G f - x) \text{ a.s.} \quad (17.47)$$

Since this is true for all $x \in (a, b)$ (and hence all $x$ in the countable set, $A$) we may conclude that

$$\mathbb{E}_G \varphi(f) \geq \sup_{x \in A} \left[\varphi(x) + \varphi'_-(x)(\mathbb{E}_G f - x)\right] \text{ a.s.}$$

By Exercise 17.1 $\mathbb{E}_G f \in (a, b)$, and hence it follows from Corollary 15.64 that

$$\sup_{x \in A} \left[\varphi(x) + \varphi'_-(x)(\mathbb{E}_G f - x)\right] = \varphi(\mathbb{E}_G f) \text{ a.s.}$$

Combining the last two estimates proves Eq. (17.44).

From Eq. (17.44) and Eq. (17.46) with $y = \mathbb{E}_G f$ and $x \in (a, b)$ fixed we find,

$$\varphi(x) + \varphi'_-(x)(\mathbb{E}_G f - x) \leq \mathbb{E}_G \varphi(f) \leq \mathbb{E}_G \varphi(f). \quad (17.48)$$

Therefore

$$|\varphi(\mathbb{E}_G f)| \leq |\mathbb{E}_G \varphi(f)| \lor |\varphi(x) + \varphi'_-(x)(\mathbb{E}_G f - x)| \in L^1(\Omega, \mathcal{G}, \mathbb{P}) \quad (17.49)$$

which implies that $\varphi(\mathbb{E}_G f) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. Taking expectations of Eq. (17.44) is now allowed and immediately gives Eq. (17.45).

Remark 17.32 (On Theorem 17.31 and its proof). *

1. From Eq. (17.46),

$$\varphi(f) \geq \varphi(\mathbb{E}_G f) + \varphi'_-(\mathbb{E}_G f)(f - \mathbb{E}_G f). \quad (17.50)$$

Therefore taking $\mathbb{E}_G$ of this equation “implies” that

$$\mathbb{E}_G \varphi(f) \geq \varphi(\mathbb{E}_G f) + \mathbb{E}_G \left[\varphi'_-(\mathbb{E}_G f)(f - \mathbb{E}_G f)\right]$$

$$= \varphi(\mathbb{E}_G f) + \varphi'_-(\mathbb{E}_G f)\mathbb{E}_G [(f - \mathbb{E}_G f)] = \varphi(\mathbb{E}_G f). \quad (17.51)$$

The technical problem with this argument is the justification that $\mathbb{E}_G \left[\varphi'_-(\mathbb{E}_G f)(f - \mathbb{E}_G f)\right] = \varphi'_-(\mathbb{E}_G f)\mathbb{E}_G [(f - \mathbb{E}_G f)]$ since there is no reason for $\varphi'_-$ to be a bounded function. The proof we give in Theorem 17.31 circumvents this technical detail.
2. On the other hand let us now suppose that \( \varphi \) is \( C^1(\mathbb{R}) \) is convex and for the moment that \( |f| \leq M < \infty \) a.s. Then \( \mathbb{E}_G f \in [-M, M] \) a.s. and hence \( \varphi'(\mathbb{E}_G f) = \varphi'(\mathbb{E}_G f) \) is bounded and Eq. (17.51) is now valid. Moreover, taking \( x = 0 \) in Eq. (17.48) shows

\[
\varphi(0) + \varphi'(0)\mathbb{E}_G f \leq \varphi(\mathbb{E}_G f) \leq \mathbb{E}_G [\varphi(f)].
\]

If \( f \) is unbounded we may apply the above inequality with \( f \) replaced by \( f_M := f \cdot 1_{|f| \leq M} \) in order to conclude,

\[
\varphi(0) + \varphi'(0)\mathbb{E}_G f_M \leq \varphi(\mathbb{E}_G f_M) \leq \mathbb{E}_G [\varphi(f_M)].
\]

If we further assume that \( \varphi(f_M) \geq 0 \) 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

\[
\varphi(0) + \varphi'(0)\mathbb{E}_G f \leq \varphi(\mathbb{E}_G f) \leq \mathbb{E}_G [\varphi(f)]
\]

where we used \( \mathbb{E}_G [\varphi(f_M)] \to \mathbb{E}_G [\varphi(f)] \) 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 \( 0 \leq f < \infty \) be a measurable function. The convexity of \( \varphi \) may be stated as,

\[
\varphi(c) + \varphi'(c)(x - c) \leq \varphi(x)
\]

for all \( x, c \geq 0 \).

So letting \( f_n := f 1_{f \leq n} \), if \( f = f_n \) and \( c = \mathbb{E}_G f_n \) in the previous equation, then

\[
\varphi(\mathbb{E}_G f_n) + \varphi'(\mathbb{E}_G f_n)(f_n - \mathbb{E}_G f_n) \leq \varphi(f_n).
\]

Under the given hypothesis, \( \varphi'(\mathbb{E}_G f_n) \) is bounded and \( G \) measurable and therefore conditioning the previous equation on \( G \) implies,

\[
\varphi(\mathbb{E}_G f_n) = \mathbb{E}_G [\varphi(\mathbb{E}_G f_n) + \varphi'(\mathbb{E}_G f_n)(f_n - \mathbb{E}_G f_n)] = \mathbb{E}_G [\varphi(f_n)].
\]

Using cMCT we may let \( n \uparrow \infty \) in order to conclude

\[
\varphi(\mathbb{E}_G f) = \varphi \left( \lim_{n \to \infty} \mathbb{E}_G f_n \right) = \lim_{n \to \infty} \varphi(\mathbb{E}_G f_n) \leq \lim_{n \to \infty} \mathbb{E}_G [\varphi(f_n)] = \mathbb{E}_G \left[ \lim_{n \to \infty} \varphi(f_n) \right] = \mathbb{E}_G [\varphi(f)].
\]

**Corollary 17.33.** The conditional expectation operator, \( \mathbb{E}_G \), maps \( L^p(\Omega, \mathcal{B}, P) \) into \( L^p(\Omega, \mathcal{B}, P) \) and the map remains a contraction for all \( 1 \leq p \leq \infty \).

**Proof.** The case \( p = \infty \) and \( p = 1 \) have already been covered in Theorem 17.5. So now suppose, \( 1 < p < \infty \), and apply Jensen’s inequality with \( \varphi(x) = |x|^p \) to find \( |\mathbb{E}_G f|^p \leq \mathbb{E}_G |f|^p \) a.s. Taking expectations of this inequality gives the desired result.

**Exercise 17.9 (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, \( M_n \), of \( \bigcup_{n=1}^\infty L^2(\Omega, \mathcal{B}_n, P) \) is \( L^2(\Omega, \mathcal{B}_\infty, P) \) where \( \mathcal{B}_\infty = \bigcup_{n=1}^\infty \mathcal{B}_n \). **Hint:** make use of Theorem 15.29.
2. For every \( X \in L^2(\Omega, \mathcal{B}, P) \), \( X_n := P[X|\mathcal{B}_n], X \to P[X|\mathcal{B}_\infty] \) in \( L^2(P) \). **Hint:** see Exercise 16.5.
3. For every \( X \in L^1(\Omega, \mathcal{B}, P) \), \( X_n := E[X|\mathcal{B}_n], X \to E[X|\mathcal{B}_\infty] \) 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 \( X_n \to E[X|\mathcal{B}_\infty] \) a.s. as well.)

**Exercise 17.10 (Martingale Convergence Theorem for general \( p \)).** Let \( 1 \leq p < \infty \), \((\Omega, \mathcal{B}, P)\) 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 \( X \in L^p(\Omega, \mathcal{B}, P) \), \( X_n := E[X|\mathcal{B}_n], X \to E[X|\mathcal{B}_\infty] \) in \( L^p(P) \). **(Hint:** show that \( \{E[X|\mathcal{B}_n]|P\}_{n=1}^\infty \) is uniformly integrable and \( E[X|\mathcal{B}_\infty] \to E[X|\mathcal{B}_\infty] \) with the aid of item 3 of Exercise 17.9.)

**Theorem 17.34.** Let \((\Omega, \mathcal{B}, \mu)\) be a probability space and assume that \( \mathcal{B} \) is countably generated. Then for all \( 1 \leq p \leq \infty \), \( L^p(\mu)^* \cong L^q(\mu) \) where \( q = p/(p-1) \).

**Proof.** The case \( 1 \leq p \leq 2 \) 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

\[
\varphi(f) = \int \varphi g \, d\mu \quad \text{for all } f \in L^p(\mu).
\]

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 Hilbert space 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

\[
\varphi_n(f) = \int f g_n \, d\mu
\]

\[\footnote{This theorem is true without this assumption or the finiteness assumption on \( \mu \), but we do not prove this here.}\]
\[ \varphi(f) = \varphi_n(f) = \int_{\Omega} f g_n \, d\mu \text{ for all } f \in L^2(\Omega, B_\mu) = L^p(\Omega, B_\mu, \mu). \]

Now taking \( f = \text{sgn}(g_n) |g_n|^{p-1} \) in the above identity shows
\[
\|g_n\|_q^p = \int_{\Omega} |g_n|^q \, d\mu = \varphi(f) \leq \|\varphi\|_{L^p(\Omega, B_\mu)} \cdot \|f\|_p
\]
where
\[
\|f\|_p^p = \int_{\Omega} |f|^{p} \, d\mu = \int_{\Omega} |g_n|^{p} \, d\mu = \|g_n\|_q^q.
\]
Therefore it follows that
\[
\|g_n\|_q^q \leq \|\varphi\|_{L^p(\Omega, B_\mu)} \cdot \left(\|g_n\|_q^q\right)^{1/p} \implies \|\varphi\|_{L^p(\Omega, B_\mu)} \geq \left(\|g_n\|_q^q\right)^{1-\frac{1}{p}} = \|g_n\|_q.
\]

Moreover it is easy to check using \( \varphi(f) = \int_{\Omega} f g_n \, d\mu \) that \( \{g_n\}_{n=1}^\infty \) is a \( (B_n)_{n=1}^\infty \)-martingale. Therefore by the martingale convergence theorem, \( g := L^\infty(\mu)\)-lim\( n \to \infty \) \( g_n \) exists and we have \( g_n = \mathbb{E}[g|B_n] \) for all \( n \). That is \( \varphi(f) = \int_{\Omega} f g \, d\mu \) for all \( f \in L^p(\Omega, B_\mu) \) for all \( n \). Using the martingale convergence theorem again; if \( f \in L^p(\mu) \), then
\[
\varphi(f) = \lim_{n \to \infty} \varphi(\mathbb{E}[f|B_n]) = \lim_{n \to \infty} \int_{\Omega} \mathbb{E}[f|B_n] \, g \, d\mu
\]
\[
= \lim_{n \to \infty} \int_{\Omega} f \mathbb{E}[g|B_n] \, d\mu = \lim_{n \to \infty} \int_{\Omega} f g \, d\mu.
\]
This completes the existence proof.

We could also use the multiplicative system theorem to finish off the proof since \( M := \bigcup_n L^p(\Omega, B_\mu) \) is a multiplicative system of bounded functions 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 \( f \in L^p(\mu) \).
17.4 Construction of Regular Conditional Distributions*

Lemma 17.37. Suppose that \( h : Q \rightarrow [0, 1] \) is an increasing (i.e. non-decreasing) function and \( H(t) := \inf \{ h(s) : t < s \in Q \} \) for all \( t \in \mathbb{R} \). Then \( H : \mathbb{R} \rightarrow [0, 1] \) is an increasing right continuous function.

**Proof.** If \( t_1 < t_2 \), then
\[
\{ h(s) : t_1 < s \in Q \} \subset \{ h(s) : t_2 < s \in Q \}
\]
and therefore \( H(t_1) \leq H(t_2) \). Let \( H(t^+) := \lim_{t \downarrow H(t)} H(t) \). Then for any \( s \in Q \) with \( s > t \) we have \( H(t) \leq H(t^+) \leq h(s) \) and then taking the infimum over such \( s \) we learn that \( H(t) \leq H(t^+) \leq H(t) \), i.e. \( H(t^+) = H(t) \).

Lemma 17.38. Suppose that \( (X, \mathcal{M}) \) is a measurable space and \( F : X \times \mathbb{R} \rightarrow \mathbb{R} \) is a function such that: 1) \( F(\cdot, t) : X \rightarrow \mathbb{R} \) is \( \mathcal{M}/\mathcal{B}_\mathbb{R} \) measurable for all \( t \in \mathbb{R} \) and 2) \( F(x, \cdot) : \mathbb{R} \rightarrow \mathbb{R} \) is right continuous for all \( x \in X \). Then \( F \) is \( \mathcal{M} \otimes \mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R} \) measurable.

**Proof.** For \( n \in \mathbb{N} \), the function,
\[
F_n(x, t) := \sum_{k=-\infty}^{\infty} F(x, (k+1)2^{-n}) \mathbb{1}_{[k2^{-n},(k+1)2^{-n})}(t),
\]
is \( \mathcal{M} \otimes \mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R} \) measurable. Using the right continuity assumption, it follows that \( F(x, t) = \lim_{n \rightarrow \infty} F_n(x, t) \) for all \( x, t \in X \times \mathbb{R} \) and therefore \( F \) is also \( \mathcal{M} \otimes \mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R} \) measurable.

Proposition 17.39. 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 12.11 we know that \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) is measure theoretically isomorphic to \( (\{0, 1\}^\mathbb{N}, \mathcal{F}) \) where \( \mathcal{F} \) is the product \( \sigma \)-algebra. As we saw in Section 6.4 \( \mathcal{F} \) is generated by the countable algebra, \( A := \bigcup_{n=1}^{\infty} A_n \) where
\[
A_n := \{ B \times \Omega : B \subset \{0, 1\}^n \} \text{ for all } n \in \mathbb{N}.
\]
According to the baby Kolmogorov Theorem 6.23 any finitely additive probability measure on \( A \) has a unique extension to a probability measure on \( \mathcal{F} \). The algebra \( A \) may now be transferred by the measure theoretic isomorphism to the desired sub-algebra, \( \mathcal{A}_\mathbb{R} \), of \( \mathcal{B}_\mathbb{R} \).

Theorem 17.40. Suppose that \( (X, \mathcal{M}) \) is a measurable space, \( X : \Omega \rightarrow X \) is a measurable function and \( Y : \Omega \rightarrow \mathbb{R} \) is a random variable. Then there exists a probability kernel, \( Q \), on \( X \times \mathbb{R} \) such that \( \mathbb{E}[f(Y) \mid X] = Q(X, f), P \)-a.s. for all bounded measurable functions, \( f : \mathbb{R} \rightarrow \mathbb{R} \).

**Proof.** First proof. For each \( r \in \mathbb{Q} \), let \( q_r : X \rightarrow [0, 1] \) be a measurable function such that
\[
\mathbb{E}[1_{Y \leq r} \mid X] = q_r(X) \text{ a.s.}
\]
Let \( \nu := P \circ X^{-1} \) be the law of \( X \). Then using the basic properties of conditional expectation, \( q_r \leq q_s, \nu \)-a.s. for all \( r \leq s \), \( \lim_{r \downarrow \infty} q_r = 1 \) and \( \lim_{r \uparrow \infty} q_r = 0, \nu \)-a.s. Hence the set, \( \mathcal{X}_0 \subset \mathbb{X} \) where \( q_r(x) \leq q_s(x) \) for all \( r \leq s \), \( \lim_{r \downarrow \infty} q_r(x) = 1 \), and \( \lim_{r \uparrow \infty} q_r(x) = 0 \) satisfies, \( \nu(\mathcal{X}_0) = P(X \in \mathcal{X}_0) = 1 \). For \( t \in \mathbb{R} \), let
\[
F(x, t) := 1_{\mathcal{X}_0}(x) \cdot \inf \{ q_r(x) : r > t \} + 1_{\mathcal{X} \setminus \mathcal{X}_0}(x) \cdot 1_{t \geq 0}.
\]
Then \( F(\cdot, t) : X \rightarrow \mathbb{R} \) is measurable for each \( t \in \mathbb{R} \) and by Lemma 17.37, \( F(x, \cdot) \) is a distribution function on \( \mathbb{R} \) for each \( x \in X \). Hence an application of Lemma 17.38 shows \( F : X \times \mathbb{R} \rightarrow [0, 1] \) is measurable.

For each \( x \in X \) 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, \( F : \mathbb{R} \rightarrow [0, 1] \).

We will now show that \( Q \) is the desired probability kernel. To prove this, let \( \mathbb{H} \) be the collection of bounded measurable functions, \( f : \mathbb{R} \rightarrow \mathbb{R} \), such that \( X \ni x \rightarrow Q(x, f) \in \mathbb{R} \) is measurable and \( \mathbb{E}[f(Y) \mid X] = Q(X, f) \), \( P \)-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, \( \mathcal{M} = \{ 1_{(-\infty, t]} : t \in \mathbb{R} \} \) so that multiplicative systems Theorem 11.2 may be applied.

Notice that \( Q(x, 1_{(-\infty, t]}) = F(x, t) \) is measurable. Now let \( r \in \mathbb{Q} \) and \( g : X \rightarrow \mathbb{R} \) be a bounded measurable function, then
\[
\mathbb{E}[1_{Y \leq r} \cdot g(X)] = \mathbb{E}[\mathbb{E}[1_{Y \leq r} \mid X] g(X)] = \mathbb{E}[q_r(X) g(X)] = \mathbb{E}[q_r(X) 1_{\mathcal{X}_0}(X) g(X)].
\]
For \( t \in \mathbb{R} \), we may let \( r \downarrow t \) in the above equality (use DCT) to learn,
\[
\mathbb{E}[1_{Y \leq t} \cdot g(X)] = \mathbb{E}[F(X, t) 1_{\mathcal{X}_0}(X) g(X)] = \mathbb{E}[F(X, t) g(X)].
\]
Since \( g \) was arbitrary, we may conclude that
\[
Q(X, 1_{(-\infty, t]}) = F(X, t) = \mathbb{E}[1_{Y \leq t}] \text{ a.s.}
\]
This completes the proof.

Second proof. Let \( A := A_2 \) be the algebra described in Proposition 17.39. For each \( A \in A \), let \( \mu_A : X \rightarrow \mathbb{R} \) be a measurable function such that \( \mu_A(X) = P(Y \in A | X) \) a.s. If \( A = A_1 \cup A_2 \) with \( A_1 \in A \) and \( A_1 \cap A_2 = \emptyset \), then
\[ \mu_{A_1}(X) + \mu_{A_2}(X) = \mathbb{P}(Y \in A_1 | X) + \mathbb{P}(Y \in A_2 | X) = \mu_{A_1 + A_2}(X) \text{ a.s.} \]

Therefore we have \( \mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1 + A_2}(x) \) for \( \nu \)-a.e. \( x \).

Since \( \mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1 + A_2}(x) \) for \( \nu \)-a.e. \( x \), we know that \( \mu_{\mathbb{R}}(x) = 1 \) for \( \nu \)-a.e. \( x \).

Thus if \( \nu := \text{Law}_P(X) \), we have \( \mu_{A_1}(x) + \mu_{A_2}(x) = \mu_{A_1 + A_2}(x) \) for \( \nu \)-a.e. \( x \).

Hence, for all bounded measurable functions, \( f \) on \( \mathbb{R} \), we have \( f \) measurable and \( \mathbb{E}(|f(Y)|X) = \mathbb{E}(f(Y)|X) \) for all disjoint \( x \in X \).

This result leads fairly immediately to the following far reaching generalization.

**Theorem 17.41.** Suppose that \( (\mathbb{X}, \mathcal{M}) \) is a measurable space and \( (\mathbb{Y}, \mathcal{N}) \) is a standard Borel space.\(^4\) Suppose that \( X : \Omega \to \mathbb{X} \) and \( Y : \Omega \to \mathbb{Y} \) are measurable functions. Then there exists a probability kernel, \( Q \), on \( (\Omega, \mathcal{G}) \times (\mathbb{Y}, \mathcal{N}) \) such that \( \mathbb{E}[f(Y) | G] = Q(\cdot, f), \) \( P \)-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 \( \mathbb{Y} \in \mathcal{B}_{\mathbb{Y}} \) and \( \mathcal{N} = \mathcal{B}_{\mathbb{Y}} \). In this case \( \mathbb{Y} \) may also be viewed to be a measurable map form \( \Omega \) to \( \mathbb{Y} \) such that \( Y(\Omega) \subset \mathbb{Y} \). By Theorem 17.40 we may find a probability kernel, \( Q_0 \), on \( \mathbb{X} \times \mathbb{Y} \) such that

\[ \mathbb{E}[f(Y) | X] = Q_0(X, f), \quad P \text{-a.s.} \]  

(17.53)

for all bounded measurable functions, \( f : \mathbb{R} \to \mathbb{R} \).

Taking \( f = 1_Y \) in Eq. (17.53) shows

\[ 1 = \mathbb{E}[1_Y(Y) | X] = Q_0(X, Y) \text{ a.s.} \]

Thus if we let \( \mathbb{X}_0 := \{ x \in \mathbb{X} : \mathbb{Q}(x, y) = 1 \} \), we know that \( P(X \in \mathbb{X}_0) = 1 \). Let us now define

\[ Q(x, B) := 1_{\mathbb{X}_0}(x) \mathbb{Q}(x, B) + 1_{\mathbb{X} \setminus \mathbb{X}_0}(x) \delta_B(B) \text{ for } (x, B) \in \mathbb{X} \times \mathcal{B}_{\mathbb{Y}}, \]

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 \( B \in \mathcal{B}_{\mathbb{Y}} \subset \mathcal{B}_\mathbb{Y} \), then

\[ Q(x, B) = 1_{\mathbb{X}_0}(x) \mathbb{Q}(x, B) = 1_{\mathbb{X}_0}(x) \mathbb{E}[1_B(Y) | X] = \mathbb{E}[1_B(Y) | X] \text{ a.s.} \]

This shows that \( Q \) is the desired regular conditional probability.

**Corollary 17.42.** Suppose \( \mathcal{G} \) is a sub-\( \mathcal{\sigma} \)-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 17.41 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 17.43.** 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-\( \mathcal{\sigma} \)-algebra \( \mathcal{B} \). Then there exists a probability kernel, \( Q \) on \( (\Omega, \mathcal{G}) \times (\mathbb{Y}, \mathcal{N}) \) 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 17.42 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 17.44.** It turns out that every standard Borel space \( (\mathbb{X}, \mathcal{M}) \) possesses a countable sub-\( \mathcal{\sigma} \)-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 17.40 extends easily to give another proof of Theorem 17.41 all in one go. As the next example shows it is a bit tricky to produce the algebra \( \mathcal{A} \).

**Example 17.45.** Let \( \Omega := \{0, 1\}^\mathbb{N}, \pi_i : \Omega \to \{0, 1\} \) be projection onto the \( i \)-th component and \( \mathcal{B} := \sigma(\pi_1, \pi_2, \ldots) \) be the product \( \mathcal{\sigma} \)-algebra on \( \Omega \). Further let \( \mathcal{A} := \bigcup_{n=1}^\infty \mathcal{A}_n \) where

\[ \mathcal{A}_n := \{ B \times \Omega : B \subset \{0, 1\}^n \} \text{ for all } n \in \mathbb{N}. \]

Suppose that \( X = \{e_n\}_{n=1}^\infty \subset \mathbb{R} \) where \( e_n(i) = \delta_{i,n} \) 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 \( \mathcal{\sigma} \)-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,

\[
P_0 (A \cup B) = 1 - \mu ([A \cup B]^c) = 1 - \mu (A^c \cap B^c) = 1 - \mu (B^c \setminus A) = 1 - (\mu (B^c) - \mu (A)) = 1 - \mu (B^c) + \mu (A) = P_0 (B) + P_0 (A).
\]

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,

\[
1 = P_0 (X) = P (X) = \sum_{n=1}^{\infty} P \left( \{ e_n \} \right) = \sum_{n=1}^{\infty} P_0 \left( \{ e_n \} \right) = \sum_{n=1}^{\infty} \mu (\{ e_n \}) = Z < 1
\]

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 \( A \subset \{ 2, 3, \ldots \} \) for \( 1 \in A \) and \( \# (A^c) < \infty \). Thus if \( A_k \in \mathcal{A} \) with \( A_k \downarrow \emptyset \) we must have that \( 1 \not\in A_k \) 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 \( \sigma (\mathcal{A}) = 2^\mathbb{N} \).
The Radon-Nikodym Theorem

**Theorem 18.1 (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 9.39 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, \(L^2(\lambda) \supseteq f \mapsto \nu(f) \in \mathbb{C}\) is a continuous linear functional on \(L^2(\lambda)\). By the Riesz representation Theorem 16.16 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 
\[\nu(f) = \text{Re} \nu(f) = \text{Re} \int_X f \rho d\lambda = \int_X f \rho d\lambda.\] 
(18.1)

Thus by replacing \(\rho\) by \(\text{Re} \rho\) if necessary we may assume \(\rho\) is real.

Taking \(f = 1_{\rho < 0}\) in Eq. (18.1) 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, \text{ \(\lambda\) – a.e., i.e. } \lambda(\rho < 0) = 0.\]
Therefore \(\rho \geq 0, \text{ \(\lambda\) – 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, \text{ \(\lambda\) – a.e.}\]

---

**Definition 18.2.** 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 
\[\nu(A) = 0 \text{ and } \mu(A^c) = 0.\]
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 18.3 (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 18.4.** 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 
\[\nu(A_i) = 0 \text{ and } \mu_j(A_i^c) = 0 \text{ for all } i.\]
Letting \(A := \cup_i A_i\) we have \(\nu(A) = 0.\) Moreover, since \(A^c = \cap_i A_i^c \subseteq A_m^c\) 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 18.5.** 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 $\nu_s \ll \mu$, we also know that $\nu_s(A) = 0$. So for $C \in \mathcal{M}$,

$$\nu(C \cap A) = \nu_s(C \cap A) + \nu(A \cap A) = \nu_s(C) = (18.2)$$

and

$$\nu(C \cap A') = \nu_s(C \cap A') + \nu(A \cap A') = \nu_s(C \cap A') = \nu_s(C'). (18.3)$$

Now suppose we have another Lebesgue decomposition, $\nu = \tilde{\nu}_a + \tilde{\nu}_s$ with $\tilde{\nu}_s \perp \mu$ and $\tilde{\nu}_s \ll \mu$. Working as above, we may choose $\tilde{A} \in \mathcal{M}$ such that $\mu(\tilde{A}) = 0$ and $\tilde{A}^c$ is $\tilde{\nu}_s$ zero measure. Then $B = A \cap \tilde{A}$ is still a $\mu$-null set and $B' = A^c \cap \tilde{A}^c$ is a null set for both $\nu_s$ and $\tilde{\nu}_s$. Therefore we may use Eqs. (18.2) and (18.3) with $A$ being replaced by $B$ to conclude,

$$\nu_s(C) = \nu(C \cap B) = \tilde{\nu}_s(C)$$

and

$$\nu_s(C') = \nu(C \cap B') = \tilde{\nu}_s(C')$$

for all $C \in \mathcal{M}$.

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 $\infty > \nu(X_n) = \nu_s(X_n) + \nu_s(X_n)$, we must have $\nu_s(X_n) < \infty$ and $\nu_s(X_n) < \infty$, showing $\nu_s$ and $\nu_s$ are $\sigma$ finite as well.

Lemma 18.6. Suppose $\mu$ is a positive measure on $(X, \mathcal{M})$ and $f, g : X \to [0, \infty]$ are functions such that the measures, $\int_A f d\mu$ and $\int_A g d\mu$ are $\sigma$ finite and further satisfy,

$$\int_A f d\mu = \int_A g d\mu$$

for all $A \in \mathcal{M}$. Then $f(x) = g(x)$ for $\mu$ - a.e. $x$. (BRUCE: this lemma is very closely related to Lemma (10.22) above.)

Proof. By assumption there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\int_A f d\mu < \infty$ and $\int_A g d\mu < \infty$ for all $n$. Replacing $A$ by $A \cap X_n$ in Eq. (18.4) implies

$$\int_A f d\mu = \int_{A \cap X_n} f d\mu = \int_{A \cap X_n} g d\mu = \int_A 1_{X_n} g d\mu$$

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 18.7. Lemma (18.6) is in general false without the $\sigma$ finite assumption. A trivial counterexample is to take $\mathcal{M} = \mathcal{P}(X)$, $\mu(A) = \infty$ for all non-empty $A \in \mathcal{M}$, $f = 1_X$ and $g = 2 \cdot 1_X$. Then Eq. (18.4) holds yet $f \neq g$.

Theorem 18.8 (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 $\nu = \nu_s + \nu_s$ relative to $\mu$ and there exists a unique (modulo sets of $\mu$ – measure zero) function $\rho : X \to [0, \infty)$ such that $d\nu_s = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$.

Proof. The uniqueness assertions follow directly from Lemmas (18.5) and (18.6)

Existence when $\mu$ and $\nu$ are both finite measures. (Von-Neumann’s Proof. See Remark (18.9) for the motivation for this proof.) First suppose that $\mu$ and $\nu$ are finite measures and let $\lambda = \mu + \nu$. By Theorem (18.1) $d\nu = h d\lambda$ with $0 < h \leq 1$ and this implies, for all non-negative measurable functions $f$, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh)$$

or equivalently

$$\nu(f(1 - h)) = \mu(fh).$$

Taking $f = 1_{\{h = 1\}}$ in Eq. (18.6) shows that

$$\mu(\{h = 1\}) = \nu(1_{\{h = 1\}}(1 - h)) = 0,$$

i.e. $0 \leq h(x) < 1$ for $\mu$ - a.e. $x$. Let

$$\rho := \frac{1_{\{h < 1\}}}{1 - h}$$

and then take $f = g1_{\{h < 1\}}(1 - h)^{-1}$ with $g \geq 0$ in Eq. (18.6) to learn

$$\nu(g1_{\{h < 1\}}) = \mu(g1_{\{h < 1\}}(1 - h)^{-1}h) = \mu(\rho g).$$

Hence if we define

$$\nu_s := 1_{\{h < 1\}} \nu$$

we then have $\nu_s \perp \mu$ (since $\nu_s$ “lives” on $\{h = 1\}$ while $\mu(h = 1) = 0$) and $\nu_s = \rho \mu$ and in particular $\nu_s \ll \mu$. Hence $\nu = \nu_s + \nu_s$ 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 $\nu = \nu_s = \rho \mu$.

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 $\nu_n^\sigma$ such that $d\nu_n = \rho_n d\mu + d\nu_n^\sigma$ with $\nu_n^\sigma \perp \mu_n$. Since $\mu_n$ and $\nu_n^\sigma$ “live” on $X_n$ there exists $A_n \in \mathcal{M}_{X_n}$ such that $\mu(A_n) = \mu_n(A_n) = 0$ and
\[ \nu_n^a (X \setminus A_n) = \nu_n^s (X_n \setminus A_n) = 0. \]

This shows that \( \nu_n^a \perp \mu \) for all \( n \) and so by Lemma 18.4 \( \nu_s := \sum_{n=1}^{\infty} \nu_n^a \) is singular relative to \( \mu \). Since

\[ \nu = \sum_{n=1}^{\infty} \nu_n = \sum_{n=1}^{\infty} (\rho_n \nu_n + \nu_n^a) = \sum_{n=1}^{\infty} (\rho_n 1X_n \mu + \nu_n^a) = \rho \mu + \nu_s, \quad (18.7) \]

where \( \rho := \sum_{n=1}^{\infty} 1X_n \rho_n \), 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 18.9. Here is the motivation for the above construction. Suppose that \( \nu \) is the desired Lebesgue decomposition of \( \nu \).

\[ \text{d} \nu = \nu_s \text{d} \mu + \rho \mu \text{d} \nu_s \]

where \( \rho := \sum_{n=1}^{\infty} 1X_n \rho_n \), it follows that \( \nu = \nu_s + \nu_a \) with \( \nu_a = \rho \mu \). Hence this is the desired Lebesgue decomposition of \( \nu \).

1. If \( f : T (\Omega) \to \mathbb{R} \) and \( g : \Omega \to \mathbb{R} \) we have

\[ \int_{\Omega} f \circ T \text{d} m = \int_{T(\Omega)} f \text{d} (m \circ T^{-1}) \quad (18.8) \]

and

\[ \int_{T(\Omega)} g \circ T^{-1} \text{d} m = \int_{\Omega} g \text{d} (m \circ T) . \quad (18.9) \]

2. Referring to the math 140 notes, show \( |m \circ T^{-1} (A)| \leq K m (A) \) and similarly \( |m \circ T (A)| \leq K m (\Omega) \).

3. There is a relationship between \( \alpha \) and \( \beta \). Indeed taking \( g = (f \beta) \circ T \) shows

\[ \int_{\Omega} f \circ T \text{d} m = \int_{T(\Omega)} f \beta \text{d} m = \int_{T(\Omega)} ((f \beta) \circ T) \circ T^{-1} \text{d} m \]

\[ = \int_{\Omega} (f \circ T) \cdot (\beta \circ T) \text{d} m \]

from which we conclude

\[ (\beta \circ T) \alpha = 1 \text{ a.e. } \iff \alpha = \frac{1}{\beta \circ T}. \quad (18.12) \]

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 \( \delta_r \ast \alpha \to \alpha \) in \( L^1_{\text{loc}} \) 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

\[ T' (x)^{-1} [T (y) - T (x)] = T' (x)^{-1} \left[ \int_0^1 T' (x + t (y - x)) (y - x) \text{d} t \right] \]

\[ = \left[ \int_0^1 T' (x)^{-1} T' (x + t (y - x)) (y - x) \text{d} t \right] \]

\[ = y - x + \varepsilon (x, y) (y - x) \]
where 

\[ \varepsilon(x, y) := \int_0^1 \left[ T'(x)^{-1} T'(x + t(y - x)) - I \right] \, dt \]

satisfies,

\[ \| \varepsilon(x, y) \| \leq \int_0^1 \left\| T'(x)^{-1} T'(x + t(y - x)) - I \right\| \, dt = \varepsilon(\|y - x\|). \]

From this it follows if \( y \in B_r(x) = x + B_r(0) \), then

\[ \left\| T'(x)^{-1} [T(y) - T(x)] \right\| \leq r \left( 1 + \varepsilon(r) \right) \]

and hence

\[ T'(x)^{-1} [T(B_r(x)) - T(x)] \subset B_r(1 + \varepsilon(r)) \]

or equivalently that

\[ T(B_r(x)) \subset T(x) + T'(x) B_r(1 + \varepsilon(r))(0) \]. \hspace{1cm} (18.13)

5. Therefore it follows that

\[ \frac{m(T(B_r(x)))}{m(B_r(x))} \leq \frac{m(T(x) + T'(x) B_r(1 + \varepsilon(r))(0))}{m(B_r(x))} = |\det T'(x)| \cdot (1 + \varepsilon(r))^d. \]

Letting \( r \downarrow 0 \) and using the Lebesgue differentiation theorem (see Theorem 5.5) which is a rather deep result!

\[ \alpha(x) = \frac{d(m \circ T)}{dm}(x) \leq |\det T'(x)| \text{ for a.e. } x \in \Omega. \hspace{1cm} (18.14) \]

Applying this result with \( T \) replaced by \( T^{-1} \) then shows,

\[ \beta(y) = \frac{d(m \circ T^{-1})}{dm}(y) \leq |\det T^{-1}(y)| \text{ for a.e. } y \in T(\Omega). \hspace{1cm} (18.15) \]

Alternatively let us observe that

\[ \frac{m(T(B_r(x)))}{m(B_r(x))} = \frac{1}{m(B_r(x))} \int_{B_r(0)} \alpha(x + y) \, dy = \delta_r \ast \alpha(x). \]

By the easier approximate identity Theorem 5.5 we know \( \delta_r \ast \alpha \rightarrow \alpha \) in \( L^1_{loc} \) and so there exists \( r_n \downarrow 0 \) such that \( \delta_{r_n} \ast \alpha \rightarrow \alpha \) a.e. as \( n \rightarrow \infty \). Thus we again learn that for a.e. \( x \),

\[ \alpha(x) = \lim_{n \rightarrow \infty} \delta_{r_n} \ast \alpha(x) = \lim_{n \rightarrow \infty} \frac{m(T(B_{r_n}(x)))}{m(B_{r_n}(x))} \leq \lim_{n \rightarrow \infty} |\det T'(x)| \cdot (1 + \varepsilon(r_n))^d = |\det T'(x)|. \]

6. We now use \( \alpha = \frac{1}{\beta \circ T} \) from Eq. (18.12) along with the inequality in Eq. (18.15) to learn

\[ \alpha(x) = \frac{1}{\beta \circ T(x)} \geq \frac{1}{|\det (T^{-1})'(T(x))|}. \]

On the other hand, since \( T^{-1} \circ T = I \), it follows by the chain rule that \( (T^{-1})'(T(x)) T'(x) = I \) and therefore

\[ \frac{1}{|\det (T^{-1})'(T(x))|} = |\det T'(x)| \]

and we may conclude \( \alpha(x) \geq |\det T'(x)| \). This result along with the inequality in Eq. (18.14) shows

\[ \alpha(x) = |\det T'(x)|. \]

7. Using this result back in Eq. (18.11) with \( g = f \circ T \) for some function on \( f : T(\Omega) \rightarrow \mathbb{R} \) 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 5.5, we could have proved the stronger version of Eq. (18.13):

\[ T(x) + T'(x) B_r(1 - \varepsilon(r))(0) \subset T(B_r(x)) \subset T(x) + T'(x) B_r(1 + \varepsilon(r))(0). \]

If we had done this we could have avoided discussing \( \beta \) altogether. Indeed, the method of step 5. would then give

\[ |\det T'(x)| \cdot (1 - \varepsilon(r))^d \leq \frac{m(T(B_r(x)))}{m(B_r(x))} \leq |\det T'(x)| \cdot (1 + \varepsilon(r))^d \]

which upon letting \( r \downarrow 0 \) would have shown \( \frac{d(m \circ T)}{dm}(x) = |\det T'(x)|. \) \hspace{1cm} \( \square \)
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 19.1 (Von-Neumann’s Mean Ergodic Theorem). Let $U : H \to H$ be an isometry on a Hilbert space $H$, $M = \text{Null}(U - I)$, $P = P_M$ be orthogonal projection onto $M$, and $S_n = \sum_{k=0}^{n-1} U^k$. Show $S_n \to P_M$ 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, $UU^* = P_{\text{Ran}(U)}$. Thus $UU^* = I$ iff $U$ is surjective, i.e. $U$ is unitary.

Before starting the proof in earnest we need to prove

$$\text{Null}(U^* - I) = \text{Null}(U - I).$$

If $x \in \text{Null}(U - I)$ then $x = Ux$ and therefore $U^* x = U^* U x = x$, i.e. $x \in \text{Null}(U^* - I)$. Conversely if $x \in \text{Null}(U^* - I)$ then $U^* x = x$ and we have

$$\|Ux - x\|^2 = 2\|x\|^2 - 2 \text{Re} (Ux, x)$$

$$= 2\|x\|^2 - 2 \text{Re} (x, U^* x) = 2\|x\|^2 - 2 \text{Re} (x, x) = 0$$

which shows that $Ux = x$, i.e. $x \in \text{Null}(U - I)$. 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 $x = (U - I)y \in \text{Ran}(U - I)$, we have

$$\frac{S_n}{n} x = \frac{1}{n} (U^n y - y) \to 0 \text{ as } n \to \infty.$$

More generally if $x \in \text{Ran}(U - I)$ and $x' \in \text{Ran}(U - I)$, we have, since $\|\frac{S_n}{n}\| \leq 1$, that

$$\left\| \frac{S_n}{n} x - \frac{S_n}{n} x' \right\| \leq \|x - x'\|$$

and hence

$$\lim \sup_{n \to \infty} \left\| \frac{S_n}{n} x \right\| = \lim \sup_{n \to \infty} \left\| \frac{S_n}{n} x - \frac{S_n}{n} x' \right\| \leq \|x - x'\|.$$

Letting $x' \in \text{Ran}(U - I)$ tend to $x \in \text{Ran}(U - I)$ allows us to conclude that

$$\lim \sup_{n \to \infty} \|\frac{S_n}{n} x\| = 0.$$

For $x \in \text{Ran}(U - I) = \text{Ran}(U - I)^\perp = \text{Null}(U^* - I) = \text{Null}(U - I) = M$ we have $\frac{S_n}{n} x = x$. So for general $x \in H$, we have $x = P_M x + y$ with $y \in M^\perp = \text{Ran}(U - I)$ 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 19.6 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 19.2. Let

$$\mathcal{B}_\theta := \{ A \in \mathcal{B} : \theta^{-1}(A) = A \} \quad \text{and} \quad \mathcal{B}_\theta' := \{ A \in \mathcal{B} : \mu (\theta^{-1}(A) \Delta A) = 0 \}$$

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. $A \Delta C \subset [A \Delta B] \cup [B \Delta C]$,
2. $[\bigcup_{n=1}^\infty A_n] \Delta [\bigcup_{n=1}^\infty B_n] \subset \bigcup_{n=1}^\infty A_n \Delta B_n$,
3. $[\bigcap_{n=1}^\infty A_n] \Delta [\bigcap_{n=1}^\infty B_n] \subset \bigcup_{n=1}^\infty \Delta B_n$,
4. $B \Delta \{A_n \ i.o.\} \subset \bigcup_{n=1}^\infty [B \Delta A_n]$. 
Lemma 19.3. The elements of $B'_g$ are the same as the elements in $B_g$ modulo null sets, i.e.

$$B'_g = \{ B \in B : \exists A \in B_g \ni \mu (A \triangle B) = 0 \}.$$  

Moreover if $B \in B'_g$, then 

$$A := \{ \omega \in \Omega : \theta^k (\omega) \in B \text{ i.o. } k \} \in B_g$$

and $\mu (A \triangle B) = 0$. (We could have just as well taken $A$ to be equal to \{ $\omega \in \Omega : \theta^k (\omega) \in B$ a.a. \}.)

Proof. If $A \in B_g$ and $B \in B$ such that $\mu (A \triangle B) = 0$, then

$$\mu (A \triangle \theta^{-1} (B)) = \mu (\theta^{-1} (A) \triangle \theta^{-1} (B)) = \mu (A \triangle B) = 0$$

and therefore it follows that

$$\mu (B \triangle \theta^{-1} (B)) \leq \mu (B \triangle A) + \mu (A \triangle \theta^{-1} (B)) = 0.$$  

This shows that $B \in B'_g$.

Conversely if $B \in B'_g$ then by the invariance of $\mu$ under $\theta$ it follows that $\mu (\theta^{-1} (B) \triangle \theta^{-(l+1)} (B)) = 0$ for all $k = 0, 1, 2, \ldots$. In particular we learn that

$$\mu (\theta^{-k} (B) \triangle B) = \mu (\{1_{\theta^{-k} (B)} - 1_B\})$$

$$\leq \sum_{l=0}^{k-1} \mu (\{1_{\theta^{-l} (B)} - 1_{\theta^{-(l+1)} (B)}\})$$

$$= \sum_{l=0}^{k-1} \mu (\theta^{-l} (B) \triangle \theta^{-(l+1)} (B)) = 0.$$  

Thus if $A = \{ \theta^{-k} (B) \text{ i.o. } k \}$ as in Eq. (19.1) we have,

$$\mu (B \triangle A) \leq \sum_{k=1}^{\infty} \mu (B \triangle \theta^{-k} (B)) = 0.$$  

This completes the proof since

$$\theta^{-1} (A) = \{ \omega \in \Omega : \theta^{k+1} (\omega) \in B \text{ i.o. } k \} = A$$

and thus $A \in B_g$.

Definition 19.4. A $B$–measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff $f \circ \theta = f$ (f a.e.).

Lemma 19.5. A $B$–measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff $f$ is $B'_g (B'_g)$ measurable. Moreover, if $f$ is almost invariant, then there exists an invariant function, $g : \Omega \to \mathbb{R}$, such that $f = g\mu$ – a.e. (This latter assertion has already been explained in Exercises [15.3 and 15.4].)

Proof. If $f$ is invariant, $f \circ \theta = f$, then $\theta^{-1} (\{ f < x \}) = \{ f \circ \theta < x \} = \{ f < x \}$ which shows that $\{ f < x \} \in B_g$ for all $x \in \mathbb{R}$ and therefore $f$ is $B_g$–measurable. Similarly if $f$ is almost invariant so that $f \circ \theta = f (\mu$–a.e.), then

$$\mu (\{ 1_{\theta^{-1} (f < x)} - 1_{f < x} \}) = \mu (\{ 1_{f \circ \theta < x} - 1_{f < x} \})$$

$$= \mu (\{ 1_{(-\infty, x)} \circ f \circ \theta - 1_{(-\infty, x)} \circ f \}) = 0$$

from which it follows that $\{ f < x \} \in B'_g$ for all $x \in \mathbb{R}$, that is $f$ is $B'_g$–measurable.

Conversely if $f : \Omega \to \mathbb{R}$ is $(B'_g) B_g$–measurable, then for all $-\infty < a < b < \infty$, $(\{ a < f \leq b \} \in B'_g) \{ a < f \leq b \} \in B_g$ from which it follows that $1_{(a < f \leq b)}$ is (almost) invariant. Thus for every $N \in \mathbb{N}$ the function defined by

$$f_N := \sum_{n=-N^2}^{N^2} \frac{n}{N} 1_{\{ \frac{n-1}{N} < f \leq \frac{n}{N} \}},$$

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 $D_n (n) \in B_g$ such that $\mu (D_n (n) \Delta \{ \frac{n-1}{N} < f \leq \frac{n}{N} \}) = 0$ 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 $g_N = f_N \mu$ a.e. and $g_N$ is $B_g$–measurable. We may thus conclude that $\bar{g} := \limsup_{N \to \infty} g_N$ is $B_g$–measurable. It now follows that $g := \bar{g} 1_{|g| < \infty}$ is $B_g$–measurable function such that $g = f$ a.e.

Theorem 19.6. Suppose that $(\Omega, B, \mu)$ 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 $U f := f \circ \theta$ is an isometry. The isometry $U$ is unitary if $\theta^{-1}$ exists as a measurable map.

2. The map

$$L^2 (\Omega, B_g, \mu) \ni f \to f \in \text{Nul} (U - I)$$

is unitary. In other words, $U f = f$ iff there exists $g \in L^2 (\Omega, B_g, \mu)$ such that $f = g$ a.e.
3. For every $f \in L^2(\mu)$ we have,

$$L^2(\mu) = \lim_{n \to \infty} \frac{f + f \circ \theta + \ldots + f \circ \theta^{n-1}}{n} = \mathbb{E}_{\theta^n}[f]$$

where $\mathbb{E}_{\theta^n}$ denotes orthogonal projection from $L^2(\Omega, \mathcal{B}, \mu)$ onto $L^2(\Omega, \mathcal{B}_0, \mu)$, i.e. $\mathbb{E}_{\theta^n}$ is conditional expectation.

**Proof.** 1. To see that $U$ is an isometry observe that

$$\|Uf\| = \|f \circ \theta\| = \|f\|$$

for all $f \in L^2(\mu)$.

2. $f \in \text{Null}(U-I)$ iff $f \circ \theta = Uf = f$ a.e., i.e. iff $f$ is almost invariant. According to Lemma [19.5], this happens iff there exists a $\mathcal{B}_0$-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 [19.1].

**Assumption 1**

From now on we will assume that $\mu = P$ is a probability measure such that $P\theta^{-1} = \theta$.

**Exercise 19.1.** For every $Z \in L^1(\mu)$, show that $\mathbb{E}[Z \circ \theta | \mathcal{B}_0] = \mathbb{E}[Z | \mathcal{B}_0]$ a.s. More generally, show for sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{B}$, that

$$\mathbb{E}[Z \circ \theta | \mathcal{B}_0] = \mathbb{E}[Z | \mathcal{G}] \circ \theta \ a.s.$$  

**Exercise 19.2.** Let $1 \leq p < \infty$. Following the ideas introduced in Exercises [17.9] and [17.10], show

$$L^p(P) = \lim_{n \to \infty} \frac{f + f \circ \theta + \ldots + f \circ \theta^{n-1}}{n} = \mathbb{E}_{\theta^n}[f]$$

for all $f \in L^p(\Omega, \mathcal{B}, P)$.

(Some of these ideas will again be used in the proof of Theorem [19.9] below.)

**Definition 19.7.** A sequence of random variables $\xi_k = (\xi_{k})_{k=1}^{\infty}$ is a stationary if $(\xi_{2}, \xi_{3}, \ldots) \overset{d}{=} (\xi_{1}, \xi_{2}, \ldots)$.

If we temporarily let

$$\theta(x_1, x_2, x_3, \ldots) = (x_1, x_2, x_3, \ldots)$$

for $(x_1, x_2, x_3, \ldots) \in \mathbb{R}^N$, then the stationarity condition states that $\theta\xi \overset{d}{=} \xi$. Equivalently if $\mu = \text{Law}_P(\xi_1, \xi_2, \ldots)$ on $(\mathbb{R}^N, \mathcal{B}_\mathbb{R}^\otimes N)$, then $\xi = (\xi_{k})_{k=1}^{\infty}$ is stationary if $\mu \circ \theta^{-1} = \mu$. 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 $x \in (x_1, x_2, x_3, \ldots) \in \mathbb{R}^N$ we will let $S_0(x) = 0$,

$$S_n(x) = x_1 + x_2 + \ldots + x_n,$$

and $S_n^* := \max(S_1, S_2, \ldots, S_n)$ for all $n \in \mathbb{N}$.

**Lemma 19.8 (Maximal Ergodic Lemma).** Suppose $\xi = (\xi_k)_{k=1}^{\infty}$ is a stationary sequence and $S_n(\xi) = \xi_1 + \cdots + \xi_n$ as above, then

$$\mathbb{E}\left[\xi_1 : \sup_n S_n(\xi) > 0 \right] \leq 0.$$  (19.3)

**Proof.** In this proof, $\theta$ will be as in Eq. (19.2). If $1 \leq k \leq n$, then

$$S_k(\xi) = \xi_1 + S_{k-1}(\theta\xi) \leq \xi_1 + S_{k-1}(\theta\xi),$$

and therefore, $S_n^*(\xi) \leq \xi_1 + S_n^*(\theta\xi)$. So we may conclude that

$$\mathbb{E}[\xi_1 : S_n^*(\xi) > 0] \leq \mathbb{E}[S_n^*(\xi) - S_n^*(\theta\xi)]_{++} \mathbb{E}[S_n^*(\theta\xi) > 0]$$

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. (19.3).

**Theorem 19.9 (Birkhoff’s Ergodic Theorem).** Suppose that $f \in L^1(\Omega, \mathcal{B}, P)$ or $f \geq 0$ and is $\mathcal{B}$-measurable, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \theta^{k-1} = \mathbb{E}[f | \mathcal{B}_0] \ a.s..$$  (19.4)

Moreover if $f \in L^p(\Omega, \mathcal{B}, P)$ for some $1 \leq p < \infty$ then the convergence in Eq. (19.4) holds in $L^p$ as well.

---

1 In other words if $\xi_{k}$ is stationary, then by lopping off the first random variable on each side of the identity, $\xi_{2}, \xi_{3}, \ldots \overset{d}{=} (\xi_{1}, \xi_{2}, \ldots)$, implies that

$$\xi_{3}, \xi_{4}, \ldots \overset{d}{=} (\xi_{2}, \xi_{3}, \ldots) \overset{d}{=} (\xi_{1}, \xi_{2}, \ldots).$$

Continuing this way inductively shows that stationarity is equivalent to

$$\xi_{n}, \xi_{n+1}, \xi_{n+2}, \ldots \overset{d}{=} (\xi_{1}, \xi_{2}, \ldots)$$

for all $n \in \mathbb{N}$. 
Proof. Let us begin with the general observation that if \( \xi = (\xi_1, \xi_2, \ldots) \) is a sequence of random variables such that \( \xi_i \circ \theta = \xi_{i+1} \) for \( i = 1, 2, \ldots \), then \( \xi \) is stationary. This is because,

\[
(\xi_1, \xi_2, \ldots) \overset{d}{=} (\xi_1, \xi_2, \ldots) \circ \theta = (\xi_1 \circ \theta, \xi_2 \circ \theta, \ldots) = (\xi_2, \xi_3, \ldots).
\]

We will first prove Eq. (19.4) under the assumption that \( f \in L^1(P) \). We now let \( g := \mathbb{E}[f|B_0] \) and \( \xi_k := f \circ \theta^k - g \) for all \( k \in \mathbb{N} \). Since \( g \) is \( B_0 \)-measurable we know that \( g \circ \theta = g \) and therefore,

\[
\xi_k \circ \theta = (f \circ \theta^k - g) \circ \theta = f \circ \theta^k - g = \xi_{k+1}
\]

and therefore \( \xi = (\xi_1, \xi_2, \ldots) \) is stationary. To simplify notation let us write \( S_n \) for \( S_n(\xi) = \xi_1 + \cdots + \xi_n \). To finish the proof we need to show that \( \lim_{n \to \infty} S_n \) is stationary. To simplify notation let us write \( E \) for \( E_0 = E \circ \theta^0 \).

Thus \( g \) is integrable. This can be done as in the second solution to Exercise 15.6 (Resnick § 6.7, #5). Here are the details.

First observe that \( \{\eta_k \} \) is uniformly integrable. Indeed, by stationarity,

\[
\mathbb{E}[|\eta_k|^p : |\eta_k|^p \geq a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \geq a]
\]

and therefore

\[
\sup_k \mathbb{E}[|\eta_k|^p : |\eta_k|^p \geq a] = \mathbb{E}[|\eta_1|^p : |\eta_1|^p \geq a] \overset{DCT}{\to} 0 \text{ as } a \to \infty.
\]

Thus if \( \varepsilon > 0 \) is given we may find \( \delta > 0 \) such that \( \mathbb{E}[|\eta_k|^p : A] \leq \varepsilon \) whenever \( A \in B \) with \( P(A) \leq \delta \). Then for such an \( A \), we have (using Jensen’s inequality relative to normalized counting measure on \( \{1, 2, \ldots, n\} \))

\[
\mathbb{E}\left[\frac{1}{n} S_n(\eta) : A\right] \leq \mathbb{E}\left[\frac{1}{n} S_n(|\eta|^p) : A\right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|\eta_k|^p : A] \leq \frac{1}{n} \varepsilon = \varepsilon.
\]

Another application of Proposition 15.48 shows \( \{\frac{1}{n} S_n(\eta)\} \) is uniformly integrable as

\[
\sup_n \mathbb{E}\left[\frac{1}{n} S_n(\eta)\right]^p \leq \sup_n \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|\eta_k|^p] = \mathbb{E}[|\eta_1|^p] < \infty.
\]

Since \( A_\varepsilon = \{\sup S_n/n > \varepsilon\} \), it follows that

\[
A_\varepsilon = \left\{\sup \frac{S_n}{n} > \varepsilon\right\} \cap A_\varepsilon = \{\sup S_n - n \varepsilon > 0\} \cap A_\varepsilon
\]

\[
= \{\inf S_n(\xi - \varepsilon) > 0\} \cap A_\varepsilon = \{\sup S_n(\xi - \varepsilon) > 0\}.
\]
Finally we need to consider the case where \( f \geq 0 \) but \( f \not\in L^1(P) \). As before, let \( g = \mathbb{E}[f|\mathcal{B}_0] \geq 0 \). For \( r \in (0, \infty) \) and let \( f_r := f \cdot 1_{g \leq r} \). We then have
\[
\mathbb{E}[f_r|\mathcal{B}_0] = \mathbb{E}[f \cdot 1_{g \leq r}|\mathcal{B}_0] = 1_{g \leq r} \mathbb{E}[f|\mathcal{B}_0] = 1_{g \leq r} \cdot g
\]
and in particular, \( \mathbb{E}f_r = \mathbb{E}(1_{g \leq r}) \leq r < \infty \). 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
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \theta^{k-1} = g \text{ a.s. on } \{ g < \infty \}.
\]
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
\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \theta^{k-1} \geq \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \left[ f_r \circ \theta^{k-1} \right] = \mathbb{E}[f \wedge r|\mathcal{B}_0].
\]
Letting \( r \uparrow \infty \) and using the eMCT implies,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \theta^{k-1} \geq \mathbb{E}[f|\mathcal{B}_0] = g
\]
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. (19.5) completes the proof. \( \square \)

As a corollary we have the following version of the strong law of large numbers, also see Theorems 23.31 and Example 21.82 below for other proofs.

**Theorem 19.10 (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 \( \mathbb{E}|X_n| < \infty \).
Part III

Stochastic Processes I
In the sequel \((\Omega, \mathcal{B}, P)\) will be a probability space and \((S, \mathcal{S})\) denote a measurable space which we refer to as state space. If we say that \(f : \Omega \rightarrow S\) is a function we will always assume that it is \(\mathcal{B}/\mathcal{S}\) measurable. We also let \(\mathcal{S}_b\) denote the bounded \(\mathcal{S}/\mathcal{B}_R\) 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 17.21 and Theorem 17.41) and the use of Kolmogorov’s extension Theorem 20.68 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, \(X := \{X_t : \Omega \rightarrow S\}_{t \in T}\), indexed by some parameter space, \(T\). The weakest description of such a stochastic process will be through its "finite dimensional distributions."

**Definition 19.11.** Given a stochastic process, \(X := \{X_t : \Omega \rightarrow S\}_{t \in T}\), and a finite subset, \(A \subset T\), we say that \(\nu_A := \text{Law}_P (\{X_t\}_{t \in A})\) on \((S^A, \mathcal{S}^A)\) 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 12.9.** 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 \(\mathcal{B}_\infty := \bigvee_{t \in T} \mathcal{B}_t := \sigma (\bigcup_{t \in T} \mathcal{B}_t)\). A four-tuple, \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in T}, P)\), 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 \(\Omega \rightarrow S\) is adapted to the filtration if \(X_t\) is \(\mathcal{B}_t/S\) 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 \(\mathcal{B}_t^X := \sigma (X_s; s \leq t)\). 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 \(Q : S \times S \rightarrow [0, 1]\)), we may associate a linear transformation, \(T = T_Q : \mathcal{S}_b \rightarrow \mathcal{S}_b\) defined by

\[
(Tf)(x) = Q(x, f) = \int_S Q(x, dy) f(y) \quad \text{for all } f \in \mathcal{S}_b.
\]  

(19.6)

It is easy to check that \(T\) satisfies;

1. \(T 1 = 1\),
2. \(T f \geq 0\) if \(0 \leq f \in \mathcal{S}_b\),
3. if \(f_n \rightarrow f\) boundedly then \(T f_n \rightarrow T f\) boundedly as well.

Notice that an operator \(T : \mathcal{S}_b \rightarrow \mathcal{S}_b\) satisfying conditions 1. and 2. above also satisfies \(T f \leq T g\) and \(T f\) is real is \(f\). Indeed if \(f = f_+ - f_-\) is real then

\[
T f = T (f_+ - f_-) = T f_+ - T f_-
\]

with \(0 \leq T f_\pm \in \mathbb{R}\) and if \(g \leq f\) then \(0 \leq f - g\) which implies

\[
T f - T g = T (f - g) \geq 0.
\]

As \(\pm f \leq |f|\) when \(f\) is real, we have \(\pm T f \leq T |f|\) and therefore \(|T f| \leq T |f|\). More generally if \(f\) is complex and \(x \in S\), we may choose \(\theta \in \mathbb{R}\) such that \(e^{i\theta} (T f)(x) \geq 0\) and therefore,

\[
|(T f)(x)| = e^{i\theta} (T f)(x) = (T [e^{i\theta} f])(x)
\]

\[
= (T \text{Re} [e^{i\theta} f]) (x) + i (T \text{Im} [e^{i\theta} f]) (x).
\]

Furthermore we must have \((T \text{Im} [e^{i\theta} f])(x) = 0\) and using \(\text{Re} [e^{i\theta} f] \leq |f|\) we find,

\[
|(T f)(x)| = |(T \text{Re} [e^{i\theta} f])(x)| \leq (T |f|)(x).
\]

As \(x \in S\) was arbitrary we have shown that \(|T f| \leq T |f|\). Thus if \(|f| \leq M\) for some \(0 \leq M < \infty\) we may conclude,

\[
|T f| \leq T |f| \leq T (M \cdot 1) = MT 1 = M.
\]

**Proposition 19.13.** If \(T : \mathcal{S}_b \rightarrow \mathcal{S}_b\) is a linear transformation satisfying the three properties listed after Eq. (19.6), then \(Q(x, A) := (T 1_A)(x)\) for all \(A \in S\) and \(x \in S\) is a probability kernel such that Eq. (19.6) holds.

The proof of this proposition is straightforward and will be left to the reader. Let me just remark that if \(Q(x, A) := (T 1_A)(x)\) for all \(x \in S\) and \(A \in S\) then \(T f = Q(\cdot, f)\) for all simple functions in \(\mathcal{S}_b\) and then by approximation for all \(f \in \mathcal{S}_b\).

**Corollary 19.14.** If \(Q_1\) and \(Q_2\) are two probability kernels on \((S, S) \times (S, S)\), then \(T_{Q_1} T_{Q_2} = T_Q\) where \(Q\) is the probability kernel given by

\[
Q(x, A) = (T_{Q_1} T_{Q_2} 1_A)(x) = Q_1(x, Q_2(\cdot, A)) = \int_S Q_1(x, dy) Q_2(y, A)
\]

for all \(A \in S\) and \(x \in S\). We will denote \(Q\) by \(Q_1 Q_2\).
From now on we will identify the probability kernel \( Q : S \times S \rightarrow [0, 1] \) 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 19.15.** Suppose that \( \nu \) is a probability measure on \((S, S)\) and \( Q_k : S \times S \rightarrow [0, 1] \) are probability kernels on \((S, S) \times (S, S)\) for \( 1 \leq k \leq n \). Then there exists a probability measure \( \mu \) on \((S^{n+1}, S^{\otimes(n+1)})\) such that for all \( f \in S^{\otimes(n+1)} \) we have

\[
\mu(f) = \int_S d\nu(x_0) \int_S Q_1(x_0, dx_1) \int_S Q_2(x_1, dx_2) \cdots \int_S Q_n(x_{n-1}, dx_n) f(x_0, \ldots, x_n). \tag{19.7}
\]

**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,**

\[
d\mu(x_0, \ldots, x_n) = d\nu(x_0) Q_1(x_0, dx_1) Q_2(x_1, dx_2) \cdots Q_n(x_{n-1}, dx_n).
\]

**Proof.** The fact that all of the iterated integrals make sense in Eq. (19.7) follows from Exercise 17.5, the measurability statements in Fubini’s theorem, and induction. The measure \( \mu \) is defined by setting \( \mu(A) = \mu(1_A) \) for all \( A \in S^{\otimes(n+1)} \). It is a simple matter to check that \( \mu \) is a measure on \((S^{n+1}, S^{\otimes(n+1)})\) and that \( \int_S f d\mu \) agrees with the right side of Eq. (19.7) for all \( f \in S^{\otimes(n+1)} \).

**Remark 19.16.** As usual the measure \( \mu \) is determined by its value on product functions of the form \( f(x_0, \ldots, x_n) = \prod_{i=0}^n f_i(x_i) \) with \( f_i \in S_b \). For such a function we have

\[
\mu(f) = \mathbb{E}_{\nu} [f_0 Q_1 M_{f_1} Q_2 M_{f_2} \cdots Q_{n-1} M_{f_{n-1}} Q_n f_n]
\]

where \( M_f : S_b \rightarrow S_b \) is defined by \( M_f g = fg \), i.e. \( M_f \) is multiplication by \( f \).
The Markov Property

For purposes of this section, $T = \mathbb{N}_0$ or $\mathbb{R}^+$, $(\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in T}, P)$ 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 20.1 (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

$$E_{\mathcal{B}_s} f(X_t) := E[ f(X_t) | \mathcal{B}_s] = E[ f(X_t) | X_s] \text{ a.s. for all } 0 \leq s < t \quad (20.1)$$

and for every $f \in \mathcal{S}_b$.

If Eq. (20.1) holds then by the factorization Lemma 2.40 there exists $F \in \mathcal{S}_b$ such that $F(X_s) = E[ f(X_t) | X_s]$. Conversely if we want to verify Eq. (20.1) it suffices to find an $F \in \mathcal{S}_b$ such that $E_{\mathcal{B}_s} f(X_t) = F(X_s)$ a.s. This is because, by the tower property of conditional expectation,

$$E_{\mathcal{B}_s} f(X_t) = F(X_s) = E[ F(X_s) | X_s] = E[ E_{\mathcal{B}_s} f(X_t) | X_s] = E[ f(X_t) | X_s] \text{ a.s.} \quad (20.2)$$

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 20.4 below) that given $X_s$ what the process did before time $s$ is independent of the what it will do after time $s$.

**Lemma 20.2.** 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 $\{B_t^X = \sigma(X_s : s \leq t)\}_{t \geq 0}$.

**Proof.** It is clear that $\{X_t\}_{t \in T}$ is $B_t^X$-adapted and that $\sigma(X_s) \subset B_t^X \subset \mathcal{B}_t$ for all $s \in T$. Therefore using the tower property of conditional expectation we have,

$$E_{B_t^X} f(X_t) = E_{B_t^X} E_{B_t} f(X_t) = E_{B_t^X} E_{\sigma(X_s)} f(X_t) = E_{\sigma(X_s)} f(X_t).$$

$$\square$$

**Remark 20.3.** If $T = \mathbb{N}_0$, a stochastic process $\{X_n\}_{n \geq 0}$ is Markov iff for all $f \in \mathcal{S}_b$,

$$E[ f(X_{m+1}) | B_m] = E[ f(X_{m+1}) | X_m] \text{ a.s. for all } m \geq 0 \quad (20.3)$$

Indeed if Eq. (20.3) holds for all $m$, we may use induction on $n$ to show

$$E[ f(X_n) | B_m] = E[ f(X_n) | X_m] \text{ a.s. for all } n \geq m. \quad (20.4)$$

It is clear that Eq. (20.4) holds for $n = m$ and $n = m + 1$. So now suppose Eq. (20.4) holds for a given $n \geq m$. Using Eq. (20.3) with $m = n$ implies

$$E_{B_n} f(X_{n+1}) = E[ f(X_{n+1}) | X_n] = F(X_n)$$

for some $F \in \mathcal{S}_b$. Thus by the tower property of conditional expectations and the induction hypothesis,

$$E_{B_m} f(X_{n+1}) = E_{B_m} E_{B_n} f(X_{n+1}) = E_{B_m} F(X_n) = E[ F(X_n) | X_m]$$

$$= E[ E_{B_n} f(X_{n+1}) | X_m] = E[ f(X_{n+1}) | X_m].$$

The next theorem and Exercise 20.1 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 20.4 (Markov Independence).** Suppose $\{X_t\}_{t \in T}$ 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. $E[G | \mathcal{B}_s] = E[G | X_s]$ for all $G \in (\mathcal{F}_s)_b$ and
2. $\mathcal{B}_s$ is independentootnote{In words, the future and past are independent given the present.} of $\mathcal{F}_s$ given $X_s$ which we abbreviate as $\mathcal{B}_s \perp \perp \mathcal{F}_s$. 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

$$E[FG | X_s] = E[F | X_s] \cdot E[G | X_s] \text{ a.s.} \quad (20.5)$$

for all $F \in (\mathcal{B}_s)_b$ and $G \in (\mathcal{F}_s)_b$ and $s \in T$. 

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 $g_i \in S_b$. Then by the Markov property and the tower property of conditional expectations,

\[
\mathbb{E}_{B_s} [G] = \mathbb{E}_{B_s} \left[ \prod_{i=0}^{n} g_i (X_{t_i}) \right] = \mathbb{E}_{B_s} \mathbb{E}_{B_{t_{n-1}}} \left[ \prod_{i=0}^{n} g_i (X_{t_i}) \right]
\]

\[
= \mathbb{E}_{B_s} \prod_{i=0}^{n-1} g_i (X_{t_i}) \cdot \mathbb{E}_{B_{t_{n-1}}} g_n (X_{t_n})
\]

\[
= \mathbb{E}_{B_s} \prod_{i=0}^{n-1} g_i (X_{t_i}) \cdot \mathbb{E} [g_n (X_{t_n}) | X_{t_{n-1}}] = \mathbb{E}_{B_s} \left[ \tilde{G} \right]
\]

where

\[
\tilde{G} := \left( \prod_{i=0}^{n-2} g_i (X_{t_i}) \right) \cdot (g_{n-1} g) (X_{t_{n-1}})
\]

where $g \in S_b$ is chosen so that $\mathbb{E} [g_n (X_{t_n}) | X_{t_{n-1}}] = g (X_{t_{n-1}})$ a.s.. Continuing this way inductively we learn that $\mathbb{E}_{B_s} [G] = F (X_s)$ a.s. for some $F \in S_b$ and therefore,

\[
\mathbb{E} [G | X_s] = \mathbb{E} [\mathbb{E}_{B_s} G | X_s] = \mathbb{E} [F (X_s) | X_s] = F (X_s) = \mathbb{E}_{B_s} [G] \text{ a.s.}
\]

An application of the multiplicative system Theorem 11.2 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,

\[
\mathbb{E}_{B_s} [FG] = F \mathbb{E}_{B_s} [G] = F \cdot \mathbb{E} [G | X_s].
\]

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. (20.5).

Exercise 20.1. Suppose that $\{X_t\}_{t \geq 0}$ is an adapted stochastic process such that Eq. (20.5) 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.

20.1 Markov Processes

If $S$ is a standard Borel space (i.e. $S$ is isomorphic to a Borel subset of $[0,1]$), we may find regular conditional probability kernels, $Q_{s,t}$ on $(S,S) \times (S,S) \to [0,1]$ for all $0 \leq s < t$ such that

\[
\mathbb{E} [f (X_t) | X_s] = Q_{s,t} (X_s; f) = (Q_{s,t} f) (X_s) \text{ a.s.}
\]

Moreover by the Markov property if $0 \leq \sigma < s < t$, then

\[
(Q_{\sigma,t} f) (X_\sigma) = \mathbb{E} [f (X_t) | X_\sigma] = \mathbb{E} \left[ \mathbb{E}_{B_\sigma} f (X_t) | X_\sigma \right] = \mathbb{E} [(Q_{\sigma,t} f) (X_s) | X_\sigma] = (Q_{\sigma,s} Q_{s,t} f) (X_\sigma) \text{ P \text{– a.s.}}
\]

\[
= Q_{\sigma,s} (X_\sigma; Q_{s,t} (\cdot ; f)) \quad P \text{– a.s.}
\]

If we let $\mu_\sigma := \text{Law}_P (X_\sigma) : S \to [0,1]$ for all $t \in T$, we have just shown that for every $f \in S_b$, that

\[
Q_{\sigma,t} (\cdot ; f) = Q_{\sigma,s} (\cdot ; Q_{s,t} (\cdot ; f)) \quad \mu_\sigma \text{ – a.s.}
\]

In the sequel we want to assume that such kernels exists and that Eq. (20.7) holds for everywhere not just $\mu_\sigma$ – a.s. Thus we make the following definitions.

Definition 20.5 (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 $Q_{s,t} (x,dy) = \delta_x (dy)$ (as an operator $Q_{s,t} = I_{S_b}$) for all $s \in T$ and the Chapman-Kolmogorov equations hold:

\[
Q_{\sigma,t} = Q_{\sigma,s} Q_{s,t} \text{ for all } 0 \leq \sigma \leq s \leq t.
\]

Recall that Eq. (20.8) is equivalent to

\[
Q_{\sigma,t} (x,A) = \int_S Q_{\sigma,s} (x,dy) Q_{s,t} (y,A) \text{ for all } x \in S \text{ and } A \in S
\]

or

\[
Q_{\sigma,t} (x;f) = Q_{\sigma,s} (x;Q_{s,t} (\cdot ; f)) \text{ for all } x \in S \text{ and } f \in S_b.
\]

Thus Markov transition kernels should satisfy Eq. (20.7) 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 $A \in S$ 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. (20.9) 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 $A \in S$ at time $t$. This interpretation is rigorously true when $S$ is a finite or countable set.

Definition 20.6 (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. (20.6) holds.
Definition 20.7. A stochastic process, \( \{X_t : \Omega \to S := \mathbb{R}^d \}_{t \in T} \), has independent increments if for all finite subsets, \( \Lambda = \{0 \leq t_0 < t_1 < \cdots < t_n \} \subseteq T \) the random variables \( \{X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}} \}_{k=1}^n \) are independent. We refer to \( X_t - X_s \) for \( s < t \) as an increment of \( X \).

Exercise 20.2. Suppose that \( \{X_t : \Omega \to S := \mathbb{R}^d \}_{t \in T} \) is a stochastic process with independent increments and let \( B_t := B_t^X \) for all \( t \in T \). Show, for all \( 0 \leq s < t \), that \( (X_t - X_s) \) is independent of \( B_s \) and then use this to show \( \{X_t\}_{t \in T} \) is a Markov process with transition kernels defined by \( 0 \leq s < t \),

\[
Q_{s,t}(x,A) := \mathbb{E} [1_A (x + X_t - X_s)] \quad \text{for all } A \in \mathcal{S} \text{ and } x \in \mathbb{R}^d. \tag{20.11}
\]

You should verify that \( \{Q_{s,t}\}_{0 \leq s \leq t} \) are indeed Markov transition kernels, i.e. satisfy the Chapman-Kolmogorov equations.

Example 20.8 (Random Walks). Suppose that \( \{\xi_n : \Omega \to S := \mathbb{R}^d \}_{n=0}^{\infty} \) are independent random vectors and \( X_n := \sum_{k=0}^{m} \xi_k \) and \( \mathcal{B}_m := \sigma(\xi_0, \ldots, \xi_m) \) for each \( m \in T = \mathbb{N}_0 \). Then \( \{X_m\}_{m \geq 0} \) has independent increments and therefore has the Markov property with Markov transition kernels being given by

\[
Q_{s,t}(x,f) = \mathbb{E} [f (x + X_t - X_s)] = \mathbb{E} \left[ f \left( x + \sum_{s < k \leq t} \xi_k \right) \right]
\]
or in other words,

\[
Q_{s,t}(x,\cdot) = \text{Law}_P \left( x + \sum_{s < k \leq t} \xi_k \right).
\]

The one step transition kernels are determined by

\[
(Q_{n,n+1})_m(x) = \mathbb{E} [f (x + \xi_{n+1})] \quad \text{for } n \in \mathbb{N}_0.
\]

Exercise 20.3. Let us now suppose that \( \{\xi_n : \Omega \to S\}_{n=0}^{\infty} \) are independent random vectors where \( (S,S) \) is a general measurable space, \( \mathcal{B}_n := \sigma(\xi_0, \xi_1, \ldots, \xi_n) \) for \( n \geq 0 \), \( u_n : S \times S \to S \) are measurable functions for \( n \geq 1 \), and \( X_n : \Omega \to S \) for \( n \in \mathbb{N}_0 \) are defined by \( X_0 = \xi_0 \) and then inductively for \( n \geq 1 \) by

\[
X_{n+1} = u_{n+1} (X_n, \xi_{n+1}) \quad \text{for } n \geq 0.
\]

Convince yourself that for \( 0 \leq m < n \) there is a measurable function, \( \varphi_{n,m} : S^{n-m+1} \to S \) determined by the \( \{u_k\} \) such that \( X_n = \varphi_{n,m} (X_m, \xi_{m+1}, \ldots, \xi_n) \). (You need not write the proof of this assertion in your solution.) In particular,

\[
X_n = \varphi_{n,0} (\xi_0, \ldots, \xi_n) \quad \text{is } \mathcal{B}_n/S - \text{measurable so that } X = \{X_n\}_{n \geq 0} \text{ is adapted. Show } \{X_n\}_{n \geq 0} \text{ is a Markov process with transition kernels,}
\]

\[
Q_{m,n}(x,\cdot) = \text{Law}_P (\varphi_{n,m} (x, \xi_{m+1}, \ldots, \xi_n)) \quad \text{for } 0 \leq m \leq n.
\]

where (by definition) \( Q_{m,n}(x,\cdot) = \delta_x (\cdot) \). Please explicitly verify that \( \{Q_{m,n}\}_{0 \leq m \leq n} \) are Markov transition kernels, i.e. satisfy the Chapman-Kolmogorov equations.

Remark 20.9. Suppose that \( T = \mathbb{N}_0 \) and \( \{Q_{m,n} : 0 \leq m \leq n\} \) are Markov transition kernels on \( S \times S \). Since

\[
Q_{m,n} = Q_{m,m+1} Q_{m+1,m+2} \cdots Q_{n-1,n},
\]

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. \( (20.12) \), then the resulting \( \{Q_{m,n} : 0 \leq m \leq n\} \) are Markov transition kernels on \( S \times S \). Moreover if \( S \) is a countable set, then we may let

\[
q_{m,n}(x,y) := Q_{m,n}(x,\{y\}) = P (X_n = y | X_m = x) \quad \text{for all } x,y \in S \tag{20.13}
\]

so that

\[
Q_{m,n}(x,A) = \sum_{y \in A} q_{m,n}(x,y).
\]

In this case it is easily checked that

\[
q_{m,n} (x,y) = \sum_{x_i \in S : m < i \leq n} q_{m,m+1} (x,x_{m+1}) q_{m+1,m+2} (x_{m+1},x_{m+2}) \cdots q_{n-1,n} (x_{n-1},y). \tag{20.14}
\]

The reader should observe that this is simply matrix multiplication!

Exercise 20.4 (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 20.10 (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 $\nu := \text{Law}_\nu (X_0)$, then for all $0 = t_0 < t_1 < t_2 < \cdots < t_n$ we have

$$\text{Law}_\nu (X_{t_0}, X_{t_1}, \ldots, X_{t_n}) (dx_0, dx_1, \ldots, dx_n) = d\nu (x_0) \prod_{i=1}^n Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

or equivalently,

$$E \left[ f (X_{t_0}, X_{t_1}, \ldots, X_{t_n}) \right] = \int_{S_{n+1}} f (x_0, x_1, \ldots, x_n) d\nu (x_0) \prod_{i=1}^n Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

for all $f \in S^\otimes (n+1)$.

Proof. Because of the multiplicative system Theorem 11.2 it suffices to prove Eq. (20.16) for functions of the form $f (x_1, \ldots, x_n) = \prod_{i=0}^n f_i (x_i)$ where $f_i \in S_b$. 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

$$E [f (X_{t_0}, X_{t_1}, \ldots, X_{t_n})] = E \left[ \prod_{i=0}^n f_i (X_{t_i}) \right] = E \left[ \prod_{i=0}^n f_i (X_{t_i}) \right] \prod_{i=0}^{n-1} Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

$$= \int_{S_{n+1}} \prod_{i=0}^n f_i (x_i) d\nu (x_0) \prod_{i=1}^{n-1} Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

as desired.$\blacksquare$

Theorem 20.11 (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 $\Omega := ST$, $X_t : \Omega \rightarrow S$ be the projection map, $X_t (\omega) = \omega (t)$ and $\mathcal{B}_{t} = \mathcal{B}^X_t = \sigma (X_s : s \leq t)$ 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) $\text{Law}_{P_\nu} (X_0) = \nu$ 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 20.68 as described in the appendix to this chapter. I will only briefly sketch the proof here.

For each $A = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$, let $P_A$ be the measure on $(S^{n+1}, \mathcal{S}^{\otimes (n+1)})$ defined by

$$dP_A (x_0, x_1, \ldots, x_n) = d\nu (x_0) \prod_{i=1}^n Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

Using the Chapman-Kolmogorov equations one shows that the $\{P_A\}_{A \subset f T}$ ($A \subset f T$ denotes a finite subset of $T$) are consistently defined measures as described in the statement of Theorem 20.68. Therefore it follows by an application of that theorem that there exists a unique measure $P_\nu$ on $(\Omega, \mathcal{B})$ such that

$$\text{Law}_{P_\nu} (X | A) = P_A$$

for all $A \subset f T$. (20.17)

In light of Theorem 20.10 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 the case the finite dimensional distributions must be given as in Eq. (20.17) and therefore $P_\nu$ is uniquely determined. So let us now verify the desired Markov property.

Again let $A = \{0 = t_0 < t_1 < t_2 < \cdots < t_n\} \subset T$ with $t_{n-1} = s < t = t_n$ and suppose that $f (x_0, \ldots, x_n) = h (x_0, \ldots, x_{n-1}) g (x_n)$ with $h \in \mathcal{S}_b^{\otimes n}$ and $f \in S_b$. By the definition of $P_\nu$ we then have (writing $\mathbb{E}_\nu$ for $\mathbb{E}_{P_\nu}$),

$$\mathbb{E}_\nu \left[ h \left( X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}} \right) g (X_{t_n}) \right]$$

$$= \int_{S^{n+1}} h (x_0, x_1, \ldots, x_{n-1}) g (x_n) d\nu (x_0) \prod_{i=1}^n Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

$$= \int_{S^{n+1}} h (x_0, x_1, \ldots, x_{n-1}) Q_{t_{n-1}, t_n} (x_{n-1}, dx_{n-1}) d\nu (x_0) \prod_{i=1}^{n-1} Q_{t_{i-1}, t_i} (x_{i-1}, dx_i)$$

$$= \mathbb{E}_\nu \left[ h \left( X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}} \right) Q_{t_{n-1}, t_n} (X_{t_{n-1}}, g) \right]$$

$$= \mathbb{E}_\nu \left[ h \left( X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}} \right) Q_{s,t} (X_s, g) \right]$$

It then follows by an application of the multiplicative system theorem that

$$\mathbb{E}_\nu \left[ H g (X_t) \right] = \mathbb{E}_\nu \left[ HQ_{s,t} (X_s, g) \right]$$

for all $H \in (\mathcal{B}_s)_b$ and therefore that

$$\mathbb{E}_\nu \left[ g (X_t) \mid \mathcal{B}_s \right] = Q_{s,t} (X_s, g) \text{ a.s.}$$
We are now going to specialize to the more manageable class of “time homogeneous” Markov processes.

**Definition 20.12.** We say that a collection of Markov transition kernels, \( \{Q_{s,t}\}_{0 \leq s \leq t} \) are time homogeneous if \( Q_{s,t} = Q_{0,t-s} \) for all \( 0 \leq s \leq t \). In this case we usually let \( Q_t := Q_{0,t-t} \). The condition that \( Q_{s,t} (x, \cdot) = \delta_x \) now reduces to \( Q_0 (x, \cdot) = \delta_x \) and the Chapmann-Kolmogorov equations reduce to

\[
Q_s Q_t = Q_{s+t} \quad \text{for all} \quad s, t \geq 0,
\]

i.e.

\[
\int_S Q_s (x, dy) Q_t (y, A) = Q_{s+t} (x, A) \quad \text{for all} \quad s, t \geq 0, \quad x \in S, \quad \text{and} \quad A \in S.
\]

A collection of operators \( \{Q_t\}_{t \geq 0} \) with \( Q_0 = \text{Id} \) satisfying Eq. (20.18) is called a one parameter semi-group.

**Definition 20.13.** A Markov process is time homogeneous if it has time homogeneous Markov transition kernels. In this case we will have,

\[
\mathbb{E} [f (X_t) | B_s] = Q_{t-s} (X_s, f) = (Q_{t-s} f) (X_s) \quad \text{a.s.} \quad (20.20)
\]

for all \( 0 \leq s \leq t \) and \( f \in \mathcal{S}_b \).

**Theorem 20.14 (The time homogeneous Markov property).** Suppose that \( (S, S) \) is a measurable space, \( Q_t : S \times S \to [0, 1] \) are time homogeneous Markov transition kernels, \( (\Omega, \mathcal{B}, \{B_t\}_{t \geq 0}) \) is a filtered measure space, \( \{X_t : \Omega \to S\}_{t \geq 0} \) 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. \( \{X_t\}_{t \geq 0} \) is a time homogeneous Markov process with transition kernels \( \{Q_t\}_{t \geq 0} \) relative to \( P_x \).

Under these assumptions we have the following conclusions.

1. If \( F \in \mathcal{S}_b^\otimes T \), then \( S \ni x \to \mathbb{E}_x F (X) \) is \( S/\mathcal{B}_\mathbb{R} \) – measurable.
2. If \( P \) is any probability measure on \( (\Omega, \mathcal{B}) \) such that \( (\Omega, \mathcal{B}, \{B_t\}_{t \geq 0}, \{X_t\}_{t \geq 0}, \{Q_t\}_{t \geq 0}, P) \) is a time homogeneous Markov process (with transition kernels being \( \{Q_t\} \)), then for all \( t \geq 0 \),

\[
\mathbb{E}_P [F (X_{t+}) | B_t] = \mathbb{E}_P [F (X_{t+}) | X_t] = \mathbb{E}_X [F (X)] \quad P \text{ – a.s.} \quad (20.21)
\]

**Warning:** In this equation \( \mathbb{E}_X \) does not denote \( \mathbb{E}_{\sigma (X_t)} = \mathbb{E} [\cdot | X_t] \) but instead \( \mathbb{E}_P \) 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,

\[
\mathbb{E}_P [F (X_{t+}) | B_t] (\omega) = \mathbb{E}_{X_t (\omega)} [F (X)] = \int_{\Omega} F (X (\omega')) P_{X_t (\omega)} (d\omega').
\]

**Proof.** Let \( F (X) := f (X_{t_0}, \ldots, X_{t_n}) \) for some \( f \in \mathcal{S}_b^\otimes (n+1) \).

1. For this \( F \) we have

\[
\mathbb{E}_x [F (X)] = \int_{Q_{0,t_1} (x, dx_1) \ldots Q_{t_{n-1},t_n} (x_{n-1}, dx_n) f (x_0, \ldots, x_n)
\]

which is \( S/\mathcal{B}_\mathbb{R} \) – 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 \( F (X_{t+}) \in \mathcal{F}_{b} \), Theorem 20.4 already implies

\[
\mathbb{E}_P [F (X_{t+}) | B_t] = \mathbb{E}_P [F (X_{t+}) | X_t] \quad P \text{ – a.s.}
\]

To compute the last conditional expectation, let us first assume that \( F (X) := f (X_{t_0}, \ldots, X_{t_n}) \) and \( g \in \mathcal{S}_b \), and let \( \nu_t = \text{Law}_P (X_t) = \nu \). Then using Theorem 20.10 twice we learn,

\[
\mathbb{E}_P [g (X_t) F (X_{t+})] = \mathbb{E}_P [g (X_t) f (X_{t_0+t}, \ldots, X_{t_n+t})]
\]

\[
= \int g (x_0) f (x_0, \ldots, x_n) \nu_0 (y, dx_0) \prod_{j=1}^n Q_{t_j-t_{j-1}} (x_{j-1}, dx_j)
\]

\[
= \int (x_0) f (x_0, \ldots, x_n) \nu_t (x, dx) \prod_{j=1}^n Q_{t_j-t_{j-1}} (x_j, dx_j)
\]

\[
= \int \nu_t (x_0) g (x_0) \mathbb{E}_x [F (X_{t_0}, \ldots, X_{t_n})]
\]

\[
= \int \nu_t (x_0) g (x_0) \mathbb{E}_x [F (X)] = \mathbb{E}_P [g (X_t) \mathbb{E}_X [F (X)]]
\]

\[2\] Unfortunately we now have a lot of different meanings for \( \mathbb{E}_x \) 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 \( \mathbb{E}_{\delta_x} \).
An application of the multiplicative systems Theorem 11.2 shows this equation is valid for all $F \in S_{\mathcal{B}}^T$ and this tells us that
\[
\mathbb{E}_P [F (X_{t+\cdot}) \mid X_t] = \mathbb{E}_X [F (X)] \quad P_x \text{ a.s.}
\]
for all $F \in S_{\mathcal{B}}^T$.

\[\tag{20.22}\]

Remark 20.15. Admittedly Theorem 20.14 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 20.4 and 20.14 to find $\{X_s\}_{s \leq t}$ and $\{X_s\}_{s \geq t}$ are $P (\cdot) = x$ -- independent and moreover,
\[
\text{Law}_{P (\cdot) = x} (X_{t+\cdot}) = \text{Law}_{P_x} (X_t).
\]

Last assertion simply states that given $X_t = x$ the process, $X$, after time $t$ behaves just like the process starting afresh from $x$.

20.2 Discrete Time Homogeneous Markov Processes

The proof of the following easy lemma is left to the reader.

Lemma 20.16. If $Q_n : S \times S \rightarrow [0, 1]$ 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 20.17 (Random Walks Revisited). Suppose that $\xi_0 : \Omega \rightarrow S := \mathbb{R}^d$ is independent of $\{\xi_n : \Omega \rightarrow S := \mathbb{R}^d\}_{n=1}^\infty$ which are now assumed to be i.i.d. If $X_m = \sum_{k=0}^m \xi_k$ is as in Example 20.8 then $\{X_m\}_{m \geq 0}$ is a time homogeneous Markov process with
\[
Q_m (x, \cdot) = \text{Law}_P (X_m - X_0)
\]
and the one step transition kernel, $Q = Q_1$, is given by
\[
Qf (x) = Q (x, f) = \mathbb{E} [f (x + \xi_1)] = \int_S f (x+y) \rho (y)
\]
where $\rho := \text{Law}_P (\xi_1)$. For example if $d = 1$ and $P (\xi_1 = 1) = p$ and $P (\xi_1 = -1) = q := 1 - p$ for some $0 \leq p \leq 1$, then we may take $S = \mathbb{Z}$ and we then have
\[
Qf (x) = Q (x, f) = pf (x+1) + qf (x-1).
\]

Example 20.18 (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,
\[
P (X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i - 1, i + 1\} \\ \frac{N}{N+1} & \text{if } j = i - 1 \\ \frac{N}{N+1} & \text{if } j = i + 1
\end{cases} := q (i, j)
\]

As these probabilities do not depend on $n$, $\{X_n\}$ is a time homogeneous Markov chain.

Exercise 20.5. 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 $q (i, j) := P (X_{n+1} = j \mid X_n = i)$, where $X_n$ denotes the room the rat is in at time $n$.

Exercise 20.6 (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. If you are at site $i$ and 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 20.1.

\[\text{Fig. 20.1. The generic jump diagram for a two state Markov chain.}\]

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) 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 the lazy chain will spend lounging at each site \( x \in S \).

Exercise 20.7 (Jump - Hold Description 1). Let \( S \) be a countable set \( (\Omega,B,\{B_n\}_{n=0}^\infty, P, \{Y_n\}_{n=0}^\infty) \) 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,

\[
\tilde{q}(x,y) := \begin{cases} 
\frac{q(x,y)}{1-q(x,x)} & \text{if } x \neq y \\
1 & \text{if } x = y.
\end{cases}
\]

Let \( j_k \) denote the time of the \( k \)-th jump of the chain \( \{Y_n\}_{n=0}^\infty \) so that

\[
j_1 := \inf \{n > 0 : Y_n \neq Y_0 \} \quad \text{and} \quad j_{k+1} := \inf \{n > j_k : Y_n \neq Y_{j_k} \}
\]

with the convention that \( j_0 = 0 \). Further let \( \sigma_k := j_k - j_{k-1} \) denote the time spent between the \( (k-1) \)-th and \( k \)-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 \( k = 1, \ldots, n \) and \( m_1, \ldots, m_k \in \mathbb{N} \), show

\[
P(\bigcap_{k=0}^n \{Y_k = x_k\}) \cap \bigcap_{k=1}^n \{\sigma_k = m_k\})
\]

\[
= P(x_0) \prod_{k=1}^n \tilde{q}(x_{k-1},x_k)(1 - q(x_{k-1},x_{k-1})) \cdot \tilde{q}(x_{k-1},x_k).
\]

(20.23)

2. Summing the previous formula on \( m_1, \ldots, m_k \in \mathbb{N} \), conclude

\[
\frac{d}{dt}|_{t=0} Q_tf = \lim_{t \to 0} Q_t f - f \cdot \frac{t}{t}. \tag{20.24}
\]

For now we make the (often unreasonable assumption) that the limit in Eq. (20.24) holds for all \( f \in S_t \). This assumption is OK when \( S \) is a finite (see Remark 20.19) or sometimes even when \( S \) is a countable state space. For more complicated states spaces we will have to restrict the set of \( f \in S_t \) that we consider when computing \( Af \) by Eq. (20.24). You should get a feeling for this issue by working through Exercise 20.10 which involves “Brownian motion.”

Remark 20.19 (L. Gårding’s trick). If \( Q_t \) is a Markov-Semi group on \( S = \{1,2,\ldots,n\} \) depending continuously on \( t \) then for \( \varepsilon > 0 \) sufficiently small,
is invertible. This is because
\[ \| (I - B_\varepsilon) \| = \left\| \frac{1}{\varepsilon} \int_0^\varepsilon (I - Q_s) ds \right\| \leq \frac{1}{\varepsilon} \int_0^\varepsilon \| I - Q_s \| ds \to 0 \quad \text{as} \quad \varepsilon \downarrow 0 \]
and as is well know \( B_\varepsilon = I - (I - B_\varepsilon) \) is invertible provided \( \| I - B_\varepsilon \| < 1 \).

Since
\[ Q_t B_\varepsilon = \frac{1}{\varepsilon} \int_0^\varepsilon Q_t Q_s ds = \frac{1}{\varepsilon} \int_0^\varepsilon Q_{t+s} ds = \frac{1}{\varepsilon} \int_t^{t+\varepsilon} Q_s d\tau \]
we may conclude by the fundamental theorem of calculus that
\[ \frac{d}{dt} Q_t B_\varepsilon = \frac{1}{\varepsilon} [Q_{t+\varepsilon} - Q_t] . \]

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 \( \frac{d}{dt} |_{t=0} Q_t f \) exists. Under these assumptions, if \( t, h > 0 \), then using the semi-group property we have
\[ Q_{t+h} - Q_t = (Q_h - I) Q_t = Q_t (Q_h - I) . \tag{20.25} \]

Therefore,
\[ Q_{t+h} f - Q_t f = (Q_h - I) Q_t f \to 0 \quad \text{as} \quad h \downarrow 0 \]
so that \( Q_t f \) is right continuous. Similarly,
\[ Q_{t-h} - Q_t = -(Q_h - I) Q_{t-h} = -(Q_h - I) Q_t \]
and
\[ |Q_{t-h} f - Q_t f| = |Q_{t-h} (Q_h - I) f| \leq |Q_{t-h}| (Q_h - I) f| \leq \sup_{S} |(Q_h - I) f| \]
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. \( \tag{20.25} \) divided by \( h \) and Eq. \( \tag{20.26} \) 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 \( \{ Q_t \}_{t \geq 0} \) is uniquely determined by its infinitesimal generator \( A \) by solving the differential equation,
\[ \frac{d}{dt} Q_t = A Q_t = Q_t A \quad \text{with} \quad Q_0 = I d. \tag{20.27} \]

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 20.20.** Suppose that \( S = \{1,2,\ldots,n\} \) and \( Q_t \) is a Markov-semigroup with infinitesimal generator, \( A \), so that \( \frac{d}{dt} Q_t = AQ_t = Q_tA \). By assumption \( Q_t \{i,j\} \geq 0 \) 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 \( Q_1 = I \) for all \( t \geq 0 \) where \( I \) denotes the identity matrix in \( \mathbb{R}^n \) with all entries being \( 1 \). Differentiating \( Q_1 = I \) 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 \( i \neq j \) we will have,
\[ A_{ij} = \lim_{t \downarrow 0} \frac{Q_t \{i,j\}}{t} \geq 0. \]

Thus we have shown the infinitesimal generator, \( A \), of \( Q_t \) must satisfy \( A_{ij} \geq 0 \) for all \( i \neq j \) 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 \( \boxed{20.9} \). 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 20.9.** Suppose that \( S = \{1,2,\ldots,n\} \) and \( A \) is a matrix such that \( A_{ij} \geq 0 \) for \( i \neq j \) and \( \sum_{j=1}^n A_{ij} = 0 \) for all \( i \). Show
\[ Q_t = e^{tA} := \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n \tag{20.28} \]
is a time homogeneous Markov kernel.

**Hints:**
1. To show \( Q_t \{i,j\} \geq 0 \) for all \( t \geq 0 \) and \( i,j \in S \), write \( Q_t = e^{-t\lambda} e^{t(M+A)} \) where \( \lambda > 0 \) is chosen so that \( M + A \) 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 20.21 (Feynmann-Kac Formula). Continue the notation in Exercise 20.9 and let $(\Omega, \{\mathcal{B}_t\}_{t \geq 0}, P_x, \{X_t\}_{t \geq 0})$ 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

$$(T_t g)(x) = \mathbb{E}_x \left[ \exp \left( \int_0^t V(X_s) \, ds \right) g(X_t) \right] \tag{20.29}$$

for all $g : S \to \mathbb{R}$. Then $T_t$ satisfies,

$$\frac{d}{dt} T_t = T_t (A + M_V) \quad \text{with} \quad T_0 = I \tag{20.30}$$

where $M_V g := V g$ for all $g : S \to \mathbb{R}$, i.e. $M_V$ is the diagonal matrix with $V(1), \ldots, V(n)$ being placed in the diagonal entries. We may summarize this result as,

$$\mathbb{E}_x \left[ \exp \left( \int_0^t V(X_s) \, ds \right) g(X_t) \right] = \left( e^{t(A+M_V)} g \right)(x). \tag{20.31}$$

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, $\frac{d}{dt} (T_t g) (x) = \frac{d}{dh} |_{h=0} (T_{t+h} g) (x)$. Then by the chain rule, the fundamental theorem of calculus, and the Markov property, we find

$$\frac{d}{dt} (T_t g)(x) = \frac{d}{dh} |_{h=0} \mathbb{E}_x \left[ \exp \left( \int_0^{t+h} V(X_s) \, ds \right) g(X_t) \right]$$

$$+ \frac{d}{dh} |_{h=0} \mathbb{E}_x \left[ \exp \left( \int_0^t V(X_s) \, ds \right) g(X_{t+h}) \right]$$

$$= \mathbb{E}_x \left[ \frac{d}{dh} |_{h=0} \exp \left( \int_0^{t+h} V(X_s) \, ds \right) g(X_t) \right]$$

$$+ \mathbb{E}_x \left[ \frac{d}{dh} |_{h=0} \exp \left( \int_0^t V(X_s) \, ds \right) \left( e^{hA} g \right)(X_t) \right]$$

$$= \mathbb{E}_x \left[ \exp \left( \int_0^t V(X_s) \, ds \right) V(X_t) g(X_t) \right]$$

$$+ \mathbb{E}_x \left[ \exp \left( \int_0^t V(X_s) \, ds \right) (Ag)(X_t) \right]$$

$$= T_t (Vg)(x) + T_t (Ag)(x) \tag{20.33}$$

which gives Eq. (20.30). [It should be clear that $(T_0g)(x) = g(x)$.]

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 $h > 0$,

$$(T_{t+h} g)(x) = \mathbb{E}_x \left[ Z_{t+h} g (X_{t+h}) \right] = \mathbb{E}_x \left[ Z_t (Z_{t+h} g)(X_{t+h}) \right]$$

$$= \mathbb{E}_x \left[ Z_t g (X_{t+h}) + Z_t \left( Z_{t+h} - 1 \right) g (X_{t+h}) \right]$$

$$= \mathbb{E}_x \left[ Z_t (e^{hA} g)(X_t) \right] + \mathbb{E}_x \left[ Z_t \left( Z_{t+h} - 1 \right) g (X_{t+h}) \right].$$

Therefore,

$$\frac{(T_{t+h} g)(x) - (T_t g)(x)}{h} = \mathbb{E}_x \left[ Z_t \left( e^{hA} g \right)(X_t) - g(X_t) \right]$$

$$+ \mathbb{E}_x \left[ Z_t \left( Z_{t+h} - 1 \right) g (X_{t+h}) \right]$$

and then letting $h \downarrow 0$ in this equation implies,

$$\frac{d}{dh} |_{h=0} (T_{t+h} g)(x) = \mathbb{E}_x \left[ Z_t (Ag)(X_t) \right] + \mathbb{E}_x \left[ Z_t V(X_t) g (X_{t+h}) \right].$$

This shows that $T_t$ is one sided differentiable and this one sided derivatives is given as in Eq. (20.30).

On the other hand for $s, t > 0$, using Theorem 20.14

$$T_{t+s} g (x) = \mathbb{E}_x \left[ Z_{t+s} g (X_{t+s}) \right] = \mathbb{E}_x \left[ Z_t Z_{t+s} g (X_{t+s}) \right]$$

$$= \mathbb{E}_x \left[ Z_t \mathbb{E}_{Z_t} (Z_{t+s} g (X_{t+s})) \right] = \mathbb{E}_x \left[ Z_t \mathbb{E}_{X_t} (Z_{0,s} g (X_s)) \right]$$

$$= (T_t T_s g)(x),$$

i.e. $\{T_t\}_{t \geq 0}$ 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

$$\frac{T_{t-h} - T_t}{-h} = T_{t-h} \frac{T_h - I}{h} \to T_t (A + V) \quad \text{as} \quad h \downarrow 0$$

using $T_t$ is continuous in $t$ and the result we have already proved. This shows $T_t$ is differentiable in $t$ and Eq. (20.30) is valid.

Example 20.22 (Poisson Process). By Exercise 20.2 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,

$$P(N_t = y | N_s = x) = P(N_s + N_t - N_s = y| N_s = x)$$

$$= P(N_s + N_t - N_s = y| N_s = x)$$

$$= P(N_t - N_s = y - x| N_s = x)$$

$$= 1_{y \geq x} \frac{(\lambda (t-s))^{y-x}}{(y-x)!} e^{-\lambda (t-s)} =: q_{t-s} (x, y).$$
With this notation it follows that

\[ P (f (N_t) | N_s) = (Q_{t-s} f) (N_s) \]

where

\[ Q_t f (x) = \sum_{y \in S} q_t (x, y) f (y) = \sum_{y \in S} 1_{y \geq x} \left( \frac{\lambda t}{(y-x)^{n-1}} \right) e^{-\lambda t} f (y) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} f (x+n). \]

In particular \( \{N_t\}_{t \geq 0} \) is a time homogeneous Markov process. It is easy (but technically unnecessary) to directly verify the semi-group property;

\[ (q_t q_s) (x, z) := \sum_{y \in S} q_t (x, y) q_s (y, z) = q_{s+t} (x, z). \]

This can done using the binomial theorem as follows;

\[
\begin{align*}
\sum_{y \in S} q_t (x, y) q_s (y, z) &= \sum_{y \in S} 1_{y \geq x} \left( \frac{\lambda t}{(y-x)^{n-1}} \right) e^{-\lambda t} \cdot 1_{z \geq y} \left( \frac{\lambda s}{(z-y)^{n-1}} \right) e^{-\lambda s} \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \cdot 1_{z \geq x+n} \frac{(\lambda s)^n}{n!} e^{-\lambda s} \\
&= 1_{z \geq x} e^{-\lambda(t+s)} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \frac{(\lambda s)^n}{n!} (z-x-n)! \\
&= 1_{z \geq x} e^{-\lambda(t+s)} \lambda(t+s)^{z-x} (z-x)! \\
&= q_{s+t} (x, z).
\end{align*}
\]

To identify infinitesimal generator, \( A = \frac{d}{dt}|_{t=0} Q_t \), in this example observe that

\[
\begin{align*}
\frac{d}{dt} Q_t f (x) &= \frac{d}{dt} \left[ e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f (x+n) \right] \\
&= -\lambda Q_t f (x) + \lambda e^{-\lambda t} \sum_{n=0}^{\infty} \frac{n (\lambda t)^{n-1}}{n!} f (x+n) \\
&= -\lambda Q_t f (x) + \lambda e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f (x+n+1) \\
&= -\lambda Q_t f (x) + \lambda (Q_t f) (x+1) \\
&= Q_t [\lambda f (\cdot + 1) - \lambda f (\cdot)] (x)
\end{align*}
\]

and hence

\[ Af (x) = \lambda f (x+1) - f (x). \]

Finally let us try to solve Eq. \((20.27)\) in order to recover \( Q_t \) from \( A \). Formally we can hope that \( Q_t = e^{tA} \) where \( e^A \) is given as its power series expansion. To simplify the computation it convenient to write \( A = \lambda (T-I) \) where \( I f = f \) and \( T f = f (\cdot + 1) \). Since \( I \) and \( T \) commute we further expect

\[ e^{tA} = e^{\lambda (T-I)} = e^{-\lambda t I} e^{\lambda T} = e^{-\lambda t} e^{\lambda T} \]

where

\[ (e^{\lambda T} f) (x) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} (T^n f) (x) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f (x+n). \]

Putting this all together we find

\[ (e^{tA} f) (x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} f (x+n) \]

which is indeed in agreement with \( Q_t f (x) \) as we saw in Eq. \((20.32)\).

**Remark 20.23 (Convolution Semi-Groups).** Here is an alternative explanation of Eq. \((20.33)\). 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

\[ \mu_{\ast} \ast \mu_s (z) = \sum_{n=0}^{\infty} z^n \mu_t (\{n\}) \sum_{k=0}^{\infty} \frac{(\lambda t)^{n-k}}{n!} \mu_s (k) = \sum_{0 \leq k \leq \infty} z^{n-k} \mu_t (n-k) \mu_s (k) \]

\[ = \sum_{0 \leq k \leq \infty} z^{n-k} k^{k} \mu_s (k) = \sum_{0 \leq k \leq \infty} \tilde{\mu}_t (z) z^k \mu_s (k) = \tilde{\mu}_s (z) \tilde{\mu}_t (z) = e^{\lambda t(1-z)} e^{\lambda s(1-z)} = e^{\lambda (s+t)(1-z)} = \tilde{\mu}_{s+t} (z) \]

from which it follows that \( \{\mu_t\}_{t \geq 0} \) is a convolution semi-group, i.e. \( \mu_t \ast \mu_s = \mu_{s+t} \) for all \( s, t > 0 \). Using
Exercise 20.10 (Brownian Motion). Assuming a Brownian motion is called a filtered probability space. A real valued adapted process, Definition 20.24 (Brownian Motion). Let it follows that

\[(q_t f) (x) = \sum_{n=0}^{\infty} \mu_t (n) f (x + n),\]

it follows that

\[(q_s q_t f) (x) = \sum_{k \in \mathbb{Z}} \mu_s (k) \sum_{n \in \mathbb{Z}} \mu_t (n) f (x + k + n)\]

\[= \sum_{k \in \mathbb{Z}} \mu_s (k - n) \sum_{n \in \mathbb{Z}} \mu_t (n) f (x + k)\]

\[= \sum_{k \in \mathbb{Z}} (\mu_s * \mu_t) (k) f (x + k) = \sum_{k} \mu_{s+t} (k) f (x + k)\]

\[= (q_{s+t} f) (x).\]

Definition 20.24 (Brownian Motion). Let \(\Omega, B, \{B_t\}_{t \in \mathbb{R}_+}, P\) 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 independent of \(B_s\) for all \(0 \leq s < t < \infty\).
2. For \(0 \leq s < t\), \(B_t - B_s \overset{d}{=} N (0, t-s)\), 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 20.10 (Brownian Motion). Assuming a Brownian motion \(\{B_t\}_{t \geq 0}\) exists as described in Definition 20.24 show:

1. The process is a time homogeneous Markov process with transition kernels given by:

\[Q_t (x, dy) = q_t (x, y) dy\]

where

\[q_t (x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2} |y-x|^2}.\]

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 is 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,

\[\frac{d}{dt} q_t (x, y) = \frac{1}{2} \frac{d^2}{dx^2} q_t (x, y) = \frac{1}{2} \frac{d^2}{dy^2} q_t (x, y)\]

for \(t > 0\).

4. Suppose that \(f : \mathbb{R} \to \mathbb{R}\) is a twice continuously differentiable function with compact support. Show

\[\frac{d}{dt} Q_t f = AQ_t f = Q_t (Af) \quad \text{for all} \ t > 0,\]

where

\[Af (x) = \frac{1}{2} f'' (x).\]

By combining Exercise 20.10 with Theorem 20.11 proves the following corollary.

Corollary 20.25. There exits a Markov process \(\{B_t\}_{t \geq 0}\) satisfying properties 1. and 2. of Definition 20.23.

To get the path continuity property of Brownian motion requires additional arguments which we will do in a number of ways later, see Theorems 29.3 and ??, Modulo technical details, Exercise 20.10 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. (20.34) and (20.35). 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 20.26. 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 Lévy process we must have \(Q_t (f) (x) = (\mu_s * f) (x)\) where \(\{\mu_t\}_{t \geq 0}\) is a one parameter convolution semi-group which is necessarily satisfies, \(\hat{\mu} (\lambda) = e^{\psi (\lambda)}\), where \(\psi (\lambda)\) is a Levy exponent as described in Eq. (26.4) of Theorem 26.7 below. In more detail \(\psi (\lambda)\) is any one of the functions of the form,

\[\psi (\lambda) = i \lambda b - \frac{1}{2} a \lambda^2 + \int_{\mathbb{R} \setminus \{0\}} \left( e^{i \lambda x} - 1 - i \lambda x \cdot 1_{|x| \leq 1} \right) d \nu (x),\]

where \(b \in \mathbb{R}, a \geq 0,\) and \(\nu\) is a measure on \(\mathbb{R} \setminus \{0\}\) such that \(\int_{\mathbb{R} \setminus \{0\}} (x^2 + 1) d \nu (x) < \infty\).

Example 20.27. If \(\{N_t\}_{t \geq 0}\) is a Poisson process and \(\{B_t\}_{t \geq 0}\) is a Brownian motion which is independent of \(\{N_t\}_{t \geq 0}\), 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;
\[
(Q_t f) (x) = E [ f (x + N_t + B_t) ] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2} y^2} E[f(x+y+N_t)] \, dy
\]
\[
= e^{-\frac{t}{2}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2} y^2} \frac{(\lambda t)^n}{n!} f(x+y+n) \, dy
\]
\[
= e^{-\frac{t}{2}} \sum_{n=0}^{\infty} \int_{\mathbb{R}} e^{-\frac{1}{2} y^2} \frac{(\lambda t)^n}{n!} f(x+y) \, dy
\]
\[
= \int_{\mathbb{R}} q_t (y-x) f(y) \, dy
\]
where
\[
q_t (y) := e^{-\frac{t}{2}} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\frac{1}{2} y^2}.
\]
The infinitesimal generator, \( A = \frac{d}{dx} |_{x=0} Q_t \) of this process satisfies,
\[
(A f) (x) = \frac{1}{2} f''(x) + \lambda (f(x+1) - f(x))
\]
for all \( f \in C^2 (\mathbb{R}) \). This example will be significantly generalized in Theorem 20.30 below.

In order to continue with giving examples in the continuous time case we will need a simple measure theoretic result.

**Lemma 20.28.** If \( \Omega \) is a set, \( \Omega_0 \subseteq \Omega \), and \( B_0 \) is a \( \sigma \)-algebra on \( \Omega_0 \), then \( B_0 : = \{ A \subseteq \Omega : A \cap \Omega_0 \in B_0 \} \) is a \( \sigma \)-algebra on \( \Omega \). Moreover, \( f : \Omega \to \mathbb{R} \) is \( B_0 \)-measurable iff \( f|_{\Omega_0} \) is \( B_0 \)-measurable.

**Proof.** It is clear that \( \emptyset, \Omega \in B_0 \) and that \( B_0 \) is closed under countable unions since \( A_n \in B_0 \) if \( \bigcup A_n \in B_0 \), which implies \( \bigcup A_n \cap \Omega_0 = \bigcup [A_n \cap \Omega_0] \in B_0 \) and this implies that \( \bigcup A_n \in B_0 \). Lastly if \( A \in B_0 \) then \( A \cap \Omega_0 \in B_0 \) implies that
\[
A' \cap \Omega_0 = \Omega_0 \setminus A = \Omega_0 \setminus [A \cap \Omega_0] \in B_0
\]
and therefore \( A \in B_0 \).

For the second assertion, let us observe that for \( W \in B_0 \) we have
\[
f^{-1} (W) \cap \Omega_0 = f|_{\Omega_0}^{-1} (W)
\]
so that \( f^{-1} (W) \cap \Omega_0 \in B_0 \) if \( f^{-1} (W) \cap \Omega_0 \in B_0 \) iff \( f|_{\Omega_0}^{-1} (W) \in B_0 \). It now clearly follows that \( f : \Omega \to \mathbb{R} \) is \( B_0 \)-measurable iff \( f|_{\Omega_0} \) is \( B_0 \)-measurable.

---

**Definition 20.29.** Suppose that \( \Omega = \bigcup_{n=0}^{\infty} \Omega_n \) and \( B_n \) is a \( \sigma \)-algebra on \( \Omega_n \) for all \( n \). Then we let \( \bigcup_{n=0}^{\infty} B_n =: B \) be the \( \sigma \)-algebra on \( \Omega \) such that \( A \subseteq \Omega \) is measurable iff \( A \cap \Omega_n \in B_n \) for all \( n \). That is to say \( B = \bigcap_{n=0}^{\infty} B_n \) with \( B_n = \{ A \subseteq \Omega : A \cap \Omega_n \in B_n \} \).

From Lemma 20.28 it follows that \( f : \Omega \to \mathbb{R} \) is \( \bigcup_{n=0}^{\infty} B_n \)-measurable iff \( f^{-1} (W) \in B_n \) for all \( n \) iff \( f|_{\Omega_n}^{-1} (W) \in B_n \) for all \( n \) iff \( f|_{\Omega_n}^{-1} (W) \in B_n \) 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 20.32 (The reader should compare the following result with Theorem 26.39 below.)

**Theorem 20.30 (Compound Poisson Process).** Suppose that \( \{ Z_i \}_{i=1}^{\infty} \) are i.i.d. random vectors in \( \mathbb{R}^d \) and \( \{ N_i \}_{i=0}^{\infty} \) be an independent Poisson process with intensity \( \lambda \). Further let \( Z_0 : \Omega \to \mathbb{R}^d \) be independent of \( \{ N_i \}_{i=0}^{\infty} \) and the \( \{ Z_i \}_{i=1}^{\infty}(N_{i-1}) \) and then define, for \( t \in \mathbb{R}^+ \),
\[
B_t = \bigcup_{n=0}^{\infty} [ \sigma (N_s : s \leq t, Z_0, \ldots, Z_n)]_{(N_{i,n})}
\]
and \( X_t := S_{N_t} \) where \( S_n = Z_0 + Z_1 + \cdots + Z_n \). Then \( \{ B_t \}_{t \geq 0} \) is a filtration (i.e. it is increasing), \( \{ X_t \}_{t \geq 0} \) is a \( B_t \)-adapted process such that for all \( 0 \leq s < t \), \( X_t - X_s \) is independent of \( B_s \). The increments are stationary and therefore \( \{ X_t \}_{t \geq 0} \) is a Lévy process. The time homogeneous transition kernel is given by
\[
(Q_t f) (x) = E [ f (x + Z_1 + \cdots + Z_{N_t}) ]
\]
\[
= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} E[f(x+Z_1+\cdots+Z_n)].
\]
If we define \( (Q_t f) (x) := E [ f (x + Z_1) ] \), the above equation may be written as,
\[
Q_t = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} Q^n = e^{\lambda t (d-1)}.
\]

**Proof.** Let us begin by showing that \( B_t \) is increasing. First observe that
\[
[ \sigma (N_s : s \leq t, Z_0, \ldots, Z_n)]_{(N_{i,n})} = \{ A \subseteq \{ N_t = n \} : A \in \sigma (N_s : s \leq t, Z_0, \ldots, Z_n) \}.
\]
If \( 0 \leq s < t \) and \( A \in B_s \), then \( A \cap \{ N_s = m \} \in \sigma (N_r : r \leq s, Z_0, \ldots, Z_m) \) therefore, for \( n \geq m \), we have
\[
A \cap \{ N_s = m \} \cap \{ N_t = n \} \in \sigma (N_r : r \leq t, Z_0, \ldots, Z_n) \cap \sigma (N_r : r \leq t, Z_0, \ldots, Z_n).
\]
and we may conclude that
\[ A \cap \{ N_t = n \} = \cup_{m \leq n} [ A \cap \{ N_s = m \} \cap \{ N_t = n \} ] \in \sigma ( N_r : r \leq t, Z_0, \ldots, Z_n ) \]
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 \( s \leq t \). Since
\[ X_s |_{N_s = n} = [ Z_0 + Z_1 + \cdots + Z_n ] |_{N_s = n} \]
is \( \sigma ( N_r : s \leq t, Z_0, \ldots, Z_n ) |_{N_s = n} \) measurable for all \( n \in \mathbb{N}_0 \), it follows by Lemma 20.28 that \( X_t \) is \( \mathcal{B}_t \)-measurable.

We now show that \( X_t - X_s \) is independent of \( \mathcal{B}_s \) for all \( t \geq s \). 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
\[ g |_{N_s = n} = G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) \]
while
\[ (X_t - X_s) |_{N_s = n} = Z_{n+1} + \cdots + Z_N \]
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,
\[ a_{m,n} = \mathbb{E} \left[ f (Z_{n+1} + \cdots + Z_{n+m}) \cdot G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) : N_s = n, N_t - N_s = m \right] \]
\[ = \mathbb{E} \left[ f (Z_{n+1} + \cdots + Z_{n+m}) 1_{N_s = n} \cdot G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) 1_{N_s = n} \right] \]
\[ = \mathbb{E} \left[ f (Z_{n+1} + \cdots + Z_{n+m}) 1_{N_s = n} \cdot \mathbb{E} \left[ G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) 1_{N_s = n} \right] \right] \]
\[ = e^{-\lambda (t-s)} \frac{(\lambda (t-s))^m}{m!} \mathbb{E} [ f (Z_1 + \cdots + Z_m) ] \cdot \mathbb{E} [ G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) 1_{N_s = n} ] \]
Therefore it follows that
\[ \mathbb{E} [ f (X_t - X_s) \cdot g] = \sum_{m,n=0}^{\infty} a_{m,n} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{m,n} \]
\[ = (Q_{t-s} f) (0) \cdot \sum_{n=0}^{\infty} \mathbb{E} \left[ G_n \left( \{ N_r \}_{r \leq s} , Z_0, \ldots, Z_n \right) 1_{N_s = n} \right] \]
\[ = (Q_{t-s} f) (0) \cdot \mathbb{E} [ g] \]
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 20.2 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 20.31.** If \( Q_t f \) is as in Theorem 20.30, then (formally),
\[ \frac{d}{dt} Q_t = \frac{d}{dt} e^{\lambda (\hat{Q} - I)} = \lambda (\hat{Q} - I) Q_t = Q_t \lambda (\hat{Q} - I) \]
where
\[ \lambda (\hat{Q} - I) f (x) = \lambda \mathbb{E} [ f (x + Z_1) - f (x) ] = \lambda \int_{\mathbb{R}^d} [ f (x + \xi) - f (x) ] d\mu (\xi) \]
and \( \mu = \text{Law} (Z_1) \). This a process which jumps at rate \( \lambda \) from position \( x \) to position \( x + \xi \) with distribution \( \mu \).

### 20.4 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 20.32.** Suppose that \( \left( \Omega, \left\{ \mathcal{B}_n \right\}_{n \in \mathbb{N}_0}, B, P, \{ Y_n : \Omega \to S \}_{n \in \mathbb{N}_0} \right) \) is a time homogeneous Markov chain with one step transition kernel, \( \tilde{Q} \). Further suppose that \( \left\{ N_t \right\}_{t \geq 0} \) is a Poisson process with parameter \( \lambda \) which is independent of \( \tilde{B}_\infty := \vee_{n \in \mathbb{N}_0} \tilde{B}_n \). Let \( \mathcal{B}_s \) be the \( \sigma \) - algebra on \( \Omega \) such that
\[ [\mathcal{B}_s]_{N_s = n} := \sigma \left( \left\{ N_r : r \leq s \not\in \tilde{B}_n \right\} \right) \]
To be more explicit, \( A \subset \Omega \) is in \( \mathcal{B}_s \) iff
\[ A \cap \{ N_s = n \} \in \sigma \left( \left\{ N_r : r \leq s \not\in \tilde{B}_n \right\} \right) \]
for all \( n \in \mathbb{N}_0 \).

Alternatively, we may describe \( \mathcal{B}_s \) as the \( \sigma \) - algebra generated by sets of the form
\[ A = B \cap C \ni B := \bigcap_{k=1}^{n} \{ N_k = k \} \mbox{ and } C \in \tilde{B}_k \]
(20.36)
where \( n \in \mathbb{N}, 0 = s_0 < s_1 < \cdots < s_n = s, k_i \in \mathbb{N} \) are such that \( k_1 \leq k_2 \leq \cdots \leq k_n \).

If \( X_t := Y_{N_t} \) for \( t \in \mathbb{R}_+ \), then \( \{B_t\}_{t \in \mathbb{R}_+} \) is a filtration, \( \{X_t\}_{t \geq 0} \) is adapted to this filtration and is a time homogeneous Markov process with transition semigroup given by

\[
Q_t = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \hat{Q}^n = e^{-\lambda t} e^{\lambda \hat{Q}} = e^{\lambda (\hat{Q} - I)}. \tag{20.37}
\]

**Proof.** Let us begin by showing that \( B_t \) is increasing. If \( 0 \leq s < t \) and \( \mathcal{A} \in \mathcal{B}_s \) then \( A \cap \{N_s = m\} \in \sigma(N_r : r \leq s \& \hat{B}_m) \) therefore, for \( n \geq m \), we have

\[
A \cap \{N_s = m\} \cap \{N_t = n\} \in \sigma(N_r : r \leq t \& \hat{B}_m) \subset \sigma(N_r : r \leq t \& \hat{B}_n)
\]

and therefore,

\[
A \cap \{N_t = n\} = \bigcup_{m \leq n} [A \cap \{N_s = m\} \cap \{N_t = n\}] \in \sigma(N_r : r \leq t \& \hat{B}_n)
\]

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 \( s \leq t \).

\( X_t|_{N_s = n} = Y_{N_t} = \mathcal{Y}_{N_t} \cap \mathcal{N}_{N_s = n} \) is \( \sigma(N_r : r \leq t \& \hat{B}_n) \) measurable for all \( n \in \mathbb{N}_0 \), it follows by Lemma 20.28 that \( X_t \) is \( \mathcal{B}_t \) measurable.

We now need to show that Markov property. To this end let \( t \geq s, f \in \mathcal{S}_0 \) and \( g \in \mathcal{B}_s \). Then for each \( n \in \mathbb{N}_0 \), we have \( g|_{N_s = n} \in \sigma(N_r : r \leq s \& \hat{B}_n) \) and therefore,

\[
\mathbb{E}[f(X_t)|g] = \sum_{m,n=0}^\infty \mathbb{E}[f(X_t) : N_s = n, N_t - N_s = m]
\]

\[
= \sum_{m,n=0}^\infty \mathbb{E}[f(Y_{n+m}) : 1_{N_t - N_s = m} \cdot g : 1_{N_s = n}]
\]

\[
= \sum_{m,n=0}^\infty P(N_t - N_s = m) \mathbb{E}[f(Y_{n+m}) : g \cdot 1_{N_s = n}]
\]

\[
= \sum_{m,n=0}^\infty e^{-\lambda(t-s)} (\frac{\lambda(t-s)}{m!})^m \mathbb{E}[(\hat{Q}^m f)(Y_n) : g \cdot 1_{N_s = n}]
\]

\[
= \sum_{n=0}^\infty \mathbb{E}[(Q_{t-s} f)(Y_s) : g \cdot 1_{N_s = n}] = \mathbb{E}[(Q_{t-s} f)(Y_s) : g]
\]

\[
= \mathbb{E}[(Q_{t-s} f)(X_s) : g].
\]

This shows that

\[
\mathbb{E}[f(X_t)|g] = (Q_{t-s} f)(X_s) \quad P \text{- a.s.}
\tag{20.38}
\]

which completes the proof.

**Second proof using Eq. (20.36).** If \( A = B \cap C \in \mathcal{B}_s \) be as in Eq. (20.36), \( f \in \mathcal{S}_0 \), and \( t > s \), then

\[
\mathbb{E}[f(X_t) : A] = \sum_{m \geq k_n} \mathbb{E}[f(Y_m) : A, N_t = m]
\]

\[
= \sum_{m=0}^\infty \mathbb{E}[f(Y_{k_n+m}) : B \cap C \cap \{N_t - N_s = m\}]
\]

\[
= \sum_{m=0}^\infty \mathbb{E}[f(Y_{k_n+m}) : C] P(B) P(N_t - N_s = m)
\]

\[
= \sum_{m=0}^\infty \mathbb{E}[(\hat{Q}^m f)(Y_{k_n}) : C] P(B) e^{-\lambda(t-s)} (\frac{\lambda(t-s)}{m!})^m
\]

\[
= e^{-\lambda(t-s)} \mathbb{E} \left[ \sum_{m=0}^\infty (\frac{\lambda(t-s)}{m!})^m (\hat{Q}^m f)(X_s) : B \cap C \right]
\]

\[
= \mathbb{E}[(e^{\lambda(t-s)(\hat{Q} - I))} f)(X_s) : A].
\]

Since this equation holds for a multiplicative system of sets, \( A \) of the form in Eq. (20.36) which generate \( \mathcal{B}_s \), we may conclude again that Eq. (20.38) holds.

Suppose \( S \) is countable or finite set and \( a : S \times S \rightarrow \mathbb{R} \) is a function such that

\[
a(x, y) \geq 0 \quad \forall \ x \neq y \quad \text{and} \quad \sum_{y \in S} a(x, y) = 0 \quad \forall \ x \in S.
\tag{20.39}
\]

We further set and

\[
a_x := \sum_{y \neq x} a(x, y) = -a(x, x) \in [0, \infty) \quad \text{and} \quad \lambda := \sup_{x \in S} a_x.
\tag{20.40}
\]

Given a bounded function \( f \) on \( S \) we let

\[
Af(x) = \sum_{y \in S} a(x, y) f(y) = \sum_{y \neq x} a(x, y) [f(y) - f(x)] \quad \forall \ x \in S.
\]

Notice that

\[
|Af(x)| \leq \sum_{y \in S} |a(x, y)| |f(y)| \leq \sum_{y \in S} |a(x, y)||f|_{\infty} = 2a_x |f|_{\infty} \leq 2\lambda |f|_{\infty}.
\]
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 \( S_0 \) for all \( t \in \mathbb{R} \).

We would now like to apply Theorem 20.32 in order to construct an associated time homogeneous Markov process with generator \( A \). To do this we must find a Markov matrix, \( \hat{Q} \), on \( S \) such that \( A = \lambda \left( \hat{Q} - \mathbb{I} \right) \) for some \( \lambda > 0 \). If this is going to work we must take
\[
\hat{Q} := I + \frac{1}{\lambda} A.
\]
(20.41)

Notice that no matter how we choose \( \lambda \), \( \hat{Q} (x, y) \geq 0 \) for \( x \neq y \) and
\[
\sum_{y \in S} \hat{Q} (x, y) = \sum_{y \in S} \left[ \delta_x (y) + \frac{1}{\lambda} a(x, y) \right] = 1 + 0 = 1.
\]
So it only remains to check \( \hat{Q} (x, x) \geq 0 \) 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. (20.40) will do the trick provided it is finite. We summarize the result in the next corollary.

**Corollary 20.33.** Let \( S \) be a countable or finite set and \( a : S \times S \to \mathbb{R} \) be a function satisfying Eq. (20.39) and \( \lambda := \sup_{x \in S} a_x < \infty \), see Eq. (20.40). Let \( \hat{Q} := I + \frac{1}{\lambda} A \) be the Markov matrix as in Eq. (20.41) so that \( \hat{Q} \) has matrix elements,
\[
\hat{q}(x, y) := \begin{cases} 
\lambda^{-1} a(x, y) & \text{if } x \neq y \\
1 - \lambda^{-1} a_x & \text{if } x = y.
\end{cases}
\]
If \( \{Y_n\}_{n=0}^{\infty} \) is a Markov chain with one step transition kernel, \( \hat{Q} \) 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,
\[
Q_t = e^{tA} = e^{-t\lambda} e^{t\lambda \hat{Q}}.
\]
In particular, \( A \) is the infinitesimal generator of a Markov transition semigroup.

**Remark 20.34.** As in Exercise 20.9 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,
\[
e^{t\lambda \hat{Q}} (x, y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \hat{Q}^n (x, y) \geq 0
\]
and
\[
\sum_{y \in S} e^{t\lambda \hat{Q}} (x, y) = \sum_{y \in S} \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \hat{Q}^n (x, y) = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} 1 = e^{t\lambda}.
\]
Therefore \( Q_t (x, y) = e^{tA} (x, y) = e^{-t\lambda} e^{t\lambda \hat{Q}} (x, y) \geq 0 \) and
\[
\sum_{y \in S} Q_t (x, y) = \sum_{y \in S} e^{-t\lambda} e^{t\lambda \hat{Q}} (x, y) = e^{-t\lambda} e^{t\lambda} = 1.
\]

**Remark 20.35.** Let us pause to describe paths of the Markov process constructed in Corollary 20.33. To do this let \( \{T_k\}_{k=1}^{\infty} \) be i.i.d. exponential random variables with intensity \( \lambda \) such that \( \{T_k\}_{n=0}^{\infty} \) are independent of the chain \( \{Y_n\}_{n=0}^{\infty} \). We may then (see Definition 14.8) 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 \( \{J_k\}_{k=0}^{\infty} \) be the jump times of the lazy chain \( \{Y_n\}_{n=0}^{\infty} \) as described in Exercise 20.7. Then \( J_k := W_{J_k} \) is the time of the \( k \)-th jump of the \( \{X_t := Y_{N_t}\}_{t \geq 0} \), see Figure 20.2 below. We further let \( S_k := J_k - J_{k-1} \) for \( k \in \mathbb{N} \) which denote the sojourn time that the process \( \{X_t\}_{t \geq 0} \) “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 \( \{Y_k\}_{k=0}^{\infty} \) and the corresponding “sojourn” times \( \{S_k\}_{k=1}^{\infty} \). The next theorem describes the joint distributions \( \{Y_0\} \cup \{Y_k, S_k\}_{k=1}^{\infty} \).

**Theorem 20.36 (Jump-Hold Description II).** Let \( \{X_t\}_{t \geq 0} \) be the Markov process as constructed in Theorem 20.32 and Corollary 20.33. 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 20.35 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;
1. \( \{X_{j_k}\}_{k=0}^\infty \) is a discrete time Markov chain with Markov transition matrix, 
   \( \tilde{q}(x,y) \) defined by
   \[
   \tilde{q}(x,y) := \begin{cases} 
   a(x,y)/a_x & \text{if } y \neq x \\
   0 & \text{if } y = x
   \end{cases}. 
   \] (20.42)

2. If \( \{x_k\}_{k=0}^n \subset S \) with \( x_{k-1} \neq x_k \) for \( 1 \leq k \leq n \), then relative to \( P \left( \cap_{k=0}^n \{X_{j_k} = x_k\} \right) \) the sojourn times, \( \{S_k\}_{k=1}^n \), are independent exponential random variables with \( S_k = \exp \{a_{x_{k-1}}\} \).

**Proof.** Recall that \( \{Y_n\}_{n=0}^\infty \) was the Markov chain with transition kernels defined by (for some sufficiently large \( \lambda > 0 \))

\[
\tilde{q}(x,y) := \begin{cases} 
   \lambda^{-1}a(x,y) & \text{if } x \neq y \\
   1 - \lambda^{-1}a_x & \text{if } x = y
   \end{cases}. 
\]

We also continue the notation set up in Remark 20.35.

1. By Exercise 20.7, \( \{X_{j_k} = Y_{j_k}\}_{k=0}^\infty \) is a Markov chain with Markov transition kernel given by \( q(x,x) = 0 \) for all \( x \in S \) and for \( x \neq y \),
   \[
   \tilde{q}(x,y) := \frac{\tilde{q}(x,y)}{1-\tilde{q}(x,x)} = \frac{\lambda^{-1}a(x,y)}{\lambda^{-1}a_x} = \frac{a(x,y)}{a_x}.
   \]

2. For \( k \geq 1 \), let \( \sigma_k := j_k - j_{k-1} \) and given \( \{x_k\}_{k=0}^n \subset S \) with \( x_{k-1} \neq x_k \) for \( 1 \leq k \leq n \), let

\[
A := \cap_{k=0}^n \{X_{j_k} = x_k\} = \cap_{k=0}^n \{Y_{j_k} = x_k\}.
\]

According to Exercise 20.7, relative to the probability measure \( P(\cdot|A) = P(\{Y_{j_k} = x_k\}) \) the random \( N \)-valued times, \( \{\sigma_k\}_{k=1}^n \), are independent with

\[
\sigma_k \overset{\text{i.i.d.}}{\sim} \text{Geo}(1 - \tilde{q}(x_{k-1},x_{k-1}))
\]

for \( 1 \leq k \leq n \). Combining this result with Exercise 20.11 below shows \( \{S_k = W_{j_k} - W_{j_{k-1}}\}_{k=1}^n \) are are independent exponential random variables with

\[
S_k \overset{d}{=} \exp \{(1 - \tilde{q}(x_{k-1},x_{k-1}))\lambda\} \overset{d}{=} \exp \{a_{x_{k-1}}\}.
\]

**Exercise 20.11.** 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 \( \sigma_\ell \overset{\text{i.i.d.}}{\sim} \text{Geo}(b_\ell) \) 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

\[
W_0 = 0, \quad W_n := T_1 + \ldots + T_n,
\]

\[
j_0 = 0, \quad j_\ell := \sigma_1 + \cdots + \sigma_\ell,
\]

\[
S_\ell := W_{j_\ell} - W_{j_{\ell-1}} \quad \text{for } 1 \leq \ell \leq n.
\]

Show \( \{S_k\}_{k=1}^n \) are independent exponential random variables with \( S_\ell \overset{d}{=} \exp \{b_\ell \lambda\} \) for all \( 1 \leq \ell \leq n \).

### 20.5 First Step Analysis and Hitting Probabilities

In this section we suppose that \( T = \mathbb{N}_0 \), \( (\Omega,B,\{B_t\}_{t\in[0,\infty)}) \) is a filtered measures space, \( X_t : \Omega \to S \) is a \( B_t/S \) measurable function for all \( t \in T \), \( Q : S \times S \to [0,1] \) is a Markov-transition kernel, and for each \( x \in S \) there exists a probability, \( P_x \) on \( (\Omega,B) \) such that \( P_x(X_0 = x) = 1 \) and \( \{X_t\}_{t\geq0} \) is a time homogenous Markov process with \( Q \) as its one step Markov transition kernel. To shorten notation we will write \( E_x \) for the expectation relative to the measure \( P_x \).

**Definition 20.37 (Hitting times).** For \( B \subset S \), let

\[
T_B(X) := \min \{n \geq 0 : X_n \in B\}
\]

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 20.38** For \( A \subset S \), let \( Q_A : A \times S \to [0,1] \) be the restriction of \( Q \) to \( A \), so that \( Q_A(x,C) := Q(x,C) \) for all \( x \in A \) and \( C \subset S \). As with probability kernels we may identify \( Q_A \) with an operator from \( (S_A)_0 \) to itself via,

\[
(Q_Af)(x) = \int_A Q(x,dy) f(y) \quad \text{for all } x \in A \text{ and } f \in (S_A)_0.
\]
Then
\[ P_x(T_B = \infty) = \lim_{n \to \infty} (Q^n_A 1)(x) \ \forall \ x \in A \]  \hspace{1cm} (20.44)
and if \( \alpha < 1 \), then \( P_x(T_B = 0) = 0 \) for all \( x \in A \). [This lemma is generalized in Theorem 20.53 below.]

**Proof.** Let \( u(x) := P_x(T_B = \infty) \) for \( x \in A := B^c \), and observe that
\[ P_x(T_B > n) = P_x(X_1 \in A, \ldots, X_n \in A) = (Q^n_A 1)(x) . \]  \hspace{1cm} (20.45)
Letting \( n \uparrow \infty \) shows that
\[ u(x) := P_x(T_B = \infty) = \lim_{n \to \infty} (Q^n_A 1)(x) \]
which proves Eq. (20.44). Passing to the limit as \( n \to \infty \) in the identity,
\[ (Q^n_A 1)(x) = \int_A Q(x, dy)(Q^{-1} A 1)(y) \]
shows (using DCT) that
\[ u(x) = \lim_{n \to \infty} (Q^n_A 1)(x) = \int_A Q(x, dy) \lim_{n \to \infty} (Q^{-1} A 1)(y) = \int_A Q(x, dy) u(y) , \]
i.e. \( u = Q_A u \). Iterating the equation \( u = Q_A u \) then implies, \( u = Q_A^k u \). However this then implies
\[ u(x) = \int_A Q^n_A(x, dy) u(y) \leq \int_A Q^n_A(x, dy) \alpha = \alpha (Q^n_A 1)(x) . \]
Letting \( n \to \infty \) then shows \( u(x) \leq \alpha u(x) \) from which it follows that \( u(x) = 0 \) for \( \alpha < 1 \). □

**Exercise 20.12.** Suppose that \( T \) is a \( \mathbb{N}_0 \) – valued random variable and \( n \in \mathbb{N} \). Show
\[ \mathbb{E} T \leq n \sum_{k=0}^{\infty} P(T > nk) . \]  \hspace{1cm} (20.46)

**Lemma 20.40 (An \( \mathbb{E} T_B < \infty \) criteria).** Let \( B \in \mathcal{S} , A := B^c \), and for \( n \in \mathbb{N} \) let
\[ \alpha_n := \sup_{y \in A} P_y(T_B > n) . \]  \hspace{1cm} (20.47)
Then
\[ \mathbb{E} x T_B \leq \frac{n}{1 - \alpha_n} \ \forall \ x \in A \]  \hspace{1cm} (20.48)
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. (20.45), we have
\[ \alpha_n = \sup_{y \in A} P_y(T_B > n) = \sup_{y \in A} (Q^n_A 1)(y) . \]
For \( k \in \mathbb{N}_0 \),
\[ Q^n_A k = |Q^n_A|^{k-1} Q^n_A 1 \leq |Q^n_A|^{k-1} \alpha_n \]
\[ \leq |Q^n_A| k^{-2} \alpha_n^2 \leq \cdots \leq \alpha_n^k . \]
Therefore, using Eq. (20.46),
\[ \mathbb{E} x T_B \leq n \sum_{k=0}^{\infty} P_x(T_B > nk) = n \sum_{k=0}^{\infty} (Q^n_A 1)(x) \]
\[ \leq n \sum_{k=0}^{\infty} \alpha_n^k \leq \frac{n}{1 - \alpha_n} \ \forall \ x \in A . \]

**Theorem 20.41.** 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
\[ \mathbb{E} x [h(X_n) : T_B = n] = (Q^n_A - 1 Q[1_B h])(x) \]
and
\[ \mathbb{E} x [h(X_{T_B}) : T_B < \infty] = \left( \sum_{n=0}^{\infty} Q^n_A Q[1_B h] \right)(x) . \]  \hspace{1cm} (20.49)
If \( g : A \to \mathbb{R}_+ \) is a measurable function, then for all \( x \in A \) and \( n \in \mathbb{N}_0 \),
\[ \mathbb{E} x [g(X_n) 1_{n<T_B}] = (Q^n_A g)(x) . \]
In particular we have
\[ \mathbb{E} x \left[ \sum_{n<T_B} g(X_n) \right] = \sum_{n=0}^{\infty} (Q^n_A g)(x) =: u(x) , \]  \hspace{1cm} (20.50)
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;
\[ \{ T_B > n \} = \{ X_i \in A \text{ for } 0 \leq i \leq n \} \text{ and } \]
\[ \{ T_B = n \} = \{ X_i \in A \text{ for } 0 \leq i \leq n - 1 \} \cap \{ X_n \in B \} . \]
From the second identity above it follows that for
That these equations are valid is the content of Corollaries 20.44 and 20.45 and if

\[ u \sum_{n=1}^{\infty} \prod_{j=1}^{n} Q(x_{j-1}, dx_j) h(x_n) \]

Thus we would like to have another way to compute these quantities. Since

\[ 292 \quad \text{The Markov Property} \]

and therefore

\[ \mathbb{E}_x [h (X_n) : T_B = n] = \mathbb{E}_x \left[ h (X_n) : (X_1, \ldots, X_{n-1}) \in A^{n-1}, X_n \in B \right] \]

\[ = \sum_{n=1}^{\infty} \int_{A^{n-1} \times B} \prod_{j=1}^{n} Q(x_{j-1}, dx_j) h(x_n) \]

\[ = (Q_A^n - 1) Q [1_B h] (x) \]

Similarly,

\[ \mathbb{E}_x [g (X_n) 1_{n<T_B}] = \int_{A^n} Q(x, dx_1) Q(x_1, dx_2) \ldots Q(x_{n-1}, dx_n) g(x_n) \]

\[ = (Q_A^n g) (x) \]

and therefore,

\[ \mathbb{E}_x \left[ \sum_{n=0}^{\infty} g (X_n) 1_{n<T_B} \right] = \sum_{n=0}^{\infty} \mathbb{E}_x [g (X_n) 1_{n<T_B}] \]

\[ = \sum_{n=0}^{\infty} (Q_A^n g) (x). \]

In practice it is not so easy to sum the series in Eqs. (20.49) and (20.50). Thus we would like to have another way to compute these quantities. Since \( \sum_{n=0}^{\infty} Q^n_A \) is a geometric series, we expect that

\[ \sum_{n=0}^{\infty} Q^n_A = (I - Q_A)^{-1} \]

which is basically correct at least when \( (I - Q_A) \) is invertible. This suggests that if \( u (x) = \mathbb{E}_x [h (X_{T_B}) : T_B < \infty] \), then (see Eq. (20.49))

\[ u = Q_A u + Q [1_B h] \quad \text{on} \quad A, \quad (20.51) \]

and if \( u (x) = \mathbb{E}_x [\sum_{n<T_B} g (X_n)] \), then (see Eq. (20.50))

\[ u = Q_A u + g \quad \text{on} \quad A. \quad (20.52) \]

That these equations are valid is the content of Corollaries 20.44 and 20.45 below which we will prove using the “first step” analysis in the next theorem. We will give another direct proof in Theorem 20.52 below as well.

**Theorem 20.42 (First step analysis).** Let us keep the assumptions in Theorem 20.14 and add the further assumption that \( T = \mathbb{N}_0 \). Then for all \( F \in \mathcal{S}^{\mathbb{N}_0}_B \) or \( F : \mathbb{S}^{\mathbb{N}_0} \rightarrow [0, \infty] \) measurable;

\[ \mathbb{E}_x [F (X_0, X_1, \ldots)] = \int_S Q (x, dy) \mathbb{E}_y F (x, X_0, X_1, \ldots). \quad (20.53) \]

This equation can be iterated to show more generally that

\[ \mathbb{E}_x [F (X_0, X_1, \ldots)] = \int_S \prod_{j=1}^{n} Q(x_{j-1}, dx_j) \mathbb{E}_{x_n} [F (x_0, x_1, \ldots, x_{n-1}, X_0, X_1, \ldots)] \]

\[ (20.54) \]

where \( x_0 := x \).

**Proof.** Since \( X_0 (\omega) = x \) for \( P_x \) - a.e. \( \omega \), we have \( F (X_0, X_1, \ldots) = F (x, X_1, X_2, \ldots) \) a.s. Therefore by Theorem 20.14 we know that

\[ \mathbb{E}_x [F (X_0, X_1, \ldots) | B_1] = \mathbb{E}_x [F (x, X_1, X_2, \ldots) | B_1] = \mathbb{E}_x F (x, X_0, X_1, \ldots). \]

Taking expectations of this equation shows,

\[ \mathbb{E}_x [F (X_0, X_1, \ldots)] = \mathbb{E}_x \left[ \mathbb{E}_X F (x, X_0, X_1, \ldots) \right] \]

\[ = \int_S Q (x, dy) \mathbb{E}_y F (x, X_0, X_1, \ldots). \]

**Remark 20.43.** Let \( T_B (x_0, x_1, x_2, \ldots) \) be the first hitting time of \( B \) of the sequence \((x_0, x_1, x_2, \ldots)\) and when \( T_B (x_0, x_1, x_2, \ldots) < \infty \), let \( L_B (x_0, x_1, x_2, \ldots) = x_{T_B (x_0, x_1, x_2, \ldots)} \) be the hitting location. In more detail, \( T_B (x_0, x_1, x_2, \ldots) = n \) and \( L_B (x_0, x_1, x_2, \ldots) = x_n \) if \( x_n \in B \) while \( x_k \in A = B^c \) for \( 0 \leq k < n \). As usual we say \( T_B (x_0, x_1, x_2, \ldots) = \infty \) if \( x_k \in A \) for all \( k \) and in this case we leave \( L_B (x_0, x_1, x_2, \ldots) \) undefined. In using Theorem 20.42 we will often implicitly make use of the following two simple observations.

1. If \( x_0 \in A \) and \( T_B (x_0, x_1, x_2, \ldots) < \infty \) then

\[ L_B (x_0, x_1, x_2, \ldots) = L_B (x_1, x_2, \ldots). \]

2. If \( x_0 \in A \), then

\[ T_B (x_0, x_1, x_2, \ldots) = 1 + 1_A (x_1) T_B (x_1, x_2, \ldots). \]
3. More generally if \( g : A \to [0, \infty] \) is a function,
\[
F(x_0, x_1, x_2, \ldots) = \sum_{0 \leq n < T_B(x_0, x_1, x_2, \ldots)} g(x_n),
\]
and \( x_0 \in A \), then
\[
F(x_0, x_1, x_2, \ldots) = g(x_0) + 1_A(x_1)F(x_1, x_2, \ldots).
\]
[Taking \( g = 1 \) gives item 2.]

The reader might interpret \( F(x_0, x_1, x_2, \ldots) \) in Eq. (20.55) as money collected in \( A \) by a traveler on the path \((x_0, x_1, x_2, \ldots)\) up to its first exit from \( A \) where \( g(x) \) is the amount of money at location \( x \in A \).

**Corollary 20.44.** 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 \( A \) and
\[
u = Q_Au + Q_Bh \text{ on } A
\]
as in Eq. (20.51). 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,
\[
u = Q_Au + Q1_B \text{ on } A.
\]
**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
\[
F(X_0, X_1, \ldots) = \mathbb{E}_x[h(X_{T_B})] = h(X_{T_B})1_{T_B < \infty},
\]
then for \( x \in A \) we have \( F(x, X_0, X_1, \ldots) = F(X_0, X_1, \ldots) \), see Remark 20.43. Therefore by the first step analysis (Theorem 20.42) we learn for \( x \in A \) that
\[
u(x) = \mathbb{E}_x[h(X_{T_B})] = \mathbb{E}_x F(x, X_1, \ldots) = \int_S Q(x, dy) \mathbb{E}_y F(x, X_0, X_1, \ldots)
\[
= \int_S Q(x, dy)E_yF(x, X_0, X_1, \ldots) = \int_S Q(x, dy)u(y)
\[
= \int_A Q(x, dy)u(y) + \int_B Q(x, dy)h(y),
\]
i.e.
\[
u = Qu = Q_Au + Q_Bh \text{ on } A.
\]

**Corollary 20.45.** Suppose that \( B \in \mathcal{S} \), \( A := B^c \in \mathcal{S} \), \( g : A \to [0, \infty] \) is a measurable function. Further let
\[
u(x) := \mathbb{E}_x\left[ \sum_{0 \leq n < T_B} g(X_n) \right] \text{ for } x \in S.
\]
Then \( u(x) = 0 \) if \( x \in B \) and \( u(x) \) satisfies Eq. (20.52), i.e.
\[
u = Qu + g = QAu + g \text{ on } A
\]
or in more detail,
\[
u(x) = \int_A Q(x, dy)u(y) + Qg(x) \text{ for all } x \in A.
\]
In particular if we take \( g \equiv 1 \) in this equation we learn that
\[
u = \mathbb{E}_x T_B = \int_A Q(x, dy)\mathbb{E}_y T_B + 1 \text{ for all } x \in A.
\]
**Proof.** Let
\[
F(X_0, X_1, \ldots) = \sum_{0 \leq n < T_B(X_0, X_1, \ldots)} g(X_n)
\]
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
\[
F(x, X_0, X_1, \ldots) = \begin{cases} g(x) & \text{if } X_0 \in B \\ g(x) + F(X_0, X_1, \ldots) & \text{if } X_0 \in A \\ g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \ldots) \end{cases}
\]
Therefore by the first step analysis (Theorem 20.42) it follows that
\[
u(x) = \mathbb{E}_x F(X_0, X_1, \ldots) = \int_S Q(x, dy)\mathbb{E}_y F(x, X_0, X_1, \ldots)
\[
= \int_S Q(x, dy)\mathbb{E}_y g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \ldots)
\[
= g(x) + \int_A Q(x, dy)\mathbb{E}_y F(X_0, X_1, \ldots)
\[
= g(x) + \int_A Q(x, dy)u(y).
\]
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 20.46.** 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

$$v(x) := \mathbb{E}_x \left[ 1_{T_B(x) < \infty} \cdot h(X_{T_B}) \right] = 1_A(x) \left( (I - QA)^{-1} Q_B h \right)(x) + 1_B(x) h(x)$$

and we have used $u = 0$ on $B$. Putting this all together shows,

$$u = g \left( Q_A (I - QA)^{-1} Q_B h + Q_B h \right) + Q_A u$$

or in other words,

$$u = (I - QA)^{-1} \left( g (I - QA)^{-1} Q_B h \right).$$

Alternatively, we may work from first principles,

$$u(x) = \mathbb{E}_x \left[ h(X_{T_B}) \sum_{0 \leq m < T_B} g(X_m) : T_B < \infty \right] = \sum_{n=1}^{\infty} \mathbb{E}_x \left[ h(X_n) \sum_{0 \leq m < n} g(X_m) : T_B = n \right]$$

and we have used $u = 0$ on $B$. Putting this all together shows,

$$u = g \left[ Q_A (I - QA)^{-1} Q_B h + Q_B h \right] + Q_A u$$

or in other words,

$$u = (I - QA)^{-1} \left[ g (I - QA)^{-1} Q_B h \right].$$

The problem with Corollaries 20.44 and 20.45 is that the solutions to Eqs. (20.51) and (20.52) may not be unique as we will see in the next examples. Theorem 20.5.2 below will explain when these ambiguities may occur and how to deal with them when they do.

**Example 20.47 (Biased random walks 0).** Let $p \in (1/2, 1)$ and consider the biased random walk $\{S_n\}_{n \geq 0}$ 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 $x \in \mathbb{Z}$. This is time homogeneous Markov chain with transition matrix,

$$Q(x, y) = p1_{y=x+1} + q1_{y=x-1}.$$  

For any conditions $X$ and $a$ Letting

$$p \frac{a}{p} + q = 1,$$

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

$$u(x) = pu(x+1) + qu(x-1) \text{ for } a < x < b \text{ with } u(a) = 0 \text{ and } u(b) = 1.$$  

From Exercise 20.15 below, we know that the general solution to Eq. (20.59) 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. (20.59) we know that one root is 1 as is easily verified. The other root is $\rho := q/p < 1$. Thus the general solution is of the form, $u(x) = a + \beta \rho^{x-a}$. 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

$$u(x) = \frac{a}{p} x + \frac{b}{q} \frac{a}{p} - q = \frac{a}{p} [q - 1 + p] = 0.$$  

For $x = b - 1$, we have $(Q_{A} u)(x) = q u(x - 1)$ while $(Q_{B} h)(x) = p h(x + 1) = p$ and so $(Q_{A} u)(x) + (Q_{B} h)(x) = q u(x - 1) + p u(x + 1)$ provided we remember that $u(x + 1) = h(b) = 1$ when $x = b - 1$.  

Indeed,

$$P_0(T_a = \infty) = \lim_{b \uparrow \infty} P_0(T_b < T_a) = \lim_{b \uparrow \infty} \frac{1 - \rho^{-a}}{1 - \rho^{(b-a)}} = 1 - \left(\frac{q}{p}\right)^{-a}.$$  

Next let us try to find $u(x) := E_x[T_{(a,b)}]$ which we know to be finite by Lemma 20.40 By the first step analysis we have,

$$u(x) = p[u(x+1) + q[u(x-1)+1] = pu(x+1) + qu(x-1)+1 \text{ for } a < x < b$$

$$u(a) = 0 = u(b).$$  

A particular solution to Eq. (20.58) may be found by trying $u(x) = cx$. Plugging this into Eq. (20.58) 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

$$0 = u(a) = \frac{a}{q-p} + \beta \left(\rho^{(a-b)} - 1\right) \implies \beta = \frac{a}{q-p} \frac{\rho^{(a-b)} - 1}{\rho^{(a-b)} - 1}$$

$$0 = u(b) = \frac{b}{q-p} + \alpha \left(\rho^{(b-a)} - 1\right) \implies \alpha = \frac{b}{q-p} \frac{\rho^{(b-a)} - 1}{\rho^{(b-a)} - 1}$$

and so we have shown,

$$E_xT_{(a,b)} = \frac{1}{p-q} \left[\frac{\rho^{(x-a)} - 1}{\rho^{(b-a)} - 1} + \frac{\rho^{(x-b)} - 1}{\rho^{(a-b)} - 1} + a - x\right].$$  

Letting $b \uparrow \infty$ shows $E_xT_a = \infty$ and letting $a \downarrow -\infty$ shows

$$E_xT_b = \frac{1}{p-q} \left[\rho^{(x-a)} - 1 \right] + \rho^{(x-b)} - 1 a - x \text{ for } x < b.$$  

Example 20.48 (Biased random walks I). Let $p \in (1/2, 1)$ and consider the biased random walk $\{S_n\}_{n \geq 0}$ on the $S = \mathbb{Z}$ where $S_n = X_0 + X_1 + \cdots + X_n$, $\{X_i\}_{i \geq 1}$ 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,

$$u(x) = pu(x+1) + qu(x-1) \text{ for } x \neq 0.$$  

(20.59)
From Exercise 20.15 below, we know that the general solution to Eq. (20.59) 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. (20.59) we know that one root is 1 as is easily verified. The other root 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 choices we will have \( u(x) = w_+(x) \) for \( x \geq 0 \) and \( u(x) = w_-(x) \) for \( x \leq 0 \) see Exercise 20.15 and Remark 20.50.

Observe that

\[ u(1) = a_+ + (1-a_+) (q/p) \quad \Rightarrow \quad a_+ = \frac{u(1) - (q/p)}{1 - (q/p)} \]

and

\[ u(-1) = a_- + (1-a_-) (p/q) \quad \Rightarrow \quad a_- = \frac{(p/q) - u(-1)}{(p/q) - 1}. \]

**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 20.44) we will still have that \( u(x) \) satisfies Eq. (20.59) 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. (20.59) 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

\[ P_x (T_0 < T_n) = \frac{(q/p)^n}{(q/p)^n - 1} + \left( 1 - \frac{(q/p)^n}{(q/p)^n - 1} \right) (q/p)^x \]

\[ = \frac{(q/p)^x - (q/p)^n}{(q/p)^n - 1} \to (q/p)^x \text{ as } n \to \infty \]

and therefore, since \( T_n \uparrow \infty \) \( P_x - \text{a.s. as } n \uparrow \infty, \)

\[ P_x (T_0 < \infty) = (q/p)^x \text{ for all } x > 0. \]

**Example 20.49 (Biased random walks II).** Continue the notation in Example 20.48. 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 \( x \neq 0 \),

\[ u(x) = p [1 + u(x + 1)] + q [1 + u(x - 1)] = pu(x + 1) + qu(x - 1) + 1 \]

(20.60)

with \( u(0) = 0 \). Notice \( u(x) = \infty \) is a solution to this equation while if \( u(0) < \infty \) for some \( a \neq 0 \) then Eq. (20.60) 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. (20.60) is therefore,

\[ u(x) = (q - p)^{-1} x + a + b (q/p)^x. \]

(20.61)

Using the boundary condition, \( u(0) = 0 \) allows us to conclude that \( a + b = 0 \) and therefore,

\[ u(x) = u_a (x) = (q - p)^{-1} x + a [1 - (q/p)^x]. \]

(20.62)

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. (20.60) is \( u(x) = \infty \) for \( x > 0 \) as we already mentioned.
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 $a \leq 0$, we may restrict our attention to $a \leq 0$. To work out which $a \leq 0$ is correct observe by MCT that

$$E_x T_0 = \lim_{n \to -\infty} E_x [T_n \wedge T_0] = \lim_{n \to -\infty} E_x [T_{(n,0)}].$$

So let $n \in \mathbb{Z}$ with $n < 0$ be fixed for the moment. By item 8. of Theorem 20.52, we may conclude that $u(x) := E_x [T_{(n,0)}] < \infty$ for all $n \leq x \leq 0$. Then by the first step analysis, $u(x)$ satisfies Eq. (20.60) 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. (20.62) implies,

$$0 = u_a (n) = (q-p)^{-1} n + a [1 - (q/p)^n]$$

so that

$$a = a_n = \frac{n}{(1 - (q/p)^n)} (p - q) \to 0 \text{ as } n \to -\infty.$$ 

Thus we conclude that

$$E_x T_0 = \lim_{n \to -\infty} E_x [T_n \wedge T_0] = \lim_{n \to -\infty} u_{a_n}(x)$$

$$= \frac{x}{q-p} = \frac{|x|}{p-q} \quad \text{for } x < 0.$$

Remark 20.50 (More on the boundary conditions). If we were to use Corollary 20.45 directly to derive Eq. (20.60) in the case that $u(x) := E_x [T_{(n,0)}] < \infty$ for all $0 \leq x \leq n$, we would find, for $x < 0$, that

$$u(x) = \sum_{y \in \{n,0\}} q(x,y) u(y) + 1$$

which implies that $u(x)$ satisfies Eq. (20.60) 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. (20.60) for $n < x \leq 0$ and agree at $0$ and $-1$ and therefore are equal for $n \leq x \leq 0$ and in particular $0 = u(n) = w(n)$. Thus correct boundary conditions on $w$ in order for $w = u$ are $w(0) = w(0) = 0$ as we have used above.

Observe from Eq. (20.60), we have for $x \neq 0$ that,

$$u(x - 1) = q^{-1} [u(x) - pu(x + 1) - 1].$$

From this equation it follows easily that $u(x)$ for $x \leq 0$ is determined by its values at $x = 0$ and $x = -1$.

Definition 20.51. 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 $\mathcal{A}/\mathbb{R}$-measurable for all $C \in \mathcal{A}$ and $\rho(x, \cdot) : A \to [0,1]$ is a measure for all $x \in A$.

As with probability kernels we will identify $\rho$ with the linear map, $\rho : \mathcal{A}_0 \to \mathcal{A}_0$ 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 $\|g\|_u := \sup_{x \in \mathcal{A}} |g(x)|$ for all $g \in \mathcal{A}_0$.

Theorem 20.52. 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 $\mathcal{A}_0$.

(When $\rho = Q_A$, $u_n(x) = P_x(T_B > n)$ and $u(x) = P(T_B = \infty)$ as $n \to \infty$.)

2. The function $u$ satisfies $\rho u = u$.

3. If $w \in \mathcal{A}_0$ and $\rho w = w$ then $|w| \leq \|g\|_u$. In particular the equation, $\rho w = w$, has a non-zero solution $w \in \mathcal{A}_0$ if $w \neq 0$.

4. If $u = 0$ and $g \in \mathcal{A}_0$, then there is at most one $w \in \mathcal{A}_0$ such that $w = \rho w + g$.

5. Let

$$U := \sum_{n=0}^{\infty} u_n = \sum_{n=0}^{\infty} \rho^n 1 : A \to [0,\infty]$$

and suppose that $U(x) < \infty$ for all $x \in A$. Then for each $g \in \mathcal{S}_A$,

$$w = \sum_{n=0}^{\infty} \rho^n g$$

is absolutely convergent,

$$|w| \leq \|g\|_u U,$$

i.e.

$$\rho(x,|w|) < \infty \text{ for all } x \in A, \text{ and } w \text{ solves } w = \rho w + g. \text{ Moreover if } v \text{ also solves } v = \rho v + g \text{ and } |v| \leq CU \text{ for some } C < \infty \text{ then } v = w.$$

Observe that when $\rho = Q_A$,

$$U(x) = \sum_{n=0}^{\infty} P_x(T_B > n) = \sum_{n=0}^{\infty} E_x (1_{T_B > n}) = E_x \left( \sum_{n=0}^{\infty} 1_{T_B > n} \right) = E_x [T_B].$$

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 $w \leq v$. Thus $w$ is the minimal non-negative solution to $v = \rho v + g$.}
7. If there exists $\alpha < 1$ such that $u \leq \alpha$ on $A$ then $u = 0$. (When $\rho = Q_A$, this state that $P_x(T_B = \infty) \leq \alpha$ 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 $u_n = \rho^n \leq \alpha$ on $A$, then there exists $C < \infty$ such that

$$u_k(x) = (\rho^k 1)(x) \leq C \beta^k$$

for all $x \in A$ and $k \in \mathbb{N}_0$ where $\beta := \alpha^{1/n} < 1$. In particular, $U \leq C (1 - \beta)^{-1}$ and $u = 0$ under this assumption.

(When $\rho = Q_A$ this assertion states; if $P_x(T_B > n) \leq \alpha$ for all $\alpha \in A$, then $P_x(T_B > k) \leq C \beta^k$ and $\mathbb{E}_x T_B \leq C (1 - \beta)^{-1}$ for all $k \in \mathbb{N}_0$.)

**Proof.** We will prove each item in turn.

1. First observe that $u_1(x) = \rho(x, A) \leq 1 = u_0(x)$ and therefore,

$$u_{n+1} = \rho^{n+1} u_1 = \rho^n u_1 \leq \rho^n 1 = u_n.$$

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 $w \in A_b$ with $\rho w = w$, then

$$|w| = |\rho^n w| \leq \rho^n |w| \leq \|w\|_{\infty} \rho^n 1 = \|w\|_{\infty} \cdot u.$$

Letting $n \to \infty$ shows that $|w| \leq \|w\|_{\infty} u$.

4. If $w_i \in A_b$ solves $w_i = \rho w_i + g$ for $i = 1, 2$ then $w := w_2 - w_1$ satisfies $w = \rho w$ and therefore $|w| \leq C u = 0$.

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

$$\rho^k U = \sum_{n=0}^{\infty} \rho^k u_n = \sum_{n=0}^{\infty} u_{n+k} = \sum_{n=k}^{\infty} u_n \leq U. \tag{20.66}$$

Since the tails of convergent series tend to zero it follows that $\lim_{k \to \infty} \rho^k U = 0$.

Now if $g \in S_b$, we have

$$\sum_{n=0}^{\infty} |\rho^n g| \leq \sum_{n=0}^{\infty} \rho^n |g| \leq \sum_{n=0}^{\infty} \rho^n \|g\|_{\infty} = \|g\|_{\infty} \cdot U < \infty \tag{20.67}$$

and therefore $\sum_{n=0}^{\infty} \rho^n g$ is absolutely convergent. Making use of Eqs. \eqref{20.66} and \eqref{20.67} we see that

$$\sum_{n=1}^{\infty} \rho |\rho^n g| \leq \|g\|_{\infty} \cdot \rho U \leq \|g\|_{\infty} U < \infty$$

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 $|v| \leq C U$ and $v = g + \rho v$, then $y := w - v$ solves $y = \rho y$ with $|y| \leq (C + \|g\|_{\infty}) U$. It follows that

$$|y| = |\rho^n y| \leq (C + \|g\|_{\infty}) \rho^n U \to 0 \text{ as } n \to \infty,$$

i.e. $0 = y = w - v$.

6. If $g \geq 0$ we may always define $w$ by Eq. \eqref{20.64} 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 $v \geq 0$ 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

$$v = \sum_{k=0}^{n} \rho^k g + \rho^{n+1} v \geq \sum_{k=0}^{n} \rho^k g.$$

Letting $n \to \infty$ in this last equation shows that $v \geq w$.

7. If $u \leq \alpha < 1$ on $A$, then by item 3. with $w = u$ we find that

$$u \leq \|u\|_{\infty} \cdot u \leq \alpha u$$

which clearly implies $u = 0$.

8. If $u_n \leq \alpha < 1$, then for any $m \in \mathbb{N}$ we have

$$u_{n+m} = \rho^m u_n \leq \alpha \rho^m 1 = \alpha u_m.$$

Taking $m = n$ in this inequality shows, $u_{(k+1)n} \leq \alpha u_{kn}$. Thus a simple induction argument shows $u_{kn} \leq \alpha^k$ for all $k \in \mathbb{N}_0$. For general $l \in \mathbb{N}_0$ we write $l = kn + r$ with $0 \leq r < n$. We then have,

$$u_l = u_{kn+r} \leq u_{kn} \leq \alpha^k = \alpha^{\frac{r}{n}} = C \alpha^{l/n}$$

where $C = \alpha^{-\frac{1}{n}}$. 


Corollary 20.53. If \( h : B \to [0, \infty) \) is measurable, then \( u(x) := E_x[ h(X_{T_B}) : T_B < \infty ] \) is the unique minimal non-negative solution to Eq. (20.57) while if \( g : A \to [0, \infty) \) is measurable, then \( u(x) := E_x[ \sum_{n<T_B} g(X_n)] \) is the unique minimal non-negative solution to Eq. (20.52).

Exercise 20.13. Keeping the notation of Example 20.48 and 20.49. Use Corollary 20.53 to show again that \( P_x(T_B < \infty) = (q/p)^n \) for all \( x > 0 \) and \( E_x T_B = x/(q-p) \) for \( x < 0 \). You should do so without making use of the extraneous hitting times, \( T_n \) for \( n \neq 0 \).

Corollary 20.54. If \( P_x(T_B = \infty) = 0 \) for all \( x \in A \) and \( h : B \to \mathbb{R} \) is a bounded measurable function, then \( u(x) := E_x[ h(X_{T_B})] \) is the unique solution to Eq. (20.57).

Corollary 20.55. 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 \( E_x T_B < \infty \) for all \( x \in A \).

Proof. We know that
\[
\lim_{n \to \infty} P_x(T_B > n) = P_x(T_B = \infty) \leq \alpha \quad \text{for all} \quad x \in A.
\]

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 20.52.

Definition 20.56 (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 20.14. Let \( x \in X \). Show:

a. for all \( n \in \mathbb{N}_0 \),
\[
P_x(\tau_x > n + 1) \leq \sum_{y \neq x} p(x,y) P_y(T_x > n).
\]

b. Use Eq. (20.68) 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. (20.68) on \( n \in \mathbb{N}_0 \) to show
\[
E_x[\tau_x] \leq P_x(\tau_x > 0) + \sum_{y \neq x} p(x,y) E_y[T_x].
\]

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. (20.69) combined with Corollary 20.55 shows that \( E_x[\tau_x] < \infty \).

20.6 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 \( x^{th} \) row and \( y^{th} \) 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 20.55 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 \( E_x T_B < \infty \) and \( P_x(T_B < \infty) = 1 \) for all \( x \in A \).

If we let \( Q_A = Q_{A,B} \) be the matrix with entries, \( Q_A = (q(x,y))_{x,y \in A} \) and \( I \) be the corresponding identity matrix, then \( (Q_A - I)^{-1} \) exists according to Theorem 20.52. 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
\[
Q = \begin{bmatrix} A & B \\ * & * \end{bmatrix} \begin{bmatrix} A \\ B' \end{bmatrix}
\]

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 20.57. 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:
The Markov Property

$$\mathbb{E}_x[h(X_{T_B})] = [(I - Q_A)^{-1} R]_{x}$$ (x) and

$$\mathbb{E}_x\left[ \sum_{n<T_B} g(X_n) \right] = [(I - Q_A)^{-1} g]_{x}.$$ (x).

Remark 20.58. 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 \( \mathbb{E}_x\left[ \sum_{0\leq n<T_B} g(X_n) \right] \) is the expected toll you have to pay before your first exit from \( A \) while \( \mathbb{E}_x[h(X_T)] \) 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

$$P_x(X_{T_B} = y) = \left[(I - Q_A)^{-1} R \delta_y \right]_{x,y}.$$

2. If \( y \in A \) and \( g = \delta_y \), then

$$\sum_{n<T_B} g(X_n) = \sum_{n<T_B} \delta_y(X_n) = \# \text{ visits to before hitting } B$$

and hence

$$\mathbb{E}_x(\# \text{ visits to before hitting } B) = \left[(I - Q_A)^{-1} \delta_y \right]_{x} = (I - Q_A)^{-1} \delta_y.$$ (x)

3. If \( g = 1 \), i.e. \( g(y) = 1 \) for all \( y \in A \), then \( \sum_{n<T_B} g(X_n) = T_B \) and we find,

$$\mathbb{E}_x T_B = \left[(I - Q_A)^{-1} 1 \right]_{x} = \sum_{y \in A} (I - Q_A)^{-1} \delta_y$$

where \( \mathbb{E}_x T_B \) is the expected hitting time of \( B \) when starting from \( x \).

Example 20.59. Let us continue the rat in the maze Exercise 20.5 and now suppose that room 3 contains food while room 7 contains a mouse trap.

$$\begin{pmatrix} 1 & 2 & 3 \text{ (food)} \\ 4 & 5 & 6 \text{ (trap)} \end{pmatrix}.$$ 

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 \( g \) in Eq. (??) leaves the matrix,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\
\end{pmatrix}$$

Deleting the 3 and 7 columns from this matrix gives

$$\begin{pmatrix} 1 & 2 & 4 & 5 & 6 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \\
\end{pmatrix}$$

and deleting the 1, 2, 4, 5, and 6 columns gives

$$\begin{pmatrix} 3 & 7 \\ 0 & 0 \\ 1/3 & 0 \\ 0 & 0 \\ 1/2 & 0 \\
\end{pmatrix}$$

and

$$R = Q_{A,B} = \begin{pmatrix} 1 & -1 & -1/3 & 0 & 0 \\ -1 & 1 & 0 & -1/3 & 0 \\ -1/3 & 0 & 1 & -1/3 & 0 \\ 0 & -1/3 & -1/3 & 1 & -1/3 \\ 0 & 0 & 0 & -1/2 & 1 \\
\end{pmatrix}.$$ 

Therefore,

$$I - Q_A = \begin{pmatrix} 1 & -1/2 & -1/2 & 0 & 0 \\ -1/2 & 1 & 0 & -1/2 & 0 \\ -1/2 & 0 & 1 & -1/2 & 0 \\ 0 & -1/2 & -1/2 & 1 & -1/2 \\ 0 & 0 & 0 & -1/2 & 1 \\
\end{pmatrix}.$$ 

and using a computer algebra package we find

$$(I - Q_A)^{-1} = \begin{pmatrix} 1 & 2 & 4 & 5 & 6 \\ 11/2 & 5 & 5 & 1 & 1 \\ 6/2 & 1/2 & 4 & 1 & 1 \\ 5/2 & 1/2 & 4 & 1 & 1 \\ 3/2 & 1/2 & 4 & 1 & 1 \\
\end{pmatrix}.$$ 

In particular we may conclude,
By the first step analysis we will have,

\[
\begin{bmatrix}
E_1 T \\
E_2 T \\
E_4 T \\
E_5 T \\
E_6 T
\end{bmatrix} = (I - Q_A)^{-1} 1 = \begin{bmatrix}
\frac{17}{18} \\
\frac{5}{18} \\
\frac{1}{6} \\
\frac{1}{6} \\
\frac{1}{6}
\end{bmatrix},
\]

and

\[
\begin{bmatrix}
P_1 (X_T = 3) & P_1 (X_T = 7) \\
P_2 (X_T = 3) & P_2 (X_T = 7) \\
P_4 (X_T = 3) & P_4 (X_T = 7) \\
P_5 (X_T = 3) & P_5 (X_T = 7) \\
P_6 (X_T = 3) & P_6 (X_T = 7)
\end{bmatrix} = (I - Q_A)^{-1} R = \begin{bmatrix}
\frac{3}{7} & \frac{4}{7} \\
\frac{4}{7} & \frac{3}{7} \\
\frac{5}{6} & \frac{1}{6} \\
\frac{1}{6} & \frac{5}{6}
\end{bmatrix}.
\]

Since the event of hitting 3 before 7 is the same as the event \(\{X_T = 3\}\), the desired hitting probabilities are

\[
\begin{bmatrix}
P_1 (X_T = 3) \\
P_2 (X_T = 3) \\
P_4 (X_T = 3) \\
P_5 (X_T = 3) \\
P_6 (X_T = 3)
\end{bmatrix} = \begin{bmatrix}
\frac{3}{7} \\
\frac{4}{7} \\
\frac{5}{6} \\
\frac{1}{6} \\
\frac{1}{6}
\end{bmatrix}.
\]

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,

\[
\begin{align*}
h_i &= \sum_j P_i (X_T = 3 | X_1 = j) P_i (X_1 = j) \\
&= \sum_j q(i,j) P_i (X_T = 3 | X_1 = j) \\
&= \sum_j q(i,j) P_j (X_T = 3) \\
&= \sum_j q(i,j) h_j
\end{align*}
\]

where \(h_3 = 1\) and \(h_7 = 0\). Looking at the jump diagram (Figure 20.3) we easily find

\[
\begin{align*}
h_1 &= \frac{1}{2} (h_2 + h_4) \\
h_2 &= \frac{1}{3} (h_1 + h_3 + h_5) = \frac{1}{3} (h_1 + 1 + h_5) \\
h_4 &= \frac{1}{3} (h_1 + h_5 + h_7) = \frac{1}{3} (h_1 + h_5) \\
h_5 &= \frac{1}{3} (h_2 + h_4 + h_6) \\
h_6 &= \frac{1}{2} (h_3 + h_5) = \frac{1}{2} (1 + h_5)
\end{align*}
\]

and the solutions to these equations are (as seen before) given by

\[
\begin{bmatrix}
h_1 = \frac{7}{12}, h_2 = \frac{3}{4}, h_4 = \frac{5}{12}, h_5 = \frac{2}{3}, h_6 = \frac{5}{6}
\end{bmatrix}.
\]

(20.70)

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,
We now claim that in fact
\begin{align*}
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
\end{align*}

and then solve to find,
\begin{equation}
\begin{bmatrix}
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}
\end{bmatrix}.
\tag{20.71}
\end{equation}

Notice that the sum of the hitting probabilities in Eqs. (20.70) and (20.71) add up to 1 as they should.

### 20.6.1 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
\begin{equation}
\tau_i := \min\{n \geq 1 : X_n = i\}
\tag{20.72}
\end{equation}

be the first passage time of the chain to site $i$.

**Proposition 20.60.** The Markov matrix $Q$ has an invariant distribution.

**Proof.** If $1 := [1 \ 1 \ \ldots \ 1]^T$, then $Q1 = 1$ from which it follows that
\[0 = \det (Q - I) = \det (Q^{T} - I).\]

Therefore there exists a non-zero row vector $\nu$ such that $Q^{T} \nu = \nu$ or equivalently that $\nu^{T} Q = \nu$. 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 $Q^n_{ij} > 0$. Alternatively put this implies that $P_i (T_j < \infty) = P_i (T_j < \infty) > 0$ for all $i \neq j$. By Corollary 20.55 we know that $E_i [\tau_i] = E_i T_j < \infty$ for all $i \neq j$ and it is not too hard to see that $E_i \tau_i < \infty$ also holds. The fact that $E_i \tau_i < \infty$ for all $i \in S$ will come out of the proof of the next proposition as well.

**Proposition 20.61.** If $Q$ is irreducible, then there is precisely one invariant distribution, $\pi$, which is given by $\pi_i = 1/(E_i \tau_i) > 0$ for all $i \in S$.

**Proof.** We begin by using the first step analysis to write equations for $E_i [\tau_j]$ as follows:
\[E_i [\tau_j] = \sum_{k=1}^{n} E_i [\tau_j | X_1 = k] Q_{ik} = E_i [\tau_j | X_1 = k] Q_{ik} + Q_{ij} 1 \]
\[= \sum_{k \neq j} (E_k [\tau_j] + 1) Q_{ik} + Q_{ij} 1 = \sum_{k \neq j} E_k [\tau_j] Q_{ik} + 1.\]

and therefore,
\[E_i [\tau_j] = \sum_{k \neq j} Q_{ik} E_k [\tau_j] + 1. \tag{20.73}\]

Now suppose that $\pi$ is any invariant distribution for $Q$, then multiplying Eq. \[20.73\] by $\pi_i$ and summing on $i$ shows
\[
\sum_{i=1}^{n} \pi_i E_i [\tau_j] = \sum_{i=1}^{n} \pi_i \sum_{k \neq j} Q_{ik} E_k [\tau_j] + \pi_i 1
\Rightarrow \sum_{k \neq j} \pi_k E_k [\tau_j] + 1.
\]

Since $\sum_{k \neq j} \pi_k E_k [\tau_j] + 1$ we may cancel it from both sides of this equation in order to learn $\pi_j \pi_j = 1$. \[\blacksquare\]

We may use Eq. \[20.73\] to compute $E_i [\tau_j]$ in examples. To do this, fix $j$ and set $v_i := E_i \tau_j$. Then Eq. \[20.73\] states that $v = Q^{j} v$ where $Q^{j}$ denotes $Q$ with the $j^{th}$ column replaced by all zeros. Thus we have
\[E_i [\tau_j]^n_{i=1} = (I - Q^{j})^{-1} 1, \tag{20.74}\]
i.e.
\[
\begin{bmatrix}
E_1 [\tau_j] \\
\vdots \\
E_n [\tau_j]
\end{bmatrix} = (I - Q^{j})^{-1}
\begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}. \tag{20.75}
\]
20.6.2 Some worked examples

Example 20.62. Let \( S = \{1, 2\} \) and \( Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \) with jump diagram in Figure 20.4. In this case \( Q^{2n} = I \) while \( Q^{2n+1} = Q \) and therefore \( \lim_{n \to \infty} Q^n \) does not exist. On the other hand it is easy to see that the invariant distribution, \( \pi \), for \( Q \) is \( \pi = [1/2 \ 1/2] \) 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

\[
\begin{bmatrix} E_1 \tau_1 \\ E_2 \tau_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}
\]

and

\[
\begin{bmatrix} E_1 \tau_2 \\ E_2 \tau_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\]

so that indeed, \( \pi_1 = 1/E_1 \tau_1 \) and \( \pi_2 = 1/E_2 \tau_2 \). Of course \( \tau_1 = 2 \) (\( P_1 \) -a.s.) and \( \tau_2 = 2 \) (\( P_2 \) -a.s.) so that it is obvious that \( E_1 \tau_1 = E_2 \tau_2 = 2 \).

Example 20.63. Again let \( S = \{1, 2\} \) and \( Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \) with jump diagram in Figure 20.5. In this case the chain is not irreducible and every \( \pi = [a \ b] \) with

\[
\begin{array}{c}
\text{1} \\
\text{2} \\
\end{array}
\]

Fig. 20.4. A non-random chain.

\[
a + b = 1 \text{ and } a, b \geq 0 \text{ is an invariant distribution.}
\]

Example 20.64. Suppose that \( S = \{1, 2, 3\} \), and

\[
Q = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}.
\]

has the jump graph given by 20.6. Notice that \( Q^7_{11} > 0 \) and \( Q^7_{11} > 0 \) that \( Q \) is

\[
\text{aperiodic.}
\]

We now find the invariant distribution,

\[
\text{Nul} \ (Q - I)^T = \text{Nul} \left[ \begin{array}{rrrr} -1 & 1/2 & 1/2 & 1 \\ -1 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & -1 \end{array} \right] = \mathbb{R} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.
\]

Therefore the invariant distribution is given by

\[
\pi = \frac{1}{5} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.
\]

Let us now observe that

\[
Q^2 = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 1/2 \end{bmatrix},
\]

\[
Q^3 = \begin{bmatrix} 1 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix},
\]

\[
Q^{20} = \begin{bmatrix} 0.399 & 0.400 & 0.399 & 0.400 \\ 0.399 & 0.400 & 0.399 & 0.400 \\ 0.399 & 0.400 & 0.399 & 0.400 \end{bmatrix}.
\]

Let us also compute \( E_{2 \tau_3} \) via,

\[
\begin{bmatrix} E_{1 \tau_3} \\ E_{2 \tau_3} \\ E_{3 \tau_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}.
\]
so that 
\[ \frac{1}{E_3^{r_3}} = \frac{1}{5} = \pi_3. \]

**Example 20.65.** The transition matrix,
\[
Q = \begin{pmatrix}
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 1 \\
\frac{1}{2} & 0 & \frac{1}{2} & 2 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 3
\end{pmatrix}
\]
is represented by the jump diagram in Figure 20.7. This chain is aperiodic. We find the invariant distribution as,

\[
\text{Nul} (Q - I)^{tr} = \text{Nul} \left( \begin{pmatrix}
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 1 \\
\frac{1}{2} & 0 & \frac{1}{2} & 2 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 3
\end{pmatrix} - \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} \right)^{tr}
\]

\[
= \text{Nul} \left( \begin{pmatrix}
-\frac{3}{4} & 1 & 1 & \frac{1}{4} \\
1 & -\frac{1}{2} & \frac{1}{2} & \frac{3}{4} \\
1 & \frac{1}{2} & \frac{1}{2} & -\frac{5}{4}
\end{pmatrix} \right) = \mathbb{R} \begin{pmatrix}
\frac{1}{6} \\
\frac{1}{6}
\end{pmatrix} = \mathbb{R} \begin{pmatrix}
6 \\
6
\end{pmatrix}
\]

\[
\pi = \frac{1}{17} \begin{pmatrix}
6 \\
5 \\
6
\end{pmatrix} = \begin{pmatrix}
0.352942 \\
0.29412 \\
0.35294
\end{pmatrix}.
\]

In this case
\[
Q^{10} = \begin{pmatrix}
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 1 \\
\frac{1}{2} & 0 & \frac{1}{2} & 2 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 3
\end{pmatrix}^{10}
\]

\[
= \begin{pmatrix}
0.35298 & 0.29404 & 0.35298 \\
0.35289 & 0.29423 & 0.35289 \\
0.35295 & 0.2941 & 0.35295
\end{pmatrix}.
\]

Let us also compute
\[
\begin{pmatrix}
E_1^{r_2} \\
E_2^{r_2} \\
E_3^{r_2}
\end{pmatrix} = \left( \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} - \begin{pmatrix}
\frac{1}{4} & 0 & \frac{1}{4} \\
\frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix} \right)^{-1} \begin{pmatrix}
1 \\
1
\end{pmatrix} = \begin{pmatrix}
11 \\
12
\end{pmatrix}
\]

so that
\[
\frac{1}{E_2^{r_2}} = \frac{5}{17} = \pi_2.
\]

**Example 20.66.** Consider the following Markov matrix,
\[
Q = \begin{pmatrix}
1 & 2 & 3 & 4 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{2} & 0 & 0 & 1 \\
0 & \frac{1}{4} & 0 & 0
\end{pmatrix}
\]

with jump diagram in Figure 20.8. Since this matrix is doubly stochastic (i.e.

\[
\sum_{i=1}^{4} Q_{ij} = 1 \text{ for all } j \text{ as well as } \sum_{j=1}^{4} Q_{ij} = 1 \text{ for all } i, \]

it is easy to check that
\[
\pi = \frac{1}{4} \begin{pmatrix}
1 \\
1 \\
1 \\
1
\end{pmatrix} \text{. Let us compute } E_3^{r_3} \text{ as follows}
\]

\[
\sum_{i=1}^{4} Q_{ij} = 1 \text{ for all } j \text{ as well as } \sum_{j=1}^{4} Q_{ij} = 1 \text{ for all } i, \]

it is easy to check that
\[
\pi = \frac{1}{4} \begin{pmatrix}
1 \\
1 \\
1 \\
1
\end{pmatrix} \text{. Let us compute } E_3^{r_3} \text{ as follows}
\]
\[
\begin{bmatrix}
E_{1T_3} \\
E_{2T_3} \\
E_{3T_3} \\
E_{4T_3}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
1/4 & 1/4 & 0 & 1/4 \\
1/4 & 0 & 3/4 & 0 \\
1/2 & 1/2 & 0 & 0 \\
0 & 1/4 & 0 & 0
\end{bmatrix}^{-1} \begin{bmatrix}
1 \\
1 \\
1 \\
1
\end{bmatrix} = \begin{bmatrix}
54/17 \\
4/17 \\
17/54 \\
17/54
\end{bmatrix}
\]
so that \(E_{3T_3} = 4 = 1/\pi_4\) as it should be. Similarly,
\[
\begin{bmatrix}
E_{1T_2} \\
E_{2T_2} \\
E_{3T_2} \\
E_{4T_2}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
1/4 & 0 & 1/4 & 0 \\
1/4 & 0 & 0 & 3/4 \\
1/2 & 0 & 0 & 0 \\
0 & 0 & 3/4 & 0
\end{bmatrix}^{-1} \begin{bmatrix}
1 \\
1 \\
1 \\
1
\end{bmatrix} = \begin{bmatrix}
54/17 \\
4/17 \\
17/54 \\
17/54
\end{bmatrix}
\]
and again \(E_{2T_2} = 4 = 1/\pi_2\).

20.6.3 Exercises

Exercise 20.15 (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:
\[
au(x + 1) + bu(x) + cu(x - 1) = 0 \tag{20.76}
\]
for \(\alpha < x < \beta\). Show:

1. For any \(\lambda \in \mathbb{C}\),
\[
a \lambda^{x+1} + b \lambda^x + c \lambda^{x-1} = \lambda^{x-1} p(\lambda) \tag{20.77}
\]
where \(p(\lambda) = a \lambda^2 + b \lambda + c\) is the **characteristic polynomial** associated to Eq. (20.76). 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. (20.76) it follows that for any choice of \(A_\pm \in \mathbb{R}\), the function,
\[
w(x) := A_+ \lambda_+^x + A_- \lambda_-^x,
\]
solves Eq. (20.76) 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
\[
u(x) := A_+ \lambda_+^x + A_- \lambda_-^x\]
for all \(\alpha \leq x \leq \beta\).

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. (20.76) for all \(x \in \mathbb{Z}\). **Hint**: Differentiate Eq. (20.77) with respect to \(\lambda\) and then set \(\lambda = \lambda_0\).

4. Again show that any function \(u\) solving Eq. (20.76) is of the form \(u(x) = (A_0 + A_1 x) \lambda_0^x\) for \(\alpha \leq x \leq \beta\) 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
\[
P(X_{n+1} = x + 1|X_n = x) = 1/2 \text{ for all } n \in \mathbb{N}_0 \text{ and } x \in \mathbb{Z}.
\]
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 \(20.55\) 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 20.16. Let \(w_x := P_x(X_{T_{a,b}} = b) := P(X_{T_{a,b}} = b|X_0 = x)\).

1. Use first step analysis to show for \(a < x < b\) that
\[
w_x = \frac{1}{2} (w_{x+1} + w_{x-1}) \tag{20.78}
\]
provided we define \(w_a = 0\) and \(w_b = 1\).

2. Use the results of Exercise 20.15 to show
\[
P_x(X_{T_{a,b}} = b) = w_x = \frac{1}{b - a} (x - a). \tag{20.79}
\]

3. Let
\[
T_b := \min\{n : X_n = b\} \text{ if } \{X_n\} \text{ hits } b
\]
\[
\infty \text{ otherwise}
\]
be the first time \(\{X_n\}\) hits \(b\). Explain why, \(\{X_{T_{a,b}} = b\} \subset \{T_b < \infty\}\) and use this along with Eq. (20.79) to conclude \(P_x(T_b < \infty) = 1\) for all \(x < b\). (By symmetry this result holds true for all \(x \in \mathbb{Z}\).)\footnote{Apply this corollary to finite walk in \([a, b] \cap \mathbb{Z}\).} \footnote{The fact that \(P_x(T_b < \infty) = 1\) is also follows from Example 13.61 above.}
Combining the last couple of exercises together shows that passing to the limit as \( u \rightarrow 0 \) in Eq. (20.80) for all \( x \) with \( w_b = 1 \).

Exercise 20.18. Let \( T = T_{a,b} \) and \( u_x := E_x T := E[T | X_0 = x] \).

1. Use first step analysis to show for \( a < x < b \) that
   \[
   u_x = \frac{1}{2} (u_{x+1} + u_{x-1}) + 1 \quad (20.80)
   \]
   with the convention that \( u_a = 0 = u_b \).

2. Show that
   \[
   u_x = A_0 + A_1 x - x^2 \quad (20.81)
   \]
   solves Eq. (20.80) 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
   \[
   E_x T_{a,b} = -ab + (b + a) x - x^2 = -a (b - x) + bx - x^2. \quad (20.82)
   \]

Exercise 20.19. Let \( T = T_b \). 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. (20.78) for all \( x \) with \( w_b = 1 \).

2. Use Exercise 20.18 to show that there is a constant, \( c \), such that
   \[
   w_x = c (x - b) + 1 \quad \text{for all} \quad x \in \mathbb{Z}.
   \]

3. Explain why \( c \) must be zero to again show that \( P_x (T_b < \infty) = 1 \) for all \( x \in \mathbb{Z} \).

Exercise 20.17. 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. (20.78) for all \( x \) with \( w_b = 1 \).

2. Use Exercise 20.18 to show that there is a constant, \( c \), such that
   \[
   w_x = c (x - b) + 1 \quad \text{for all} \quad x \in \mathbb{Z}.
   \]

3. Explain why \( c \) must be zero to again show that \( P_x (T_b < \infty) = 1 \) for all \( x \in \mathbb{Z} \).

Exercise 20.20. Let us write \( A \subset T \) to mean \( A \subset T \) and \( A \) is at most countable. Show

\[
S = \bigcup_{A \subset T} X_{\tilde{A}}^{-1} (S_A). \quad (20.84)
\]

Hint: Verify that \( X_{\tilde{A}}^{-1} (S_A) = \bigcup_{A' \subset T} X_{\tilde{A}}^{-1} (S_A) \) for all \( A' \subset T \) by showing that \( S_0 := \bigcup_{A \subset T} X_{\tilde{A}}^{-1} (S_A) \) is a \( \sigma \)-algebra.

20.7 Appendix: Kolmogorov’s extension theorem II

The Kolmogorov extension Theorem [12.58] 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, \( \{ (S_i, S_i) \}_{i \in T} \) be a collection of standard Borel spaces, \( S = \prod_{i \in T} S_i \), \( S := \otimes_{i \in T} S_i \), and for \( A \subset T \) let

\[
\left( S_A := \prod_{i \in A} S_i, S_A := \otimes_{i \in A} S_i \right)
\]

and \( X_A : S \rightarrow S_A \) be the projection map, \( X_A (x) := x \vert_A \). If \( A \subset A' \subset T \), also let \( X_{A', A} : S_{A'} \rightarrow S_A \) be the projection map, \( X_{A', A} (x) := x \vert_A \) for all \( x \in S_{A'} \).

**Theorem 20.68 (Kolmogorov).** For each \( A \subset T \) (i.e., \( A \subset T \) and \( \# (A) < \infty \)), let \( \mu_A \) be a probability measure on \( (S_A, S_A) \). We further suppose \( \{ \mu_A \}_{A \subset T} \) satisfy the following compatibility relations;

\[
\mu_A \circ X_{A, A'}^{-1} = \mu_A \quad \text{for all} \quad A \subset A' \subset T. \quad (20.83)
\]

Then there exists a unique probability measure, \( P \), on \( (S, S) \) such that \( P \circ X_A^{-1} = \mu_A \) for all \( A \subset f T \).

**Proof.** (For slight variation on the proof of this theorem given here, see Exercise 20.21.) Let

\[
A := \bigcup_{A \subset T} X_A^{-1} (S_A)
\]

and for \( A = X_A^{-1} (A') \in A \), let \( P (A) := \mu_A (A') \). The compatibility conditions in Eq. (20.83) imply \( P \) is a well defined finitely additive measure on the algebra, \( A \). We now complete the proof by showing \( P \) is continuous on \( A \).

To this end, suppose \( A_n := X_{A_n}^{-1} (A_n) \in A \) with \( A_n \downarrow \emptyset \) as \( n \to \infty \). Let \( \Lambda := \bigcup_{n=1}^{\infty} A_n = A \) be a countable subset of \( T \). Owing to Theorem 12.58 there is a unique probability measure, \( P_A \), on \( (S_A, S_A) \) such that \( P_A (X_T^{-1} (A)) = \mu_A (A) \) for all \( \Gamma \subset A \). Then we have

\[
P (A_n) = \mu_{A_n} (A'_n) = P_A (\tilde{A}_n)
\]

with \( \tilde{A}_n \downarrow \emptyset \) as \( n \to \infty \). Since \( P_A \) is a measure, we may conclude

\[
\lim_{n \to \infty} P (A_n) = \lim_{n \to \infty} P_A (\tilde{A}_n) = 0.
\]

\[\blacksquare\]

Exercise 20.20. Let us write \( A \subset T \) to mean \( A \subset T \) and \( A \) is at most countable. Show

\[
S = \bigcup_{A \subset T} X_{\tilde{A}}^{-1} (S_A). \quad (20.84)
\]

Hint: Verify that \( X_{\tilde{A}}^{-1} (S_A) = \bigcup_{A' \subset T} X_{\tilde{A}}^{-1} (S_A) \) for all \( A' \subset T \) by showing that \( S_0 := \bigcup_{A \subset T} X_{\tilde{A}}^{-1} (S_A) \) is a \( \sigma \)-algebra.
Exercise 20.21. For each $A \subset T$, let $S_A' := X_A^{-1}(S_A) = \sigma(X_i : i \in A) \subset S$.
Show:
1. if $U, V \subset T$ then $S_U' \cap S_V' = S_{U \cap V}'$.
2. By Theorem 20.85 if $U, V \subset T$, there exists unique probability measures, $P_U$ and $P_V$ on $S_U'$ and $S_V'$ respectively such that $P_U \circ X_A^{-1} = \mu_A$ for all $A \subset \subset T$ and $P_V \circ X_A^{-1} = \mu_A$ for all $A \subset \subset V$. Show $P_U = P_V$ on $S_U' \cap S_V'$.
Hence for any $A \in S$ we may define $P(A) := P_U(A)$ provided $A \in S_U'$.
3. Show $P$ defined in the previous item is a countably additive measure on $S$.

20.8 Removing the standard Borel restriction

Theorem 20.69. Let \{$(S_n, S_n)$\}$\in\mathbb{N}$ be a collection of measurable spaces, $S = \prod_{n=0}^{\infty} S_n$ and $S := \otimes_{n=0}^{\infty} S_n$. Moreover for each $n \in \mathbb{N}$ let $S_n := S_n \times \cdots \times S_n$ and $S_n := S_0 \otimes \cdots \otimes S_n$. We further suppose that $\mu_0$ is a given probability measure on $(S_0, S_0)$ and $T_n : S^{n-1} \times S_n \to [0,1]$ for $n = 1, 2, \ldots$ are given probability kernels on $S^{n-1} \times S_n$. Finally let $\nu_n$ be the probability measure on $(S^n, S_n)$ defined inductively by,

$$\mu_n(dx_0, \ldots, dx_n) = \mu_{n-1}(dx_0, \ldots, dx_{n-1}) T_n(x_0, \ldots, x_{n-1}, x_n) \quad \forall n \in \mathbb{N}.$$  

Then there exists a unique probability measure, $P$ on $(S, S)$ such that

$$P(f) = \int_{S^n} F d\mu_n$$

whenever $f(x) = F(x_0, \ldots, x_n)$ for some $F \in (S_n)_b$.

Remark 20.70 (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 $B_n := \sigma(X_0, \ldots, X_n)$. If $P$ exists, then

$$P[F(x_0, \ldots, x_n) | B_n] = T_n(x_0, \ldots, x_n, F(x_0, \ldots, x_n))$$

and

$$= (TF(x_0, \ldots, x_n))(x_0, \ldots, x_n).$$

Indeed,

$$E[T_n(x_0, \ldots, x_n; F(x_0, \ldots, x_n)) G(x_0, \ldots, x_n)]$$

$$= \int T_n(x_0, \ldots, x_n, dx_{n+1}) F(x_0, \ldots, x_n, x_{n+1}) G(x_0, \ldots, x_n) d\mu_n(x_0, \ldots, x_n)$$

$$= \int F(x_0, \ldots, x_n, x_{n+1}) G(x_0, \ldots, x_n) d\mu_{n+1}(x_0, \ldots, x_n, x_{n+1})$$

$$= E[F(x_0, \ldots, x_n, x_{n+1}) G(x_0, \ldots, x_n)].$$

Now suppose that $f_n = F_n(x_0, \ldots, x_n)$ 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 $f_n \geq f_{\infty}$ for all $n$ and therefore $f_n \geq E[f\infty | B_n] := f_n$. We also use

$$f_n(x_0, x_1, \ldots, x_n)$$

$$= E[f\infty | B_n] = E[E[f\infty | B_{n+1}] | B_n]$$

$$= E[f_{n+1} | B_n] = \int f_{n+1}(x_0, x_1, \ldots, x_{n+1}) T_{n+1}(x_0, \ldots, x_n, dx_{n+1})$$

and $P(f_n) = P(f\infty) = \lim_{n \to \infty} P(f_m) = \varepsilon > 0$ (we only use the case where $n = 0$ here). Since $P(f_0(X_0)) = \varepsilon > 0$, there exists $x_0 \in S_0$ such that

$$\varepsilon \leq f_0(x_0) = E[f_1 | B_0] = \int f_1(x_0, x_1) T_1(x_0, dx_1)$$

and so similarly there exists $x_1 \in S_1$ such that

$$\varepsilon \leq f_1(x_0, x_1) = \int f_2(x_0, x_1, x_2) T_2(x_0, x_1, dx_2).$$

Again it follows that there must exists an $x_2 \in S_2$ such that $\varepsilon \leq f_2(x_0, x_1, x_2)$.

We continue on this way to find and $x \in S$ such that

$$f_n(x) \geq f_n(x_0, \ldots, x_n) \geq \varepsilon$$

for all $n$.

Thus if $P(f_n) \downarrow \varepsilon > 0$ then $\lim_{n \to \infty} f_n(x) \geq \varepsilon \neq 0$ as desired.

Proof. Now onto the formal proof. Let $S$ denote the space of finitely based bounded cylinder functions on $S$, i.e. functions of the form $f(x) = F(x_0, \ldots, x_n)$ with $F \in S^n_0$. 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 $S$.

Now suppose that $0 \leq f_n \in S$ are forms a decreasing sequence of functions such that $\lim_{n \to \infty} I(f_n) = \varepsilon > 0$. We wish to show that $\lim_{n \to \infty} f_n(x) \neq 0$ for every $x \in S$. By assumption, $f_n(x) = F_n(x_0, \ldots, x_N_n)$ for some $N_n \in \mathbb{N}$ of which we may assume $N_0 < N_1 < N_2 < \ldots$. Moreover if $N_0 = 2 < N_1 = 5 < N_2 = 7 < \ldots$, we may replace $(f_0, f_1, \ldots)$ by

$$(g_0, g_1, g_2, \ldots) = (1, 1, f_0, f_0, f_0, f_1, f_1, f_2, \ldots).$$

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 $g_n(x) = G_n(x_0, \ldots, x_n)$ for some $G_n \in S^n_0$, we may now assume that $f_n(x) = F_n(x_0, \ldots, x_n)$ with $F_n \in S^n_0$. 

For any $k \leq n$ let
\[
F_n^k (x_0, \ldots, x_k) := \int \cdots \int F_n(x_0, \ldots, x_n) \prod_{l=k}^{n-1} T_l(x_0, \ldots, x_l, dx_{l+1})
\]
which is an explicit version of $P_n [F_n (x_0, \ldots, x_n) | x_0, \ldots, x_k] = \mathbb{E} [f_n | B_k] (x)$. By construction of the measures $P_n$ it follows that
\[
P_k F_n^k = P_n F_n = I (f_n) \text{ for all } k \leq n. \tag{20.86}
\]
Since
\[
F_n (x_0, \ldots, x_n) = f_n (x) \leq f_{n+1} (x) = F_{n+1} (x_0, \ldots, x_n, x_{n+1}),
\]
it follows that
\[
F_n^k (x_0, \ldots, x_k) = \int F_n(x_0, \ldots, x_n) \prod_{l=k}^{n} T_l(x_0, \ldots, x_l, dx_{l+1})
\leq \int F_{n+1} (x_0, \ldots, x_n, x_{n+1}) \prod_{l=k}^{n} T_l(x_0, \ldots, x_l, dx_{l+1})
= F_{n+1}^k (x_0, \ldots, x_k).
\]
Thus we may define $F^k (x_0, \ldots, x_k) := \lim_{n \to \infty} F_n^k (x_0, \ldots, x_k)$ which is formally equal to $\mathbb{E} [f | B_k] (x)$. Hence we expect that
\[
F^k (x_0, \ldots, x_k) = \int F^{k+1} (x_0, \ldots, x_k, x_{k+1}) T_k (x_0, \ldots, x_k, dx_{k+1}) \tag{20.87}
\]
by the tower property for conditional expectations. This is indeed that case since,
\[
\int F^{k+1} (x_0, \ldots, x_k, x_{k+1}) T_k (x_0, \ldots, x_k, dx_{k+1})
= \lim_{n \to \infty} \int F^{k+1} (x_0, \ldots, x_k, x_{k+1}) T_k (x_0, \ldots, x_k, dx_{k+1})
\]
while
\[
\int F_{n+1}^{k+1} (x_0, \ldots, x_k, x_{k+1}) T_k (x_0, \ldots, x_k, dx_{k+1})
= \int \left[ \ldots \int F_n (x_0, \ldots, x_n) \prod_{l=k+1}^{n} T_l(x_0, \ldots, x_l, dx_{l+1}) \right] T_k (x_0, \ldots, x_k, dx_{k+1})
= \int \cdots \int F_n (x_0, \ldots, x_n) \prod_{l=k}^{n} T_l(x_0, \ldots, x_l, dx_{l+1})
= F_n^k (x_0, \ldots, x_k).
\]
We may now pass to the limit as $n \to \infty$ in Eq. (20.86) to find
\[
P_k (F^k) = \epsilon > 0 \text{ for all } k.
\]
For $k = 0$ it follows that $F^0 (x_0) \geq \epsilon > 0$ for some $x_0 \in S_0$ for otherwise $P_0 (F_0) < \epsilon$. But
\[
\epsilon \leq F^0 (x_0) = \int F^1 (x_0, x_1) T_1 (x_0, dx_1)
\]
and so there exists $x_1$ such that
\[
\epsilon \leq F^1 (x_0, x_1) = \int F^2 (x_0, x_1, x_2) T_2 (x_0, x_1, dx_2)
\]
and hence there exists $x_2$ such that $\epsilon \leq F^2 (x_0, x_1, x_2)$, etc. etc. Thus in the end we find an $x = (x_0, x_1, \ldots) \in S$ such that $F^k (x_0, \ldots, x_n) \geq \epsilon$ for all $k$. Finally recall that
\[
F_n^k (x_0, \ldots, x_k) \geq F^k (x_0, \ldots, x_k) \geq \epsilon \text{ for all } k \leq n.
\]
Taking $k = n$ then implies,
\[
f_n (x) = F_n^n (x_0, \ldots, x_n) \geq F^n (x_0, \ldots, x_n) \geq \epsilon \text{ for all } n.
\]
Therefore we have constructed a $x \in S$ such that $f (x) = \lim_{n \to \infty} f_n (x) \geq \epsilon > 0$.

We may now use the Caratheodory extension theorem to show that $P$ extends to a countably additive measure on $(S, S)$. Indeed suppose $A_n \in \mathcal{A} (X_i : i \in \mathbb{N})$. 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 20.71 (Infinite Product Measures).** Let $\{ (S_n, \mu_n) \}_{n \in \mathbb{N}_0}$ be a collection of measurable spaces, then there exists $P$ on $(S, S)$ such that
\[
P (f) = P (f) = \int_{S^n} F (x_0, \ldots, x_n) dv_0 (x_0) \cdots dv_n (x_n)
\]
whenever $f (x) = F (x_0, \ldots, x_n)$ for some $F \in (S_n)_b$.

**Proof.** Let $\mu_0 = \nu_0$ and
\[
\mu_n (dx_0, \ldots, dx_{n-1}, dx_{n+1}) = \nu_n (dx_n).
\]
Then in this case we will have
\[
\mu_n (dx_0, \ldots, dx_n) = dv_0 (x_0) dv_1 (dx_1) \cdots \nu_n (dx_n)
\]
as desired.
20.9 *Appendix: More Probability Kernel Constructions

Lemma 20.72. 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

\[
y \mapsto \int_Z R(y, dz) F(y, z)
\]

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,

\[
((Q \otimes R) \otimes S)(x, A \times B \times C)
\]

\[
= \int_Y Q(x, dy) \int_{Z \times W} RS(y, dz, dw) 1_{A \times B \times C}(y, z, w)
\]

\[
= \int_Y Q(x, dy) 1_A(y) \int_{Z \times W} RS(y, B \times C)
\]

\[
= \int_Y Q(x, dy) 1_A(y) \int_{Z \times W} R(y, dz) S(z, dw) 1_{B \times C}(z, w)
\]

\[
= \int_Y Q(x, dy) 1_A(y) \int_Z R(y, dz) S(z, C) 1_B(z)
\]

while

\[
(Q \otimes R)(x, A \times B \times C)
\]

\[
= \int_Y Q(x, dy) \int_{Z \times W} S(z, dw) 1_{A \times B \times C}(y, z, w)
\]

\[
= \int_Y Q(x, dy) 1_A(y) \int_{Z \times W} S(z, C)
\]

\[
= \int_Y Q(x, dy) 1_A(y) \int_Z S(z, C) 1_B(z).
\]

\[\blacksquare\]

Corollary 20.73. Keeping the notation in Lemma 20.72, 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

\[
Q(RS)(x; C) = Q \otimes (RS)(x; Y \times C) = \int_Y Q(x, dy) (RS)(y; C)
\]

\[
= \int_Y Q(x, dy) (R \otimes S)(y; Z \times C) = [Q \otimes (R \otimes S)](Y \times Z \times C).
\]

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 20.72.

\[\blacksquare\]
(Sub and Super) Martingales

Let us start with a reminder of a few key notions that were already introduced in Chapter 20. As usual we will let \((S, \mathcal{S})\) denote a measurable space called state space. (Often in this chapter we will take \((S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}_2)\).) As in Chapter 20 we will fix a filtered probability space, \((\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})\), i.e. \(\mathcal{B}_n \subset \mathcal{B}_{n+1} \subset \mathcal{B}\) for all \(n = 0, 1, 2, \ldots\). We further define

\[
\mathcal{B}_\infty := \bigvee_{n=0}^\infty \mathcal{B}_n := \sigma(\bigcup_{n=0}^\infty \mathcal{B}_n) \subset \mathcal{B}. \tag{21.1}
\]

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 \(\mathcal{B}_n/S\)-measurable for all \(n\).

**Definition 21.1.** 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, or 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 21.4.

By induction one shows that \(X\) is a supermartingale, martingale, or submartingale if

\[
\mathbb{E}[X_n|\mathcal{B}_m] \leq \geq X_m \text{ a.s for all } n \geq m, \tag{21.2}
\]

to be read from top to bottom respectively. This last equation may also be expressed as

\[
\mathbb{E}[X_n|\mathcal{B}_m] \leq \geq X_m \land m \text{ a.s for all } m, n \in \mathbb{N}_0. \tag{21.3}
\]

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 21.2 (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 \(\sigma(X_0, \ldots, X_n) \subset \mathcal{B}_n' \subset \mathcal{B}_n\) 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 \(\sigma(X_0, \ldots, X_n) \subset \mathcal{B}_n' \subset \mathcal{B}_n\), for all \(n\),

\[
\mathbb{E}_{\mathcal{B}_n'} X_{n+1} = \mathbb{E}_{\mathcal{B}_n} \mathbb{E}_{\mathcal{B}_n} X_{n+1} \leq \geq \mathbb{E}_{\mathcal{B}_n} X_n = X_n,
\]

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.

### 21.1 (Sub and Super) Martingale Examples

**Example 21.3.** 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, \(\mathcal{B}_n^Z := \sigma(Z_0, \ldots, Z_n)\). 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 \(\{S_n\}_{n\geq0}\) is a submartingale if \(\mathbb{E} Z_n \geq 0\) and supermartingale if \(\mathbb{E} Z_n \leq 0\) for all \(n\).

**Exercise 21.1.** 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, \(\{M_n\}_{n=0}^\infty\) will be a martingale which is not of the form \(M_n = \mathbb{E}_{\mathcal{B}_n} X\) for some \(X \in L^1(\mathbb{P})\).) **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 \(\mathbb{E}|Z_k|\) growing rather rapidly.
Example 21.4 (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 $n$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.”

Now we assume that $Z_n$ is independent of $(W, Z_1, \ldots, Z_{n-1})$ for each $n$, i.e. the dice are not influenced by the previous plays or your whims. If we let $B_n := \sigma(W, Z_1, \ldots, Z_n)$ with $B_0 = \sigma(W)$, then we are assuming the $Z_n$ is independent of $B_{n-1}$ for each $n \in \mathbb{N}$.

As a gambler, you are allowed to choose before the $n$th - game is played, the amounts $\{(C_n(s))_{s \in S}\}$ that you want to bet on each of the possible outcomes of the $n$th - game. Assuming that you are not clairvoyant (i.e. can not see the future), these amounts may be random but must be $B_{n-1}$ - measurable, that is $C_n(s) = C_n(W, Z_1, \ldots, Z_{n-1}, s)$, i.e. $\{C_n(s)\}_{n=1}^{\infty}$ is “previsible” process (see Definition 21.5 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 $n$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 $n$th - game and $C_n(Z_n) \alpha(Z_n)$ represents the house’s payoff to you for the $n$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 $B_n$ - measurable for each $n$, and

$$\mathbb{E}_{B_{n-1}}[X_n - X_{n-1}] = -\sum_{s \in S} C_n(s) + \mathbb{E}_{B_{n-1}}[C_n(Z_n) \alpha(Z_n)]$$

$$= -\sum_{s \in S} C_n(s) + \sum_{s \in S} C_n(s) \alpha(s) p(s)$$

$$= \sum_{s \in S} C_n(s) (\alpha(s) p(s) - 1).$$

Thus it follows, that no matter the choice of the betting “strategy,” $\{C_n(s) : s \in S\}_{n=1}^{\infty}$ we will have

$$\mathbb{E}_{B_{n-1}}[X_n - X_{n-1}] = \begin{cases} 0 & \text{if } \alpha(\cdot) p(\cdot) \geq 1 \\ \geq 0 & \text{if } \alpha(\cdot) p(\cdot) = 1 \\ \leq 0 & \text{if } \alpha(\cdot) p(\cdot) \leq 1 \end{cases},$$

that is $\{X_n\}_{n=0}^{\infty}$ is a sub-martingale, martingale, or supermartingale depending on whether $\alpha \cdot p \geq 1$, $\alpha = 1$, 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 21.5.** We say $\{C_n : \Omega \to S\}_{n=1}^{\infty}$ is predictable or previsible if each $C_n$ is $B_{n-1}/S$ - 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 $B_n := \sigma(X_0, \ldots, X_n)$. An application of Lemma 17.1 shows that a sequence of random variables, $\{Y_n\}_{n=0}^{\infty}$, is adapted to the filtration iff there are $\mathcal{S}_{\Omega(n+1)/\mathbb{R}}$ - measurable functions, $f_n : S^{n+1} \to \mathbb{R}$, such that $Y_n = f_n(X_0, \ldots, X_n)$ 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 $Z_n = f_n(X_0, \ldots, X_{n-1})$ for all $n \in \mathbb{N}$.

**Example 21.6 (Regular martingales).** Suppose that $(\Omega, \mathcal{B}, \{B_n\}_{n=0}^{\infty}, P)$ is a filtered probability space and $X \in L^1(\Omega, \mathcal{B}, P)$. Then $X_n := \mathbb{E}[X|B_n]$ is a martingale. Indeed, by the tower property of conditional expectations,

$$\mathbb{E}[X_{n+1}|B_n] = \mathbb{E}[\mathbb{E}[X|B_{n+1}]|B_n] = \mathbb{E}[X|B_n] = X_n$$

as.s.

When $X_n := \mathbb{E}[X|B_n]$ for some $X \in L^1(P)$ we say that $\{X_n\}_{n=1}^{\infty}$ is a regular martingale.

**Example 21.7.** Suppose that $\Omega = (0, 1]$, $\mathcal{B} = \mathcal{B}_{(0, 1]}$, and $P = \text{Lebesgue measure.}$ Let $P_n = \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right\}_{k=0}^{2^n-1}$ and $B_n := \sigma(P_n)$ for each $n \in \mathbb{N}$. Then $M_n := 2^n 1_{(0, 2^{-n}]}$ for $n \in \mathbb{N}$ is a martingale (Exercise 21.2) such that $\mathbb{E}[M_n] = 1$ for all $n$. However, there is no $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|B_n]$. To verify this last assertion, suppose an $X$ existed. We would then have for $2^n > k > 0$ and any $m > n$, that

$$\mathbb{E}\left[X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = \mathbb{E}\left[\mathbb{E}_{B_m} X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = \mathbb{E}\left[M_m : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = 0.$$
Using $E[X : A] = 0$ for all $A$ in the $\pi$–system, $Q := \bigcup_{k=1}^{\infty} \{ \lambda_{2^k} : 0 \leq k < 2^n \}$, an application of the $\pi$–$\lambda$ theorem shows $E[X : A] = 0$ for all $A \in \sigma(Q) = B$. Therefore $X = 0$ a.s. by Proposition 10.22. But this is impossible since $1 = EM_n = EX$.

**Moral:** not all $L^1$–bounded martingales are regular, i.e. as in Example 21.6. Proposition 21.8 shows what is missing from this martingale in order for it to be of the form in Example 21.6. See the comments after Example 21.11 for another $L^1$–bounded martingale which is not of the form in example 21.6.

**Exercise 21.2.** Show that $M_n := 2^n 1_{[0,2^{-n}]}$ for $n \in \mathbb{N}$ as defined in Example 21.7 is a martingale.

**Proposition 21.8.** Suppose $1 \leq p < \infty$ and $X \in L^p (\Omega, B, P)$. Then the collection of random variables, $\Gamma := \{ E[X|G] : G \subset B \}$ is a bounded subset of $L^p (\Omega, B, P)$ which is also uniformly integrable.

**Proof.** Since $E_G$ is a contraction on all $L^p$–spaces it follows that $\Gamma$ is bounded in $L^p$ with

$$\sup_{G \subset B} \|E[X|G]\|_p \leq \|X\|_p .$$

For the $p > 1$ the uniform integrability of $\Gamma$ follows directly from Lemma 15.54.

We now concentrate on the $p = 1$ case. Recall that $|E_G X| \leq E_G |X|$ a.s. and therefore,

$$E[|E_G X| : |E_G X| \geq a] \leq E[|X| : |E_G X| \geq a] \text{ for all } a > 0.$$  

But by Chebyshev’s inequality,

$$P(|E_G X| \geq a) \leq \frac{1}{a} E[|E_G X|] \leq \frac{1}{a} E[|X|].$$

Since $\{|X|\}$ is uniformly integrable, it follows from Proposition 15.48 that, by choosing $a$ sufficiently large, $E[|X| : |E_G X| \geq a]$ is as small as we please uniformly in $G \subset B$ and therefore,

$$\lim_{a \to \infty} \sup_{G \subset B} E[|E_G X| : |E_G X| \geq a] = 0.$$  

**Example 21.9.** This example generalizes Example 21.7. Suppose $(\Omega, B, \{B_n\}_{n=0}^{\infty}, P)$ is a filtered probability space and $Q$ is another probability measure on $(\Omega, B)$. Let us assume that $Q|B_n \ll P|B_n$ for all $n$, which by the Raydon-Nikodym Theorem 18.8 implies there exists $0 \leq X_n \in L^1 (\Omega, B_n, P)$ with $EX_n = 1$ such that $dQ|B_n = X_n dP|B_n$, or equivalently put, for any $B \in B_n$ we have

$$Q(B) = \int_B X_n dP = E[X_n : B].$$

Since $B \in B_n \subset B_{n+1}$, we also have $E[X_n+1 : B] = Q(B) = E[X_n : B]$ for all $B \in B_n$, and hence $E[X_{n+1}|B_n] = X_n$ a.s., i.e. $X = \{X_n\}_{n=0}^{\infty}$ is a positive martingale.

Example 21.10 is of this form with $Q = \delta_0$. Notice that $\delta_0|B_n \ll m|B_n$ for all $n < \infty$ while $\delta_0 \perp m$ on $B_{[0,1]} = B_\infty$. See Section 22.4 for more in the direction of this example.

**Lemma 21.10.** Let $X := \{X_n\}_{n=0}^{\infty}$ be an adapted process of integrable random variables on a filtered probability space, $(\Omega, B, \{B_n\}_{n=0}^{\infty}, P)$ and let $d_n := X_n - X_{n-1}$ with $X_{-1} := EX_0$. Then $X$ is a martingale (respectively submartingale or supermartingale) iff $E[d_n+1|B_n] = 0$ ($E[d_n+1|B_n] \geq 0$ or $E[d_n+1|B_n] \leq 0$ respectively) for all $n \in \mathbb{N}$.

Conversely if $\{d_n\}_{n=1}^{\infty}$ is an adapted sequence of integrable random variables and $X_0$ is a $B_0$–measurable integrable random variable. Then $X_n = X_0 + \sum_{j=1}^{n} d_j$ is a martingale (respectively submartingale or supermartingale) iff $E[d_{n+1}|B_n] = 0$ ($E[d_{n+1}|B_n] \geq 0$ or $E[d_{n+1}|B_n] \leq 0$ 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

$$0 = E[X_{n+1}|B_n] - X_n = E[X_{n+1} - X_n|B_n] = E[d_{n+1}|B_n].$$

The second assertion is an easy consequence of the first assertion.

**Example 21.11.** Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a sequence of independent integrable random variables, $\bar{X}_n = Z_0 \ldots Z_n$, and $B_n := \sigma(Z_0, \ldots, Z_n)$. (Observe that $E[X_n] = \prod_{k=0}^{n} E[Z_k] < \infty$.) Since

$$E[X_{n+1}|B_n] = E[X_n Z_{n+1}|B_n] = X_n E[Z_{n+1}|B_n] = X_n E[Z_{n+1}]$$

it follows that $\{X_n\}_{n=0}^{\infty}$ is a martingale if $EZ_n = 1$. If we further assume, for all $n$, that $Z_n \geq 0$ so that $X_n \geq 0$, then $\{X_n\}_{n=0}^{\infty}$ is a supermartingale (submartingale) provided $EZ_n \leq 1$ ($EZ_n \geq 1$) for all $n$.

Let us specialize the above example even more by taking $Z_n \overset{d}{=} p + U$ where $p \geq 0$ and $U$ is the uniform distribution on $[0,1]$. In this case we have by the strong law of large numbers that

$$\frac{1}{n} \ln X_n = \frac{1}{n} \sum_{k=0}^{n} \ln Z_k \to E[\ln (p + U)] \text{ a.s. (21.4)}$$

An elementary computation shows

\begin{align*}
Q (B) &= \int_B X_n dP = E[X_n : B].
\end{align*}
Suppose that $\{X_n: n \in \mathbb{N} \}$ is a submartingale. If $\varphi$ is a convex function and increasing, then $\varphi$ preserves the inequality, $X_n \leq E_{B_n} X_{n+1}$, and hence

$$\varphi(X_n) \leq \varphi(E_{B_n} X_{n+1}) \leq E_{B_n} [\varphi(X_{n+1})]$$

which shows $\varphi(X)$ is a submartingale. Similarly, if $X$ is a submartingale and $\varphi$ is convex and increasing, then $\varphi$ preserves the inequality, $X_n \leq E_{B_n} X_{n+1}$, and hence

$$\varphi(X_n) \leq \varphi(E_{B_n} X_{n+1}) \leq E_{B_n} [\varphi(X_{n+1})]$$

so again $\varphi(X)$ is a submartingale.

**Proposition 21.13 (Markov Chains and Martingales).** Suppose that $(\Omega, \mathcal{F}, \{\mathcal{B}\}_{n \in \mathbb{N}}, \{X_n: \Omega \to S\}_{n \geq 0}, Q, P)$ is a time homogeneous Markov chain and $f: \mathbb{N} \times S \to \mathbb{R}$ is measurable function which is either non-negative or satisfies $E[|f(n, X_n)|] < \infty$ for all $n$ and let $Z_n := f(n, X_n)$. Then $\{Z_n\}_{n=0}^\infty$ is a (sub-martingale) martingale if $(Qf(n,1) \leq f(n))$ $Qf(n,1) = f(n)$ for all $n \geq 0$. In particular if $f: S \to \mathbb{R}$ is a function such that $(Qf \leq f)$ $Qf = f$ then $Z_n = f(X_n)$ is a (sub-martingale) martingale. (Also see Exercise 21.5 below.)

**Proof.** Using the Markov property and the definition of $Q$, we have

$$E[Z_{n+1}|\mathcal{B}_n] = E[f(n+1, X_{n+1})|\mathcal{B}_n] = [Qf(n+1, \cdot)](X_n).$$

The latter expression is (less than or equal) equal to $Z_n$ if $(Qf(n+1, \cdot) \leq f(n))$ $Qf(n+1, \cdot) = f(n)$ for all $n \geq 0$.

One way to find solutions to the equation $Qf(n+1, \cdot) = f(n)$ 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

$$f(n, y) := (Q^{T-n}g)(y) \quad \text{for } 0 \leq n \leq T.$$  

Then $Qf(n+1, \cdot) = Q(Q^{T-n-1}g) = Q^{T-n}g = f(n, \cdot)$ and we will have that

$$Z_n = f(n, X_n) = (Q^{T-n}g)(X_n)$$

is a Martingale for $0 \leq n \leq T$. 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 21.14.** In special cases one can often make sense of these expressions (see Exercise 21.5). 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 a $S$-valued random variable independent of $\{X_i\}_{i=1}^\infty$ as in Exercise 20.48. 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:
\[
\mathbb{E}_B_n \left( \frac{q}{p} \right)^{S_{n+1}} = \mathbb{E}_B_n \left( \frac{q}{p} \right)^{S_n + X_{n+1}} = \mathbb{E}_B_n \left( \frac{q}{p} \right)^{S_n} \mathbb{E}_B_n \left( \frac{q}{p} \right)^{X_{n+1}} = \left( \frac{q}{p} \right)^{S_n} \mathbb{E}_B_n \left( \frac{q}{p} \right)^{X_{n+1}} = \left( \frac{q}{p} \right)^{S_n} \cdot \left[ \left( \frac{q}{p} \right)^1 p + \left( \frac{q}{p} \right)^{-1} q \right] = \left( \frac{q}{p} \right)^{S_n} \cdot [q + p] = \left( \frac{q}{p} \right)^{S_n}.
\]

Now suppose that \( \lambda \neq 0 \) and observe that \( Q^\lambda x = (p \lambda + q \lambda^{-1}) \lambda^x \). Thus it follows that we may set \( Q^{-1} \lambda^x = (p \lambda + q \lambda^{-1})^{-1} \lambda^x \) and therefore conclude that
\[
f(n, x) := Q^{-n} \lambda^x = (p \lambda + q \lambda^{-1})^{-n} \lambda^x
\]
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 \( \{M_n = (p \lambda + q \lambda^{-1})^{-n} \lambda^S_n \}_{n \geq 0} \) is a martingale for all \( \lambda \neq 0 \).

**Exercise 21.3.** For \( \theta \in \mathbb{R} \) let
\[
f_\theta(n, x) := Q^{-n} e^{\theta x} = (pe^{\theta} + qe^{-\theta})^{-n} e^{\theta x}
\]
so that \( Qf_\theta(n+1, \cdot) = f_\theta(n, \cdot) \) for all \( \theta \in \mathbb{R} \). Compute;
1. \( f^{(k)}_\theta(n, x) := (\frac{d}{d\theta})^k f_\theta(n, x) \) 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 21.15. If \( \{M_n(\theta)\}_{n=0}^{\infty} \) is a martingale depending differentiability on a parameter \( \theta \in \mathbb{R} \). Then for all \( A \in \mathcal{B}_n \),
\[
\mathbb{E} \left[ \frac{d}{d\theta} M_{n+1}(\theta) : A \right] = \frac{d}{d\theta} \mathbb{E} [M_{n+1}(\theta) : A] = \frac{d}{d\theta} \mathbb{E} [M_n(\theta) : A] = \mathbb{E} \frac{d}{d\theta} M_n(\theta) : A
\]
provided it is permissible to interchange \( \frac{d}{d\theta} \) with the expectations in this equation. Thus under “suitable” hypothesis, we will have \( \{\frac{d}{d\theta} M_n(\theta)\}_{n \geq 0} \) is another martingale.

21.2 Decompositions

**Notation 21.16** Given a sequence \( \{Z_k\}_{k=0}^{\infty} \), let \( \Delta_k Z := Z_k - Z_{k-1} \) for \( k = 1, 2, \ldots \).

**Lemma 21.17 (Doob Decomposition).** Each adapted sequence, \( \{Z_n\}_{n=0}^{\infty} \), of integrable random variables has a unique decomposition,
\[
Z_n = M_n + A_n
\]
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 \( A_0 = 0 \),
\[
A_n := \sum_{k=1}^{n} \mathbb{E}_{B_{k-1}} [\Delta_k Z] \text{ for } n \geq 1
\]
and
\[
M_n = Z_n - A_n = Z_n - \sum_{k=1}^{n} \mathbb{E}_{B_{k-1}} [\Delta_k Z]
\]
(21.7)
\[
= Z_0 + \sum_{k=1}^{n} (Z_k - \mathbb{E}_{B_{k-1}} Z_k).
\]
(21.8)

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. (21.5), then
\[
\mathbb{E}_{B_n} [\Delta_{n+1} Z] = \mathbb{E}_{B_n} [\Delta_{n+1} M + \Delta_{n+1} A] = \Delta_{n+1} A
\]
(21.9)
wherein we have used \( M \) is a martingale and \( A \) is predictable so that \( \mathbb{E}_{B_n} [\Delta_{n+1} M] = 0 \) and \( \mathbb{E}_{B_n} [\Delta_{n+1} A] = \Delta_{n+1} A \). Hence we must define, for \( m \geq 1 \),
Remark 21.18. Suppose that \( X = \{X_n\}_{n=0}^\infty \) is a submartingale and \( X_n = M_n + A_n \) is its Doob decomposition. Then \( A_\infty = \lim_{n \to \infty} A_n \) exists a.s.,

\[
E A_n = E |X_n - M_n| = E X_n - EM_0 = E |X_n - X_0| \tag{21.10}
\]

and hence by MCT,

\[
E A_\infty = \lim_{n \to \infty} E |X_n - X_0|. \tag{21.11}
\]

Hence if \( \lim_{n \to \infty} E |X_n - X_0| = \sup_n E |X_n - X_0| < \infty \), then \( E A_\infty < \infty \) and so by DCT, \( A_n \to A_\infty \) in \( L^1(\Omega,B,P) \). In particular if \( \sup_n E |X_n| < \infty \), we may conclude that \( \{X_n\}_{n=0}^\infty \) is \( L^1(\Omega,B,P) \) convergent iff \( \{M_n\}_{n=0}^\infty \) is \( L^1(\Omega,B,P) \) convergent. (We will see below in Corollary 21.56 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 \( \sup_n E |X_n| < \infty \).)

Example 21.19. Suppose that \( N = \{N_n\}_{n=0}^\infty \) is a square integrable martingale, i.e. \( EN_n^2 < \infty \) for all \( n \). Then from Proposition 21.12 \( X := \{X_n = N_n^2\}_{n=0}^\infty \) is a positive submartingale. In this case

\[
\mathbb{E} B_{k-1} \Delta_k X = \mathbb{E} B_{k-1} (N_k^2 - N_{k-1}^2) = \mathbb{E} B_{k-1} [(N_k - N_{k-1})(N_k + N_{k-1})] \\
= \mathbb{E} B_{k-1} [(N_k - N_{k-1})(N_k - N_{k-1})] \\
= \mathbb{E} B_{k-1} (N_k - N_{k-1})^2
\]

wherein the second to last equality we have used

\[
\mathbb{E} B_{k-1} (N_k - N_{k-1}) N_{k-1} = N_{k-1} \mathbb{E} B_{k-1} (N_k - N_{k-1}) = 0 \text{ a.s.}
\]

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

\[
A_n = \sum_{k \leq n} \mathbb{E} B_{k-1} \Delta_k X = \sum_{k \leq n} \mathbb{E} B_{k-1} (\Delta_k N)^2. \tag{21.12}
\]

Exercise 21.4 (Very similar to above example?). Suppose \( \{M_n\}_{n=0}^\infty \) is a square integrable martingale. Show;

1. \( \mathbb{E} [M_{n+1}^2 - M_n^2 | B_n] = \mathbb{E} [(M_{n+1} - M_n)^2 | B_n] \). Conclude from this that the Doob decomposition of \( M_n^2 \) is of the form,

\[
M_n^2 = N_n + A_n
\]

where

\[
A_n := \sum_{1 \leq k \leq n} \mathbb{E} [(M_k - M_{k-1})^2 | B_{k-1}].
\]

2. If further assume that \( M_k - M_{k-1} \) is independent of \( B_{k-1} \) for all \( k = 1, 2, \ldots \), explain why,

\[
A_n = \sum_{1 \leq k \leq n} \mathbb{E} (M_k - M_{k-1})^2.
\]

The next exercise shows how to characterize Markov processes via martingales.

Exercise 21.5 (Martingale problem I). Suppose that \( \{X_n\}_{n=0}^\infty \) is an \((S,S)\) valued adapted process on some filtered probability space \((\Omega,B,\{B_n\}_{n\in\mathbb{N}_0},P)\) 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 \( \{M^n_f\}_{n \geq 0} \) is a martingale for each \( f \in \mathcal{S}_b \).

2. Conversely if \( \{M^n_f\}_{n \geq 0} \) 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 21.20. 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

\[
\sup_n \mathbb{E}X_n^+ \leq \sup_n \mathbb{E}|X_n| \leq 2\sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0.
\]

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 21.21 (Krickeberg Decomposition). Suppose that \( X \) is an integrable submartingale such that \( C := \sup_n \mathbb{E}|X_n^+| \leq \infty \) or equivalently \( \sup_n \mathbb{E}|X_n| < \infty \), see Eq. (21.13). Then

\[
M_n := \lim_{p \to \infty} \mathbb{E}[X_p^+|\mathcal{B}_n] \text{ exists a.s.,}
\]

\( M = \{M_n\}_{n=0}^{\infty} \) 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 21.12 we know that \( X^+ = \{X_n^+\} \) is a still a positive submartingale. Therefore for each \( n \in \mathbb{N} \), and \( p \geq n \),

\[
\mathbb{E}_{\mathcal{B}_n} [X_{p+1}^+] = \mathbb{E}_{\mathcal{B}_n} \mathbb{E}_{\mathcal{B}_p} [X_{p+1}^+] \geq \mathbb{E}_{\mathcal{B}_n} X_p^+ \text{ a.s.}
\]

Therefore \( \mathbb{E}_{\mathcal{B}_n} X_p^+ \) is increasing in \( p \) for \( p \geq n \) and therefore, \( M_n := \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} [X_p^+] \text{ exists in } [0, \infty) \). By Fatou’s lemma, we know that

\[
\mathbb{E}M_n \leq \liminf_{p \to \infty} \mathbb{E} \mathbb{E}_{\mathcal{B}_n} [X_p^+] \leq \liminf_{p \to \infty} \mathbb{E} X_p^+ = C < \infty
\]

which shows \( M \) is integrable. By cMCT and the tower property of conditional expectation,

\[
\mathbb{E}_{\mathcal{B}_n} M_{n+1} = \mathbb{E}_{\mathcal{B}_n} \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_{n+1}} [X_p^+] = \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} \mathbb{E}_{\mathcal{B}_{n+1}} [X_p^+]
= \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} [X_p^+] = M_n \text{ a.s.,}
\]

which shows \( M = \{M_n\} \) is a martingale.

We now define \( Y_n := M_n - X_n \). Using the submartingale property of \( X^+ \) implies,

\[
Y_n = M_n - X_n = \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} [X_p^+] - X_n = \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} [X_p^+] - X_n^+ + X_n^-
= \lim_{p \to \infty} \mathbb{E}_{\mathcal{B}_n} [X_p^+ - X_n^+] + X_n^+ \geq 0 \text{ a.s.}
\]

Moreover,

\[
\mathbb{E}Y_n^{+1}|\mathcal{B}_n] = \mathbb{E}[M_n^{+1} - X_n^{+1}|\mathcal{B}_n] = M_n - \mathbb{E}[X_n^{+1}|\mathcal{B}_n] \geq M_n - X_n = Y_n
\]

wherein we have use \( M \) is a martingale in the second equality and \( X \) is submartingale the last inequality. □

21.3 Stopping Times

Definition 21.22. Again let \( \{B_n\}_{n=0}^{\infty} \) be a filtration on \( (\Omega, \mathcal{B}) \) and assume that \( \mathcal{B} = \mathcal{B}_{\infty} := \bigvee_{n=0}^{\infty} \mathcal{B}_n := \sigma (\bigcup_{n=0}^{\infty} \mathcal{B}_n) \). A function, \( \tau : \Omega \to \mathbb{N} := \mathbb{N} \cup \{0, \infty\} \) is said to be a stopping time if \( \{\tau \leq n\} \in \mathcal{B}_n \) for all \( n \in \mathbb{N} \). Equivalently put, \( \tau : \Omega \to \mathbb{N} \) is a stopping time iff the process, \( n \to \mathbb{I}_{\tau \leq n} \) is adapted.

Lemma 21.23. Let \( \{B_n\}_{n=0}^{\infty} \) be a filtration on \( (\Omega, \mathcal{B}) \) and \( \tau : \Omega \to \mathbb{N} \) be a function. Then the following are equivalent:

1. \( \tau \) is a stopping time.
2. \( \{\tau \leq n\} \in \mathcal{B}_n \) for all \( n \in \mathbb{N} \).
3. \( \{\tau > n\} \in \mathcal{B}_n \) for all \( n \in \mathbb{N} \).
4. \( \{\tau = n\} \in \mathcal{B}_n \) for all \( n \in \mathbb{N} \).

Moreover if any of these conditions hold for \( n \in \mathbb{N} \) then they also hold for \( n = \infty \).

Proof. (1. \( \iff \) 2.) Observe that if \( \{\tau \leq n\} \in \mathcal{B}_n \) for all \( n \in \mathbb{N} \), then \( \{\tau < \infty\} = \bigcup_{n=0}^{\infty} \{\tau \leq n\} \in \mathcal{B}_{\infty} \) 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 \( n \in \mathbb{N} \).

The equivalence of 2., 3., and 4. follows from the identities
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 \(B_n = \sigma (X_0, \ldots, X_n)\), 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 \(1_{\{\tau = n\}} = f_n (X_0, \ldots, X_n)\). In other words, if \(\tau(\omega) = n\) and \(\omega'\) is any other point in \(\Omega\) such that \(X_k (\omega) = X_k (\omega')\) for \(k \leq n\) then \(\tau(\omega') = n\). Here is another common example of a stopping time.

**Example 21.24 (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, \((\Omega, \mathcal{F}, \{B_n\}_{n=0}^\infty)\) 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

\[
\{\tau = n\} = \{X_0 \in A^c, \ldots, X_{n-1} \in A^c, X_n \in A\} \in \sigma (X_0, \ldots, X_n) \subset B_n.
\]

More generally if \(\sigma\) is a stopping time, then the first hitting time after \(\sigma\),

\[
\tau := \inf \{k \geq \sigma : X_k \in A\},
\]

is also a stopping time. Indeed,

\[
\{\tau = n\} = \{\sigma \leq n\} \cap \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} = \bigcup_{0 \leq k \leq n} \{\sigma = k\} \cap \{X_k \notin A, \ldots, X_1 \notin A, X_0 \in A\}
\]

which is in \(B_n\) for all \(n\). Here we use the convention that

\[
\{X_k \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} = \{X_n \in A\} \text{ if } k = n.
\]

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

\[
\{\tau = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \ldots\} \in \sigma (X_n, X_{n+1}, \ldots)
\]

which typically will not be in \(B_n\).

**Proposition 21.25 (New Stopping Times from Old).** Let \((\Omega, \mathcal{F}, \{B_n\}_{n=0}^\infty)\) 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, \(\sup_k \tau_k = \lim_{k \to \infty} \max \{\tau_1, \ldots, \tau_k\}\) and \(\inf_k \tau_k = \lim_{k \to \infty} \min \{\tau_1, \ldots, \tau_k\}\) are also stopping times.

**Proof.**
1. Since \(\{\tau \wedge \sigma > n\} = \{\tau > n\} \cap \{\sigma > n\} \in \mathcal{B}_n\), \(\{\tau \vee \sigma \leq n\} = \{\tau \leq n\} \cap \{\sigma \leq n\} \in \mathcal{B}_n\) for all \(n\), and

\[
\{\tau + \sigma > n\} = \bigcup_{k=0}^n \{\tau = k, \sigma = n - k\} \in \mathcal{B}_n
\]

for all \(n, \tau \wedge \sigma, \tau \vee \sigma, \tau + \sigma\) are all stopping times.
2. If \(\tau_k \uparrow \tau_\infty\), then \(\{\tau_\infty \leq n\} = \bigcap_k \{\tau_k \leq n\} \in \mathcal{B}_n\), and so \(\tau_\infty\) is a stopping time. Similarly, if \(\tau_k \downarrow \tau_\infty\), then \(\{\tau_\infty > n\} = \bigcap_k \{\tau_k > n\} \in \mathcal{B}_n\) and so \(\tau_\infty\) is a stopping time. (Recall that \(\{\tau_\infty > n\} = \{\tau_\infty \geq n + 1\}\).
3. This follows from items 1. and 2.

**Lemma 21.26.** If \(\tau\) is a stopping time, then the processes, \(f_n := 1_{\{\tau \leq n\}}\), 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 \(f_n := 1_{\sigma < n \leq \tau}\) is predictable.

**Proof.** These are all trivial to prove. For example, if \(f_n := 1_{\sigma < n \leq \tau}\), then \(f_n\) is \(\mathcal{B}_{n-1}\) measurable since,

\[
\{\sigma < n \leq \tau\} = \{\sigma < n\} \cap \{n \leq \tau\} = \{\sigma < n\} \cap \{\tau < n\}^c \in \mathcal{B}_{n-1}.
\]

**Notation 21.27 (Stochastic intervals)** If \(\sigma, \tau: \Omega \to \bar{\mathbb{N}}\), let

\[
(\sigma, \tau) := \{ (\omega, n) \in \Omega \times \bar{\mathbb{N}} : \sigma(\omega) < n \leq \tau(\omega) \}
\]

and we will write \(1_{(\sigma, \tau)}\) for the process, \(1_{\sigma < n \leq \tau}\).

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 \(n \in \mathbb{N}_0\), \(\mathcal{B}_n := \sigma (X_0, \ldots, X_n)\), and \(\tau: \Omega \to \mathbb{N}_0\) is a \(\mathcal{F}\) – stopping time. Recalling that a function \(Y : \Omega \to \mathbb{R}\) is \(\mathcal{B}_n\) measurable iff \(Y (\omega) = f_n (X_0 (\omega), \ldots, X_n (\omega))\) 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 \(Y (\omega) = f_\tau(\omega) (X_0 (\omega), \ldots, X_\tau(\omega))\), where \(f_\tau : \mathbb{R}^{n+1} \to \mathbb{R}\) are measurable random variables. If this is the case, then we would have \(1_{\tau=n} Y = f_n (X_0, \ldots, X_n)\) is \(\mathcal{B}_n\) – measurable for all \(n\). Hence we should define \(A \subset \Omega\) to be in \(\mathcal{B}_\tau\) if \(1_A\) 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 21.28 (Stopped $\sigma$-algebra). Given a stopping time $\tau$ on a filtered measure space $(\Omega, B, \{B_n\}_{n=0}^\infty)$ with $B_\infty := \bigvee_{n=0}^\infty B_n := \sigma(\bigcup_{n=0}^\infty B_n)$, let

$$B_\tau := \{ A \subset \Omega : \{ \tau = n \} \cap A \in B_n \text{ for all } n \leq \infty \}.$$  \hfill (21.14)

Lemma 21.29. Suppose $\sigma$ and $\tau$ are stopping times.

1. A set, $A \subset \Omega$ is in $B_\tau$ iff $A \cap \{ \tau \leq n \} \in B_n$ for all $n \leq \infty$.

2. $B_\tau$ is a sub-$\sigma$-algebra of $B_\infty$.

3. $\tau$ is $B_\tau$-measurable.

4. For all $n \in \mathbb{N}_0$, $\{ \tau = n \} \in B_\tau \cap B_n$ and $B_\tau = B_n$ on $\{ \tau = n \}$.

5. If $\sigma \leq \tau$, then $B_\sigma \subset B_\tau$.

Proof. We take each item in turn.

1. Since

$$A \cap \{ \tau \leq n \} = \bigcup_{k \leq n} [A \cap \{ \tau \leq k \}]$$

and

$$A \cap \{ \tau = n \} = [A \cap \{ \tau \leq n \}] \setminus [A \cap \{ \tau \leq n-1 \}],$$

it easily follows that $A \subset \Omega$ is in $B_\tau$ iff $A \cap \{ \tau \leq n \} \in B_n$ for all $n \leq \infty$.

2. Since $\Omega \cap \{ \tau \leq n \} = \{ \tau \leq n \} \in B_n$ for all $n$, it follows that $\Omega \in B_\tau$. If $A \in B_\tau$, then, for all $n \in \mathbb{N}_0$,

$$A^c \cap \{ \tau \leq n \} = \{ \tau \leq n \} \setminus A = \{ \tau \leq n \} \setminus [A \cap \{ \tau \leq n \}] \in B_n.$$ 

This shows $A^c \in B_\tau$. Similarly if $\{A_k\}_{k=1}^\infty \subset B_\tau$, then

$$\bigcap_{k=1}^\infty A_k \subset B_\tau \cap B_n$$

and hence $\bigcap_{k=1}^\infty A_k \in B_\tau$. This completes the proof the $B_\tau$ is a $\sigma$-algebra.

3. For $n, k \in \mathbb{N}_0$ we have

$$\{ \tau = n \} \cap \{ \tau = k \} = \begin{cases} 0 & \text{if } k \neq n \\ \{ \tau = k \} & \text{if } k = n \end{cases}$$

and so in any case $\{ \tau = n \} \cap \{ \tau = k \} \in B_k$. This show $\{ \tau = n \} \in B_\tau$ for all $n \in \mathbb{N}_0$ and therefore $\tau$ is $B_\tau$-measurable.

4. For all $n \in \mathbb{N}_0$, $\{ \tau = n \} \in B_n$ as $\tau$ is a stopping time and $\{ \tau = n \} \in B_\tau$. By the second assertion, $A \in B_\tau \cap B_n$, and $A \subset \{ \tau = n \}$ which implies $A \subset A \cap \{ \tau = n \} \in B_n$ and so $A \in (B_n)_{\{ \tau = n \}}$. Conversely, if $A \in (B_n)_{\{ \tau = n \}}$, then $A \in B_n$ and $A \subset \{ \tau = n \}$. Thus for $k \in \mathbb{N}_0$ we have

$$A \cap \{ \tau = k \} = \begin{cases} \emptyset & \text{if } k \neq n \\ A & \text{if } k = n \end{cases}$$

from which it follows that $A \cap \{ \tau = k \} \in B_k$ for all $k \in \mathbb{N}_0$, i.e. $A \in [B_\tau]_{\{ \tau = n \}}$.

5. Now suppose $\sigma \leq \tau$ and $A \in B_\sigma$. Since $A \cap \{ \sigma \leq n \} \subset \{ \tau \leq n \}$ for all $n \leq \infty$, we find

$$A \cap \{ \tau \leq n \} = [A \cap \{ \sigma \leq n \}] \cap \{ \tau \leq n \} \in B_n \forall n \leq \infty$$

which shows $A \in B_\tau$.

\begin{flushright}
\hfill $\blacksquare$
\end{flushright}

Proposition 21.30 ($B_\tau$-measurable random variables). Let $(\Omega, B, \{B_n\}_{n=0}^\infty)$ 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 $B_\tau$-measurable,

2. $1_{\{\tau \leq n\}} Z$ is $B_n$-measurable for all $n \leq \infty$,

3. $1_{\{\tau=n\}} Z$ is $B_n$-measurable for all $n \leq \infty$.

4. There exists, $Y_n : \Omega \to \mathbb{R}$ which are $B_n$-measurable for all $n \leq \infty$ such that

$$Z = Y_\tau = \sum_{n \in \mathbb{N}} 1_{\{\tau=n\}} Y_n.$$ 

Proof. 1. $\implies$ 2. By definition, if $A \in B_\tau$, then $1_{\{\tau \leq n\}} 1_A = 1_{\{\tau \leq n\} \cap A}$ is $B_n$-measurable for all $n \leq \infty$. Consequently any simple $B_\tau$-measurable function, $Z$, satisfies $1_{\{\tau \leq n\}} Z$ is $B_n$-measurable for all $n$. So by the usual limiting argument (Theorem 9.39), it follows that $1_{\{\tau \leq n\}} Z$ is $B_n$-measurable for all $n$ for any $B_\tau$-measurable function, $Z$.

2. $\implies$ 3. This property follows from the identity,

$$1_{\{\tau=n\}} Z = 1_{\{\tau \leq n\}} Z - 1_{\{\tau < n\}} Z.$$ 

3. $\implies$ 4. Simply take $Y_n = 1_{\{\tau=n\}} Z$.

4. $\implies$ 1. Since $Z = \sum_{n \in \mathbb{N}} 1_{\{\tau=n\}} Y_n$, it suffices to show $1_{\{\tau=n\}} Y_n$ is $B_n$-measurable if $Y_n$ is $B_n$-measurable. Further, by the usual limiting arguments using Theorem 9.39, it suffices to assume that $Y_n = 1_A$ for some $A \in B_n$. In this case $1_{\{\tau=n\}} Y_n = 1_{A \cap \{\tau=n\}}$. Hence we must show $A \cap \{\tau = n\} \in B_n$ which indeed is true because

$$A \cap \{\tau = n\} \cap \{\tau = k\} = \begin{cases} \emptyset & \text{if } k \neq n \\ A \cap \{\tau = n\} \in B_k & \text{if } k = n \end{cases}$$

Alternatively proof for 1. $\implies$ 2. If $Z$ is $B_\tau$ measurable, then $\{Z \in B\} \cap \{\tau \leq n\} \in B_n$ for all $n \leq \infty$ and $B \in B_\infty$. Hence if $B \in B_k$ with $0 \notin B$, then

$$\{1_{\{\tau \leq n\}} Z \in B\} = \{Z \in B\} \cap \{\tau \leq n\} \in B_n$$

and similarly,
Exercise 21.6. Suppose $\tau$ is a stopping time, $(S, S)$ is a measurable space, and $Z : \Omega \to S$ is a function. Show that $Z$ is $B_\tau / S$ measurable iff $Z|_{(\tau = n)}$ is $(B_n)_{(\tau = n)} / S$ measurable for all $n \in \mathbb{N}_0$.

Lemma 21.31 ($B_\sigma$ – conditioning). Suppose $\sigma$ is a stopping time and $Z \in L^1(\Omega, B, P)$ or $Z \geq 0$, then
\[
E[Z|B_\sigma] = \sum_{n \leq \infty} 1_{\tau = n} E[Z|B_n] = Y_\sigma
\]  
(21.15)

where
\[
Y_n := E[Z|B_n] \text{ for all } n \in \mathbb{N}.
\]  
(21.16)

Proof. From Lemma 21.29, we know $B_\sigma = B_n$ on $\{\sigma = n\}$ and therefore by the localization Lemma 17.29,
\[
1_{\{\sigma = n\}} E_{B_\sigma} Z = 1_{\{\sigma = n\}} E_{B_n} Z \text{ a.s.}
\]
Summing this equation on $n$ shows
\[
E_{B_\sigma} Z = \sum_{n \leq \infty} 1_{\{\sigma = n\}} E_{B_n} Z = Y_\sigma.
\]

Alternative direct proof. By Proposition 21.30, $Y_\sigma$ is $B_\sigma$ – measurable. Moreover, if $Z$ is integrable, then
\[
\sum_{n \leq \infty} E\left[1_{\{\sigma = n\}} |Y_n|\right] = \sum_{n \leq \infty} E\left[1_{\{\sigma = n\}} E[Z|B_n]\right]
\]
\[
\leq \sum_{n \leq \infty} E\left[1_{\{\sigma = n\}} E[Z|B_n]\right]
\]
\[
= \sum_{n \leq \infty} E\left[E\left[1_{\{\sigma = n\}} |Z|\right]|B_n\right]
\]
\[
= \sum_{n \leq \infty} E\left[1_{\{\sigma = n\}} |Z|\right] = E[Z] < \infty
\]
(21.17)

and therefore
\[
E[Y_\sigma] = E\left[\sum_{n \leq \infty} 1_{\{\sigma = n\}} Y_n\right]
\]
\[
\leq \sum_{n \leq \infty} E\left[1_{\{\sigma = n\}} |Y_n|\right] \leq E[Z] < \infty.
\]

Furthermore if $A \in B_\sigma$, then
\[
E[Z : A] = \sum_{n \leq \infty} E[Z : A \cap \{\sigma = n\}] = \sum_{n \leq \infty} E[Y_n : A \cap \{\sigma = n\}]
\]
\[
= \sum_{n \leq \infty} E[1_{\{\sigma = n\}} Y_n : A] = E\left[\sum_{n \leq \infty} 1_{\{\sigma = n\}} Y_n : A\right]
\]
\[
= E[Y_\sigma : A],
\]

wherein the interchange of the sum and the expectation in the second to last equality is justified by the estimate in 21.17 or by the fact that everything in sight is positive when $Z \geq 0$.

Theorem 21.32 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 21.7. Suppose $\sigma$ and $\tau$ are two stopping times. Show:

1. $\{\sigma < \tau\}$, $\{\sigma = \tau\}$, and $\{\sigma \leq \tau\}$ are all in $B_\sigma \cap B_\tau$,
2. $B_{\sigma \wedge \tau} = B_\sigma \cap B_\tau$,
3. $B_{\sigma \vee \tau} = B_\sigma \vee B_\tau := \sigma (B_\sigma \cup B_\tau)$ and
4. $B_\sigma = B_{\sigma \wedge \tau}$ on $C$ where $C$ is any one of the following three sets: $\{\sigma \leq \tau\}$, $\{\sigma < \tau\}$, or $\{\sigma = \tau\}$.

As an example, since
\[
\{\sigma \leq \tau\} \cap \{\sigma \wedge \tau = n\} = \{\sigma \leq \tau\} \cap \{\sigma = n\} = \{n \leq \tau\} \cap \{\sigma = n\} \in B_n
\]
for all $n \in \mathbb{N}_0$, it follows that $\{\sigma \leq \tau\} \in B_\sigma \cap B_\tau$.

Theorem 21.32 (Tower Property II). Let $X \in L^1(\Omega, B, P)$ or $X : \Omega \to [0, \infty]$ be a $B$ – measurable function. If $\sigma$ and $\tau$ are any two stopping times, then
\[
1_{\tau \leq \sigma} E_{B_\tau} = 1_{\tau \leq \sigma} E_{B_{\sigma \wedge \tau}}, \quad 1_{\tau > \sigma} E_{B_\tau} = 1_{\tau > \sigma} E_{B_{\sigma \wedge \tau}}, \quad \text{and}
\]
\[
E_{B_\tau} E_{B_\sigma} X = E_{B_\sigma \wedge \tau} X = E_{B_{\sigma \wedge \tau}} X.
\]
(21.18)
(21.19)

Proof. As usual it suffices to consider the case where $X \geq 0$ and this case there will be now convergence issues to worry about. Equation 21.18 follows from the localization Lemma 17.29 as explained in the second proof. Nevertheless let us first give a self-contained proof.

First Proof. Notice that
\footnote{In fact, you will likely show in your proof that every set in $B_\sigma \vee B_\tau$ may be written as a disjoint union of a set from $B_\sigma$ with a set from $B_\tau$.
}
and similarly,
\[ 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}} = 1_{\tau > \sigma} \sum_{n \leq \infty} 1_{\tau = n} \mathbb{E}_{B_{\sigma}} = 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma} \wedge \tau} \]

Using these remarks and the fact that \( \{ \tau \leq \sigma \} \) and \( \{ \tau > \sigma \} \) are both in \( B_{\sigma \wedge \tau} = B_{\sigma} \cap B_{\tau} \), we find:

\[
\begin{align*}
\mathbb{E}_{B_{\sigma}} \mathbb{E}_{B_{\tau}} &= \mathbb{E}_{B_{\sigma}} (1_{\tau \leq \sigma} + 1_{\tau > \sigma}) \mathbb{E}_{B_{\tau}} = \mathbb{E}_{B_{\sigma}} 1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau}, \sigma} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}} \mathbb{E}_{B_{\tau}} \\
&= 1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau}, \sigma} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau} \\
&= 1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau} \wedge \sigma} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma} \wedge \tau} = \mathbb{E}_{B_{\sigma \wedge \tau}}.
\end{align*}
\]

Second Proof. In this proof we are going to make use of the localization Lemma [17.29] Since \( B_{\sigma \wedge \tau} \subset B_{\sigma} \), it follows by item 4. of Exercise [21.7] that \( B_{\sigma} = B_{\sigma \wedge \tau} \) on \( \{ \sigma \leq \tau \} \) and on \( \{ \sigma < \tau \} \). We will actually use the first statement in the form, \( B_{\sigma} = B_{\sigma \wedge \tau} \) on \( \{ \tau \leq \sigma \} \). From Lemma [21.31] we have

\[
\begin{align*}
1_{\tau \leq \sigma} \mathbb{E}_{B_{\sigma}} &= 1_{\tau \leq \sigma} \mathbb{E}_{B_{\sigma}, \sigma} \\
1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}} &= 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau}.
\end{align*}
\]

Using these relations and the basic properties of conditional expectation we arrive at

\[
\begin{align*}
\mathbb{E}_{B_{\sigma}} \mathbb{E}_{B_{\tau}} X &= \mathbb{E}_{B_{\sigma}} \mathbb{E}_{B_{\tau}} [1_{\tau \leq \sigma} \mathbb{E}_{B_{\sigma}} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau} X] \\
&= \mathbb{E}_{B_{\sigma}} [1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau}, \sigma} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau} X] \\
&= \mathbb{E}_{B_{\sigma}} [1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau}, \sigma} \mathbb{E}_{B_{\sigma}, \sigma} + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau} X] \\
&= 1_{\tau \leq \sigma} \mathbb{E}_{B_{\tau}, \sigma} \mathbb{E}_{B_{\sigma}, \sigma} X + 1_{\tau > \sigma} \mathbb{E}_{B_{\sigma}, \tau} X = \mathbb{E}_{B_{\sigma \wedge \tau}} X \text{ a.s.}
\end{align*}
\]

Exercise 21.8. Show, by example, that it is not necessarily true that

\[ \mathbb{E}_{G_{1}} \mathbb{E}_{G_{2}} = \mathbb{E}_{G_{1} \wedge G_{2}} \]

for arbitrary \( G_{1} \) and \( G_{2} \) – sub-sigma algebras of \( B \).

Hint: It suffices to take \((\Omega, B, P)\) with \( \Omega = \{1, 2, 3\} \) and \( B = 2^{\Omega} \) and \( P \{1\} = \frac{1}{3} \) for \( j = 1, 2, 3 \).

Exercise 21.9 (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 \( P_{1} = P_{H_{1}} \) denote orthogonal projection onto \( H_{1} \) 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 \( M_{1} = \{ h \in H_{1} : h \perp M_{0} \} \) so that \( H_{1} = M_{0} \oplus M_{1} \) and \( H_{2} = M_{0} \oplus M_{2} \). Then \( M_{0} = H_{1} \cap H_{2} \) and \( P_{1} P_{2} = P = P_{2} P_{1} \).

2. If \( P_{1} P_{2} = P_{2} P_{1} \), then \( P_{1} P_{2} = P = P_{2} P_{1} \). Moreover if we let \( M_{0} = H_{1} \cap H_{2} \) and \( M_{1} \) be as above, then \( M_{1} \perp M_{2} \).

Exercise 21.10. Let \( \sigma \) and \( \tau \) be stopping times and apply the results of Exercise [21.9] with \( M_{0} := L^{2}(\Omega, B_{\sigma \wedge \tau}, P) \), \( H_{1} := L^{2}(\Omega, B_{\sigma}, P) \), and \( H_{2} = L^{2}(\Omega, B_{\tau}, P) \) to give another proof of Theorem [21.32].

21.3.1 Summary of some of the more notable Chapter 21 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 \( M_{n} = \mathbb{E}[X|B_{n}] \) 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 [21.56].

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 \( X_{n} \geq 0 \) for all \( n \). Then for every stopping time, \( \tau \), \( X^{\tau} \) is a \( \{B_{n}\}_{n=0}^{\infty} \) – supermartingale, martingale, or submartingale respectively, see the optional stopping Theorem [21.39].

3. Suppose that \( \sigma \) and \( \tau \) are two stopping times and \( \sigma \leq \tau \) 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

\[ \mathbb{E}[X_{\tau}|B_{\sigma}] \leq_{\mathrm{a.s.}} X_{\tau \wedge \sigma} \]

4. Suppose that \( M := \{M_{n}\}_{n=0}^{\infty} \) is an \( L^{1} \) – bounded martingale then (see Theorem [21.67]) the following are equivalent;
\begin{enumerate}
  \item \( M := \{M_{n}\}_{n=0}^{\infty} \) is a regular martingale.
  \item \( M_{n} = \mathbb{E}[M_{\infty}|B_{n}] \) for \( n \in \mathbb{N} \).
  \item \( M_{n} \to M_{\infty} \) in \( L^{1}(\Omega, B, P) \).
  \item \( \{M_{n}\}_{n=0}^{\infty} \) is uniformly integrable.
\end{enumerate}

5. Suppose that \( M := \{M_{n}\}_{n=0}^{\infty} \) 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 [21.69].
6. Suppose that $M = \{M_t\}_{t=0}^{\infty}$ is a regular martingale, $\sigma$ and $\tau$ are arbitrary stopping times, then (see the optional stopping Theorem \[21.70\])

$$
M_\tau = \mathbb{E}[M_{\infty}|B_\tau], \quad \mathbb{E}[M_\tau] \leq \mathbb{E}[M_{\infty}] < \infty \quad \text{and} \quad \mathbb{E}[M_\tau|B_\sigma] = M_{\sigma \wedge \tau} \ a.s.
$$

7. Let $\{B_n : n \leq 0\}$ be a reverse filtration (still have $B_n \subset B_0$ if $m \leq n$) and $\{X_n\}_{n \leq 0}$ be a backwards submartingale, i.e. $\mathbb{E}[X_n|B_m] \geq X_m$ if $m < n$. The Backwards (or reverse) submartingale convergence theorem \[21.79\] 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

$$
C := \lim_{n \to -\infty} \mathbb{E}X_n = \inf_{n \leq 0} \mathbb{E}X_n > -\infty,
$$

then $X_n = M_n + A_n$ where

a) $\{M_n\}_{-\infty < n \leq 0}$ is a martingale, $\{A_n\}_{-\infty < n \leq 0}$ is a predictable process such that $A_{-\infty} = \lim_{n \to -\infty} A_n = 0$;

b) $\{X_n\}_{n \leq 0}$ is uniformly integrability;

c) $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$, and

d) $\lim_{n \to -\infty} \mathbb{E}|X_n - X_{-\infty}| = 0$.

### 21.4 Stochastic Integrals and Optional Stopping

**Notation 21.33** Suppose that $\{c_n\}_{n=1}^{\infty}$ and $\{x_n\}_{n=0}^{\infty}$ are two sequences of numbers, let $(c \cdot \Delta x)_n = \{(c \cdot \Delta x)_n\}_{n \in \mathbb{N}_0}$ denote the sequence of numbers defined by $(c \cdot \Delta x)_0 = 0$ and

$$
(c \cdot \Delta x)_n = \sum_{j=1}^{n} c_j (x_j - x_{j-1}) = \sum_{j=1}^{n} c_j \Delta_j x \quad \text{for} \quad n \geq 1.
$$

(For convenience of notation later we will interpret $\sum_{j=0}^{0} c_j \Delta_j x = 0.$)

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 $c_k = 5 \cdot 1_{3 \leq k \leq 9}$ so that

$$
(c \cdot \Delta x)_9 = 5 \cdot \sum_{3 \leq k \leq 9} \Delta_k x = 5 \cdot (x_9 - x_3)
$$

would represent your profit (loss) for this transaction. The next example formalizes this observation.

**Example 21.34.** Suppose that $0 \leq \sigma \leq \tau$ where $\sigma, \tau \in \mathbb{N}_0$ and let $c_n := 1_{\sigma < n \leq \tau}$. Then

$$
(c \cdot \Delta x)_n = \sum_{j=1}^{n} 1_{\sigma < j \leq \tau} (x_j - x_{j-1}) = \sum_{j=1}^{\infty} 1_{\sigma < j \leq \tau \wedge n} (x_j - x_{j-1})
$$

$$
= \sum_{j=1}^{\infty} 1_{\sigma \wedge n < j \leq \tau \wedge n} (x_j - x_{j-1}) = x_{\tau \wedge n} - x_{\sigma \wedge n}.
$$

More generally if $\sigma, \tau \in \mathbb{N}_0$ are arbitrary and $c_n := 1_{\sigma < n \leq \tau}$ we will have $c_n := 1_{\sigma \wedge n < \tau \wedge n}$ and therefore

$$
(c \cdot \Delta x)_n = x_{\tau \wedge n} - x_{\sigma \wedge n}.
$$

**Proposition 21.35 (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 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 $C_n \geq 0$, 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

$$
\mathbb{E}[|C \cdot \Delta X_n|] \leq \mathbb{E}[X_n] < \infty
$$

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

$$
\mathbb{E}[\{(C \cdot \Delta X)_{n+1}\}|B_n] = \mathbb{E}[\{(C \cdot \Delta X)_n + C_{n+1} (X_{n+1} - X_n)\}|B_n]
$$

$$
= (C \cdot \Delta X)_n + C_{n+1} \mathbb{E}[(X_{n+1} - X_n)|B_n].
$$

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, $\{C_n\}_{\infty}$. Taking expectations of Eq. \[21.21\] then allows us to conclude that

$$
\mathbb{E}[C_{n+1} \mathbb{E}[(X_{n+1} - X_n)|B_n]] = 0
$$

for all bounded $B_n$ – measurable random variables, $C_{n+1}$. Taking $C_{n+1} := \text{sgn}(\mathbb{E}[(X_{n+1} - X_n)|B_n])$ shows $\mathbb{E}[\{(X_{n+1} - X_n)|B_n]\] = 0$ a.s. and hence $X$ is
a martingale. Similarly, if for all non-negative, predictable $C$, Eq. (21.20) holds for all $n \geq 1$, and $C_n \geq 0$, then taking $A \in \mathcal{B}_n$ and $C_k = \delta_{k,n+1} A_A$ in Eq. (21.13) allows us to conclude that

$$\mathbb{E} [X_{n+1} - X_n : A] = \mathbb{E} [(C \cdot \Delta X)_{n+1}] \leq 0,$$

i.e. $X$ is a supermartingale, martingale, or submartingale respectively.

**Example 21.36.** Suppose that $\{X_n\}_{n=0}^\infty$ are 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

$$Y_n := \sum_{k=1}^n f_k (X_0, \ldots, X_{k-1}) (X_k - X_{k-1}) \quad \text{for } n \in \mathbb{N},$$

(21.22)
is a martingale sequence relative to $\{\mathcal{B}_n^X\}_{n \geq 0}$.

**Notation 21.37.** Given an adapted process, $X$, and a stopping time $\tau$, let $X^\tau_n := X_{\tau \wedge n}$. We call $X^\tau := \{X^\tau_n\}_{n=0}^\infty$ the process $X$ stopped by $\tau$.

Observe that

$$|X^\tau_n| = |X_{\tau \wedge n}| = \sum_{0 \leq k \leq n} 1_{\tau = k} |X_k| \leq \sum_{0 \leq k \leq n} 1_{\tau = k} |X_k| \leq \sum_{0 \leq k \leq n} |X_k|,$$

so that $X^\tau_n \in L^1(\mathbb{P})$ for all $n$ provided $X_n \in L^1(\mathbb{P})$ for all $n$.

**Example 21.38.** 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

$$A \cap \{\sigma < n \leq \tau\} = (A \cap \{\sigma < n\} \cap \{n \leq \tau\}$$

$$= (A \cap \{\sigma < n - 1\} \cap \{n \leq \tau\} \cap \{\tau \leq n - 1\}) \in \mathcal{B}_{n-1}.$$ 

Therefore by Proposition 21.35, $\{(C \cdot \Delta X)_n\}_{n=0}^\infty$ is a supermartingale, martingale, or submartingale respectively where

$$(C \cdot \Delta X)_n = \sum_{k=1}^n 1_A \cdot 1_{\sigma < k \leq \tau} \Delta_k X = 1_A \sum_{k=1}^n 1_{\sigma \wedge \tau < k \leq \tau} \Delta_k X$$

$$= \sum_{k=1}^\infty 1_A \cdot 1_{\sigma \wedge \tau \wedge n < k \leq \tau \wedge n} \Delta_k X = 1_A (X^\tau_n - X^\sigma_n \wedge \tau_n).$$

**Theorem 21.39 (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 21.38 in order to learn that $\{X^\tau_n - X_0\}_{n=0}^\infty$ is a supermartingale, martingale, or submartingale respectively and therefore so is $\{X^\tau_n = X_0 + X^\tau_n - X_0\}_{n=0}^\infty$. 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

$$\mathbb{E}_{\mathcal{B}_n} [X^\tau_{n+1}] = \mathbb{E}_{\mathcal{B}_n} [X_{\tau \wedge (n+1)}] = 1_{\tau \leq n} \mathbb{E}_{\mathcal{B}_n} X_\tau + 1_{\tau > n} \mathbb{E}_{\mathcal{B}_n} X_{n+1}$$

$$\{\leq \} 1_{\tau \leq n} \mathbb{E}_{\mathcal{B}_n} X_\tau + 1_{\tau > n} \mathbb{E}_{\mathcal{B}_n} X_{n+1} = X_{\tau \wedge n},$$

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 21.41.)

**Theorem 21.40 (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

$$\mathbb{E} [X_\tau | \mathcal{B}_\sigma] \leq X_{\sigma \wedge \tau} \text{ a.s.}$$

(21.23)

respectively 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 21.38 we know that $1_A (X^\tau_n - X^\sigma_n \wedge \tau_n)$ is a supermartingale, martingale, or submartingale respectively and in particular for all $n \in \mathbb{N}_0$ we have

$$\mathbb{E} [1_A (X^\tau_n - X^\sigma_n \wedge \tau_n)] \leq 0$$

respectively.

\footnote{This is the natural generalization of Eq. (21.13) to the stopping time setting.}
Combining this result with Lemma 21.31 then implies
\[ E[(X_\tau - X_{\sigma \wedge \tau}) : A] \leq 0 \]
and this is equivalent to Eq. (21.23).

When we only assume that \( X_n \geq 0 \) 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 21.39 in order to conclude that \( X_\tau \) is a supermartingale, martingale, or submartingale respectively and in particular that
\[ E[X_\tau | B_n] = E[X_N^+ | B_n] \leq X_{\tau \wedge N} \text{ for all } n \leq \infty. \]

Combining this result with Lemma 21.31 then implies
\[ E[X_\tau | B_n] = \sum_{n \leq \infty} 1_{\sigma = n} E[X_\tau | B_n] \leq \sum_{n \leq \infty} 1_{\sigma = n} X_{\tau \wedge N} = X_{\tau \wedge N}. \quad (21.24) \]

(This second proof again covers both cases at once!)

**Exercise 21.11.** Give another proof of Theorem 21.40 when \( E[X_n] < \infty \) by using the tower property in Theorem 21.32 along with the Doob decomposition of Lemma 21.17.

**Exercise 21.12.** Give yet another (full) proof of Theorem 21.40 using the following outline:

1. Show by induction on \( n \) starting with \( n = N \) that
\[ E[X_\tau | B_n] \leq X_{\tau \wedge N} \text{ a.s. for all } 0 \leq n \leq N. \quad (21.25) \]

2. Observe the above inequality holds as an equality for \( n > N \) as well.

3. Combine this result with Lemma 21.17 to complete the proof.

This argument makes it clear why we must at least initially assume that \( \tau \leq N \) for some \( N \in \mathbb{N} \). To relax this restriction will require a limiting argument which will be the topic of Section 21.8 below.

**Remark 21.41.** Theorem 21.40 can be used to give a simple proof of the Optional Stopping Theorem 21.39. For example, if \( X = \{X_n\}_{n=0}^\infty \) is a submartingale and \( \tau \) is a stopping time, then
\[ E[B_n X_{\tau \wedge (n+1)}] \geq E[X_{\tau \wedge (n+1)}] = X_{\tau \wedge n}, \]
i.e. \( X_\tau \) is a submartingale.

### 21.5 Submartingale Maximal Inequalities

**Notation 21.42 (Running Maximum)** If \( X = \{X_n\}_{n=0}^\infty \) is a sequence of (extended) real numbers, we let
\[ X_N^* := \max \{X_0, \ldots, X_N\}. \quad (21.26) \]

**Proposition 21.43 (Maximal Inequalities of Bernstein and Lévy).** Let \( \{X_n\} \) be a submartingale on a filtered probability space, \( (\Omega, \mathcal{F}, \{B_n\}_{n=0}^\infty, P) \). Then for any \( a \geq 0 \) and \( N \in \mathbb{N} \),
\[ a P(X_N^* \geq a) \leq E[X_N^* \geq a] \leq E[X_N^+] \quad (21.27) \]
and
\[ a P(\min_{n \leq N} X_n \leq -a) \leq E[X_N^* \min_{k \leq N} X_k > -a] - E[X_0] \quad (21.28) \]
and
\[ a P(|X_N^*| \geq a) \leq 2E[X_N^+] - E[X_0]. \quad (21.30) \]

**Proof.** Let \( \tau := \inf \{n : X_n \geq a\} \) and observe that
\[ X_N^* \geq X_\tau \geq a \text{ on } \{\tau \leq N\} = \{X_N^* \geq a\} \]
and (by the optional sampling Theorem 21.40)
\[ E[X_N | B_\tau] \geq X_N^* \quad (21.31) \]

where \( \tau \leq N \in \mathbb{N} \). Since \( \{\tau \leq N\} \in B_{\tau \wedge N} \subseteq \mathcal{F}_N \),
\[ E[X_N : \tau \leq N] = E[X_N^* : \tau \leq N] \leq E[E[X_N | B_\tau] : \tau \leq N] = E[X_N : \tau \leq N] \]
which combined with Eq. (21.31) implies,
\[ a \cdot P(X_N^* \geq a) = E[a : X_N^* \geq a] = E[a : \tau \leq N] \leq E[X_\tau : \tau \leq N] \leq E[X_N^* : \tau \leq N] = E[X_N^* : X_N^* \geq a] \leq E[X_N^* : X_N^* \geq a] \leq E[X_N^*]. \]

i.e. Eq. (21.27) holds.

**More generally** if \( X \) is any integrable process and \( \tau \) is the random time defined by, \( \tau := \inf \{n : X_n \geq a\} \), we still have Eq. (21.31) and
\[ a P(X_N^* \geq a) = E[a : \tau \leq N] \leq E[X_\tau : \tau \leq N] = E[X_N^* : X_N^* \geq a] \leq E[X_N^* : X_N^* \geq a] \leq E[X_N^*]. \quad (21.33) \]

\(^4\) The first inequality is the most important.
Let me emphasize again that in deriving Eq. (21.33), 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 (21.40), \( \mathbb{E}_{\tau \wedge N} X_N \geq X_{\tau \wedge N} \) and in particular \( \mathbb{E} [X_N - X_{\tau \wedge N}] \geq 0 \). Combining this observation with Eq. (21.33) and Eq. (21.31) again gives Eq. (21.27).

Secondly we may apply Eq. (21.33) with \( X_n \) replaced by \(-X_n\) to find
\[
aP \left( \min_{n \leq N} X_n \leq -a \right) = aP \left( -\min_{n \leq N} X_n \geq a \right) = aP \left( \max_{n \leq N} (-X_n) \geq a \right)
\leq -\mathbb{E} [X_N : \tau \leq N] + \mathbb{E} [X_N - X_{\tau \wedge N}] \tag{21.34}
\]
where now,
\[
\tau := \inf \{ n : X_n \geq a \} = \inf \{ n : X_n \leq -a \}.
\]
By the optional sampling Theorem (21.40), \( \mathbb{E} [X_{\tau \wedge N} - X_0] \geq 0 \) and adding this to the right side of Eq. (21.34) gives the estimate
\[
aP \left( \min_{n \leq N} X_n \leq -a \right) \leq -\mathbb{E} [X_N : \tau \leq N] + \mathbb{E} [X_N - X_{\tau \wedge N}] + \mathbb{E} [X_{\tau \wedge N} - X_0]
\leq \mathbb{E} [X_N - X_0] - \mathbb{E} [X_N : \tau \leq N]
= \mathbb{E} [X_N : \tau > N] - \mathbb{E} [X_0]
= \mathbb{E} \left[ X_N : \min_{k \leq N} X_k > -a \right] - \mathbb{E} [X_0]
\]
which proves Eq. (21.28) and hence Eq. (21.29). Adding Eqs. (21.27) and (21.29) gives the estimate in Eq. (21.30) since
\[
\{|X|^*_N \geq a\} = \{X^*_N \geq a\} \cup \left\{ \min_{n \leq N} X_n \leq -a \right\}.
\]

Remark 21.44. It is of course possible to give a direct proof of Proposition 21.43. For example,
\[
\mathbb{E} \left[ X_N : \max_{n \leq N} X_n \geq a \right] = \sum_{k=1}^{N} \mathbb{E} [X_N : X_1 < a, \ldots, X_{k-1} < a, X_k \geq a]
\geq \sum_{k=1}^{N} \mathbb{E} [X_k : X_1 < a, \ldots, X_{k-1} < a, X_k \geq a]
\geq \sum_{k=1}^{N} \mathbb{E} [a : X_1 < a, \ldots, X_{k-1} < a, X_k \geq a]
= aP \left( \max_{n \leq N} X_n \geq a \right)
\]
which proves Eq. (21.27).

Corollary 21.45. Suppose that \( \{Y_n\}_{n=1}^{\infty} \) is a non-negative supermartingale, \( a > 0 \) and \( N \in \mathbb{N} \), then
\[
aP \left( \max_{n \leq N} Y_n \geq a \right) \leq \mathbb{E} [Y_0 \wedge a] - \mathbb{E} \left[ Y_N : \max_{n \leq N} Y_n < a \right] \leq \mathbb{E} [Y_0 \wedge a]. \tag{21.35}
\]

Proof. Let \( X_n := -Y_n \) in Eq. (21.28) to learn
\[
aP \left( \min_{n \leq N} (-X_n) \leq -a \right) \leq \mathbb{E} \left[ -Y_N : \min_{n \leq N} (-Y_n) > -a \right] + \mathbb{E} [Y_0]
\]
or equivalently that
\[
aP \left( \max_{n \leq N} Y_n \geq a \right) \leq \mathbb{E} [Y_0] - \mathbb{E} \left[ Y_N : \max_{n \leq N} Y_n < a \right] \leq \mathbb{E} [Y_0]. \tag{21.36}
\]
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. (21.36) with \( Y_n \) replaced by \( Y_n \wedge a \) proves Eq. (21.35). \( \blacksquare \)

Lemma 21.46. 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 \( p \in (1, \infty) \),
\[
\mathbb{E} Y^p \leq \left( \frac{p}{p-1} \right)^p \mathbb{E} X^p. \tag{21.37}
\]

Proof. We will begin by proving Eq. (21.37) under the additional assumption that \( Y \in L^p (\Omega, \mathcal{B}, P) \). Since
\[
\mathbb{E} Y^p = p \int_0^\infty 1_{y \leq Y} \cdot y^{p-1} dy = p \int_0^\infty \mathbb{E} [1_{y \leq Y}] \cdot y^{p-1} dy
= p \int_0^\infty \mathbb{E} P (Y \geq y) \cdot y^{p-1} dy \leq p \int_0^\infty \frac{1}{y} \mathbb{E} [X : Y \geq y] \cdot y^{p-1} dy
= p \mathbb{E} \int_0^\infty X 1_{y \leq Y} \cdot y^{p-2} dy = \left( \frac{p}{p-1} \right)^p \mathbb{E} [X Y^{p-1}]^p.
\]
Now apply Hölder’s inequality, with \( q = p (p-1)^{-1} \), to find
\[
\mathbb{E} [X Y^{p-1}] \leq \|X\|_p \cdot \|Y^{p-1}\|_q = \|X\|_p \cdot \|E [Y]^p\|^{1/q}.
\]
Combining the two inequalities shows and solving for \(\|Y\|_p\) shows \(\|Y\|_p \leq \frac{p}{p-1} \|X\|_p\), which proves Eq. \((21.37)\) under the additional restriction of \(Y\) being in \(L^p(\Omega, \mathcal{F}, P)\).

To remove the integrability restriction on \(Y\), for \(M > 0\) let \(Z := Y \wedge M\) and observe that

\[
P(Z \geq y) = P(Y \geq y) \leq \frac{1}{y} E[X; Y \geq y] = \frac{1}{y} E[X; Z \geq y] \quad \text{if } y \leq M
\]

while

\[
P(Z \geq y) = 0 = \frac{1}{y} E[X; Z \geq y] \quad \text{if } y > M.
\]

Since \(Z\) is bounded, the special case just proved shows

\[
E[(Y \wedge M)^p] = EZ^p \leq \left(\frac{p}{p-1}\right)^p EX^p.
\]

We may now use the MCT to pass to the limit, \(M \uparrow \infty\), and hence conclude that Eq. \((21.37)\) holds in general. □

**Corollary 21.47 (Doob’s Inequality).** If \(X = \{X_n\}_{n=0}^\infty\) be a non-negative submartingale and \(1 < p < \infty\), then

\[
E[X_N]^p \leq \left(\frac{p}{p-1}\right)^p E[X_N^p]. \quad (21.38)
\]

**Proof.** Equation \((21.38)\) follows by applying Lemma 21.46 with the aid of Proposition 21.43. □

**Corollary 21.48 (Doob’s Inequality).** If \(\{M_n\}_{n=0}^\infty\) is a martingale and \(1 < p < \infty\), then for all \(a > 0\),

\[
P(\{M_N^* \geq a\}) \leq \frac{1}{a} E[|M_N|; M_N^* \geq a] \leq \frac{1}{a} E[|M_N|] \quad (21.39)
\]

and

\[
E[M_N]^p \leq \left(\frac{p}{p-1}\right)^p E[M_N^p]. \quad (21.40)
\]

**Proof.** By the conditional Jensen’s inequality, it follows that \(X_n := |M_n|\) is a submartingale. Hence Eq. \((21.39)\) follows from Eq. \((21.27)\) and Eq. \((21.40)\) follows from Eq. \((21.38)\). □

**Example 21.49.** Let \(\{X_n\}\) be a sequence of independent integrable random variables with mean zero, \(S_0 = 0, S_n := X_1 + \cdots + X_n\) for \(n \in \mathbb{N}\), and \(|S_n^*| = \max_{j \leq n} |S_j|\). Since \(\{S_n\}_{n=0}^\infty\) is a martingale, by cJensen’s inequality, \(\{\|S_n\|_1\}_{n=1}^\infty\) is a (possibly extended) submartingale for any \(p \in [1, \infty)\). Therefore an application of Eq. \((21.27)\) of Proposition 21.43 show

\[
P(|S_N^*| \geq a) = p(\{S_N^* \geq a\}) \leq \frac{1}{a} E[|S_N|; S_N^* \geq a].
\]

(When \(p = 2\), this is Kolmogorov’s inequality in Theorem 23.46 below.) From Corollary 21.48 we also know that

\[
E[S_N]^p \leq \left(\frac{p}{p-1}\right)^p E[S_N^p].
\]

In particular when \(p = 2\), this inequality becomes,

\[
E[S_N]^2 \leq 4 \cdot E[S_N]^2 = 4 \cdot \sum_{n=1}^N E[X_n]^2.
\]

### 21.6 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 21.50.** 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

\[
W_n = W_0 + \sum_{k=1}^n C_k \Delta_k X_k.
\]

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 21.51 (“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

\[
E\left(\sum_{k=1}^n C_k \Delta_k X_k\right) \leq E[X_n - X_0]
\]

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 21.35,
\[
\mathbb{E} \left( \sum_{k=1}^{n} (1 - C_k) \Delta_k X \right) \geq 0.
\]
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
\[
\mathbb{E} [X_n - X_0] = \mathbb{E} \left( \sum_{k=1}^{n} C_k \Delta_k X \right) + \mathbb{E} \left( \sum_{k=1}^{n} (1 - C_k) \Delta_k X \right) \geq \mathbb{E} \left( \sum_{k=1}^{n} C_k \Delta_k X \right).
\]

We are now going to apply Lemma 21.51 to the time honored gambling strategy of buying low and selling high in order to prove the important “upcrossing” inequality of Doob, see Theorem 21.53. 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, \( N_0 \geq n \to X_n \in \mathbb{R} \) and \(-\infty < a < b < \infty\), let
\[
\tau_0 = \inf \{ n \geq 0 : X_n \leq a \}, \quad \tau_1 = \inf \{ n \geq \tau_0 : X_n \geq b \}
\]
\[
\tau_2 = \inf \{ n \geq \tau_1 : X_n \leq a \}, \quad \tau_3 = \inf \{ n \geq \tau_2 : X_n \geq b \}
\]
\[
\vdots
\]
\[
\tau_{2k} = \inf \{ n \geq \tau_{2k-1} : X_n \leq a \}, \quad \tau_{2k+1} = \inf \{ n \geq \tau_{2k} : X_n \geq b \} \quad (21.41)
\]
\[
\vdots
\]
with the usual convention that \( \inf \emptyset = \infty \) in the definitions above, see Figures 21.2 and 21.3.

In terms of these stopping time our betting strategy may be describe as,
\[
C_n = \sum_{k=0}^{\infty} 1_{\tau_{2k} < n \leq \tau_{2k+1}} \text{ for } n \in \mathbb{N}, \quad (21.42)
\]
see Figure 21.3 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 \( N \in \mathbb{N} \) let

\[
W_N := \sum_{k=1}^{N} C_k \Delta_k X \geq (b-a) \sum_{k=1}^{N} U^X_N (a,b) + (a-X_0)_+ - (a-X_0)_+ \quad (21.44)
\]
\[
= (b-a) \sum_{k=1}^{N} U^X_N (a,b) + (X_0-a)_+ - (X_0-a)_+. \quad (21.45)
\]

In words the inequality in Eq. (21.45) 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. 21.2. A sample path or the positive part of a random walk with level crossing of \( a=1 \) and \( b=2 \) being marked off.](image-url)
Here are the three cases to consider.

1. If \( X_n > a \) for all \( 0 \leq n \leq N \), 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 \( X_0 \leq a \) and \( X_n < b \) for all \( 0 \leq n \leq N \), then \( C_n = 1 \) for all \( n \) so that

\[
W_N = X_N - X_0 = (X_N - a) - (X_0 - a) = (X_N - a)_+ + (X_0 - a)_-.
\]

3. If \( X_0 > a \), but \( \tau_1 \leq N \) and \( X_n < b \) for all \( 0 \leq n \leq N \), then

\[
W_N = X_N - X_\tau_1 \geq X_N - a \geq - (X_N - a)_- = - (X_N - a)_- + (X_0 - a)_-.
\]

**Lemma 21.52.** 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, \( X_n \leq X_{n+1}^+ \) and \( 0 \leq X_{n+1}^+ \) and therefore,

\[
X_n \leq \mathbb{E}_{\mathcal{B}_n} X_{n+1} \leq \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+ \quad \text{and} \quad 0 \leq \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+.
\]

and therefore \( X_n^+ = \max\{0, X_n\} \leq \mathbb{E}_{\mathcal{B}_n} X_{n+1}^+ \). \( \Box \)

**Theorem 21.53 (Doob’s Upcrossing Inequality).** If \( \{X_n\}_{n=0}^\infty \) is a submartingale and \( -\infty < a < b < \infty \), then for all \( N \in \mathbb{N} \),

\[
\mathbb{E}[U_N^X(a, b)] \leq \frac{1}{b-a} \left[ \mathbb{E}(X_N - a)^+ - \mathbb{E}(X_0 - a)^+ \right] .
\]

**First Proof.** Let \( \{C_k\}_{k=1}^\infty \) be the buy low sell high strategy defined in Eq. \((21.42)\). Taking expectations of the inequality in Eq. \((21.45)\) making use of Lemma \(21.51\) implies,

\[
\mathbb{E}[X_N - a - (X_0 - a)] = \mathbb{E}[X_N - X_0] \geq \mathbb{E}[\{C \cdot \Delta X\}_N] \geq (b-a) \mathbb{E}[U_N^X(a, b)] + \mathbb{E}(X_0 - a)_- - \mathbb{E}(X_N - a)_-.
\]

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 \(21.52\) it follows that \( \{(X_n - a)^+\}_{n=0}^\infty \) is still a submartingale.\(^5\) 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 \( \{(X_n - a)^+\}_{n=0}^\infty \) then it is easily\(^6\) seen that

\[
W_N - W_0 \geq (b-a) U_N^{X-a}(0, b-a) = (b-a) U_N^X(a, b).
\]

Therefore it follows from Lemma \(21.51\) that

\[
(b-a) \mathbb{E}[U_N^X(a, b)] \leq \mathbb{E}[W_N - W_0] \leq \mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+.
\]

\(^5\) 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 21.54 (*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
\[
\tau_0 = \inf \{ n \geq 0 : X_n = 0 \}, \quad \tau_1 = \inf \{ n \geq \tau_0 : X_n \geq b \}
\]
\[
\tau_2 = \inf \{ n \geq \tau_1 : X_n = 0 \}, \quad \tau_3 = \inf \{ n \geq \tau_2 : X_n \geq b \}
\]
\[
\tau_{2k} = \inf \{ n \geq \tau_{2k-1} : X_n = 0 \}, \quad \tau_{2k+1} = \inf \{ n \geq \tau_{2k} : X_n \geq b \}
\]
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 \( \tau' = \tau_n \wedge N \), then \( \tau' = N \) for all \( n \geq 2k \). Therefore,
\[
X_N - X_0 = \sum_{n=1}^{2k} \left( X_{\tau'_n} - X_{\tau'_{n-1}} \right)
\]
\[
= \sum_{n=1}^{k} \left( X_{\tau_{2n-1}} - X_{\tau_{2n-2}} \right) + \sum_{n=1}^{k} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} \right)
\]
\[
\geq bU_N^X(0, b) + \sum_{n=1}^{k} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} \right), \tag{21.46}
\]
wherein we have used \( X_{\tau_{2n-1}} - X_{\tau_{2n-2}} \geq b \) if there were an upcrossing in the interval \( [\tau'_{2n-2}, \tau'_{2n-1}] \) and \( X_{\tau_{2n}} - X_{\tau_{2n-1}} \geq 0 \) otherwise, see Figure 21.4.

Taking expectations of Eq. (21.46) 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 21.53 \( \{(X_n - a)_+\}_{n=0}^{\infty} \) is still a sub-martingal and
\[
U_N^X(a, b) = U_N^{(X-a)_+}(0, b - a)
\]
and therefore
\[
(b - a) \mathbb{E} [U_N^X(a, b)] = (b - a) \mathbb{E} \left[ U_N^{(X-a)_+}(0, b - a) \right]
\]
\[
\leq \mathbb{E} (X_N - a)_+ - \mathbb{E} (X_0 - a)_+.
\]

**21.6 Submartingale Upcrossing Inequality and Convergence Theorems**

The third proof is now complete, nevertheless it is worth contemplating a bit how is that \( \mathbb{E} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} \right) \geq b \) and therefore,
\[
0 \leq \mathbb{E} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} \right)
\]
\[
= \mathbb{E} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} : \tau_{2n} \leq N \right) + \mathbb{E} \left( X_{\tau_{2n}} - X_{\tau_{2n-1}} : \tau_{2n} > N \right)
\]
\[
\leq - bP (\tau_{2n} \leq N) + \mathbb{E} \left( X_N - X_{\tau_{2n-1}} : \tau_{2n} > N \right).
\]

Therefore we must have
\[
\mathbb{E} \left( X_N - X_{\tau_{2n-1}} : \tau_{2n} > N \right) \geq bP (\tau_{2n} \leq N)
\]
so that \( X_N \) must be sufficiently large sufficiently often on the set where \( \tau_{2n} > N \).

Fig. 21.4. 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 \( X_{\tau_{70}} - X_{\tau_{70}} \geq 2 \) while \( X_{\tau_{70}} - X_{\tau_{70}} = 0 \).
Lemma 21.55. Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a sequence of extended real numbers such that $U_X^N (a, b) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$. Then $X_{\infty} := \lim_{n \to \infty} X_n$ exists in $\mathbb{R}$.

Proof. If $\lim_{n \to \infty} X_n$ does not exist in $\mathbb{R}$, 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, $U_X^N (a, b) = \infty$.

Corollary 21.56. 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 21.20), then $X_{\infty} := \lim_{n \to \infty} X_n$ exists in $\mathbb{R}$ a.s. and $X_{\infty} \in L^1 (\Omega, \mathcal{F}, \mathbb{P})$. Moreover $\{X_n\}_{n=0}^{\infty}$ is a submartingale (that is we also have $X_n \leq \mathbb{E} [X_\infty | \mathcal{B}_n]$ a.s. for all $n$, if $\{X_n^+\}_{n=1}^{\infty}$ is uniformly integrable.

Proof. For any $\infty < a < b < \infty$, by Doob’s upcrossing inequality (Theorem 21.53) and the MCT,

$$\mathbb{E} [U_X^N (a, b)] \leq \frac{1}{b-a} \left[ \sup_N \mathbb{E} (X_N - a)_+ - \mathbb{E} (X_0 - a)_+ \right] < \infty$$

where

$$U_X^N (a, b) := \lim_{N \to \infty} U_X^N (a, b)$$

is the total number of upcrossings of $X$ across $[a, b]$. In particular it follows that

$$\Omega_0 := \cap \{U_X^N (a, b) < \infty : a, b \in \mathbb{Q} \text{ with } a < b \}$$

has probability one. Hence by Lemma 21.55 for $\omega \in \Omega_0$ we have $X_{\infty} (\omega) := \lim_{n \to \infty} X_n (\omega)$ exists in $\mathbb{R}$. By Fatou’s lemma we know that

$$\mathbb{E} [\|X_{\infty}\|] = \mathbb{E} \left[ \liminf_{n \to \infty} |X_n| \right] \leq \liminf_{n \to \infty} \mathbb{E} [\|X_n\|] \leq C < \infty$$

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 $L^1 (\mathbb{P})$ by Vitalli’s convergence Theorem 21.51. Thus if $A \in \mathcal{B}_n$, then using $\mathbb{E} X_m \geq X_n$ a.s. for $m \geq n$ it follows that

$$\mathbb{E} [X_{\infty}^+] \leq \limsup_{m \to \infty} \mathbb{E} [X_m^+] = \limsup_{m \to \infty} \left( \mathbb{E} [X_m^+] - \mathbb{E} [X_{m-1}^-] \right) = \mathbb{E} [X_{\infty}^+] - \liminf_{m \to \infty} \mathbb{E} [X_{m-1}^-] \leq \mathbb{E} [X_{\infty}^+] - \mathbb{E} [X_{\infty}^-] = \mathbb{E} [X_{\infty}]$$

wherein Fatou’s lemma for the second inequality above. As $A \in \mathcal{B}_n$ was arbitrary, it now follows that $X_n \leq \mathbb{E} [X_\infty | \mathcal{B}_n]$ a.s. for $n$.

Conversely if we suppose that $X_n \leq \mathbb{E} [X_\infty | \mathcal{B}_n]$ a.s. for $n$, then by Lemma 21.52 (or cJensen’s inequality with $\varphi (x) = x \vee 0$ being an increasing convex function),

$$X_n^+ \leq \mathbb{E} [X_\infty | \mathcal{B}_n]^+ \leq \mathbb{E} [X_{n+1}^+ | \mathcal{B}_n] \text{ a.s. for all } n$$

and therefore $\{X_n^+\}_{n=1}^{\infty}$ is uniformly integrable by Proposition 21.8 and Exercise 15.5.

Second Proof. We may also give another proof of the first assertion based on the Krickeberg decomposition Theorem 21.21 and the supermartingale convergence Corollary 21.65 below. Indeed, by the Krickeberg decomposition Theorem 21.21 $X_n = M_n - Y_n$ where $M$ is a positive martingale and $Y$ is a positive supermartingale. Hence by two applications of Corollary 21.65 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 21.57. 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 21.7. Then $X_n$ is a negative (sub)martingle and hence $X_n^+ = 0$ 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

$$\lim_{n \to \infty} \mathbb{E} X_n = 1 \neq 0 = \mathbb{E} \lim_{n \to \infty} X_n$$

in this case.

Notation 21.58. Given a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$ and $A, B \in \mathcal{B}$, we say $A = B$ a.s. if $P(A \triangle B) = 0$ or equivalently if $1_A = 1_B$ a.s.

Corollary 21.59 (Localizing Corollary 21.56). Suppose $M = \{M_n\}_{n=0}^{\infty}$ is a martingale and $c < \infty$ such that $\Delta_n M := M_n - M_{n-1} \leq c$ a.s. for all $n$. Then

$$\left\{ \lim_{n \to \infty} M_n \text{ exists in } \mathbb{R} \right\} = \left\{ \sup_n M_n < \infty \right\} \text{ a.s.}$$
Corollary 21.60. 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
\[
C := \left\{ \lim_{n \to \infty} M_n \text{ exists in } \mathbb{R} \right\} \quad \text{and} \quad D := \left\{ \limsup_{n \to \infty} M_n = -\infty \text{ and } \liminf_{n \to \infty} M_n = -\infty \right\}.
\]
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 21.59, 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,
\[
C^c = \left\{ \sup_n M_n = \infty \right\} = \left\{ \inf_n M_n = -\infty \right\}
\]
\[
= \{ \sup_n M_n = \infty \} \cap \{ \inf_n M_n = -\infty \} = D.
\]

Corollary 21.61. Suppose \((\Omega, \mathcal{B}, \{B_n\}_{n=0}^{\infty}, P)\) is a filtered probability space and \( A_n \in \mathcal{B}_n \) for all \( n \). Then
\[
\left\{ \sum_n 1_{A_n} = \infty \right\} = \left\{ A_n \text{ i.o.} \right\} = \left\{ \sum_n \mathbb{E}[1_{A_n}|\mathcal{B}_{n-1}] = \infty \right\} \text{ a.s.} \quad (21.48)
\]
\footnote{If \( n < \tau_n \) then \( M_n < a \) and if \( n \geq \tau_n \) then \( M_n^{\tau_n} = M_{\tau_n} \leq M_{\tau_n-1} + c < a + c \).
Since \( \{ n < \tau \} \) and \( \{ n \geq \tau \} \) are \( B_n \) - measurable, it now follows from the supermartingale property of \( X \) and \( Y \) that
\[
E_{B_n} Z_{n+1} \leq E_{B_n} [1_{n<\tau} X_{n+1} + 1_{n\geq \tau} Y_{n+1}]
= 1_{n<\tau} E_{B_n} [X_{n+1}] + 1_{n\geq \tau} E_{B_n} [Y_{n+1}]
\leq 1_{n<\tau} X_n + 1_{n\geq \tau} Y_n = Z_n.
\]

\section{21.7.1 Maximal Inequalities}

\textbf{Theorem 21.63 (Supermartingale maximal inequality).} Let \( X \) be a positive supermartingale (in the extended sense) and \( a \in B_0 \) with \( a \geq 0 \), then
\[
 a P \left[ \sup_n X_n \geq a | B_0 \right] \leq a \wedge X_0 \tag{21.49}
\]
and moreover
\[
P \left[ \sup_n X_n = \infty | B_0 \right] = 0 \quad \text{on} \quad \{ X_0 < \infty \}. \tag{21.50}
\]

In particular if \( X_0 < \infty \) a.s. then \( \sup_n X_n < \infty \) a.s.

\textbf{Proof.} Simply apply Corollary 21.45 with \( Y_n = ((2a) \wedge X_n) \cdot 1_A \) where \( A \in B_0 \) to find
\[
a E \left( P \left[ \sup_n X_n \geq a | B_0 \right] : A \right) = a P \left[ \sup_n X_n \geq a : A \right] \leq E \left[ a \wedge X_0 : A \right].
\]

Since this holds for all \( A \in B_0 \), Eq. (21.49) follows.

\textbf{Second Proof.} Let \( \tau := \inf \{ n : X_n \geq a \} \) which is a stopping time since,
\[
\{ \tau \leq n \} = \{ X_n \geq a \} \in B_n \quad \text{for all} \quad n.
\]
Since \( X_\tau \geq a \) on \( \{ \tau < \infty \} \) and \( Y_\tau := a \) is a supermartingale, it follows by the switching Lemma 21.62 that
\[
 Z_n := 1_{n<\tau} X_n + a 1_{n\geq \tau}
\]
is a supermartingale (in the extended sense). In particular it follows
\[
 a P (\tau \leq n | B_0) = E_{B_0} [a 1_{n\geq \tau}] \leq E_{B_0} Z_n \leq Z_0,
\]
and
\[
 Z_0 = 1_{0<\tau} X_0 + a 1_{\tau=0} = 1_{X_0 < a} X_0 + 1_{X_0 \geq a} = a \wedge X_0.
\]
Therefore, using the cMCT,
\[
a P \left[ \sup_n X_n \geq a | B_0 \right] = a P (\tau < \infty | B_0) = \lim_{n \to \infty} a P (\tau \leq n | B_0) \leq Z_0 = a \wedge X_0
\]
which proves Eq. (21.49).

For the last assertion, take \( a > 0 \) to be constant in Eq. (21.49) and then use the cDCT to let \( a \uparrow \infty \) to conclude
\[
P \left[ \sup_n X_n = \infty | B_0 \right] = \lim_{a \to \infty} P \left[ \sup_n X_n \geq a | B_0 \right] \leq \lim_{a \to \infty} 1 \wedge X_0 / a = 1_{X_0 = \infty}.
\]
Multiplying this equation by \( 1_{X_0 < \infty} \) and then taking expectations implies
\[
 E \left[ 1_{\sup_n X_n = \infty} 1_{X_0 < \infty} \right] = E \left[ 1_{X_0 = \infty} 1_{X_0 < \infty} \right] = 0
\]
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 \( \{ X_0 < \infty \} \).

\section{21.7.2 The upcrossing inequality and convergence result}

\textbf{Theorem 21.64 (Dubin’s Upcrossing Inequality).} Suppose \( X = (X_n)_{n=0}^{\infty} \) is a positive supermartingale and \( 0 < a < b < \infty \). Then
\[
P \left( U^X_{\infty} (a,b) \geq k | B_0 \right) \leq \left( \frac{a}{b} \right)^k \left( 1 + \frac{X_0}{a} \right), \quad \text{for} \quad k \geq 1 \tag{21.51}
\]
and \( U_\infty (a,b) < \infty \) a.s. and in fact
\[
 E \left[ U^X_{\infty} (a,b) \right] \leq \frac{1}{b/a - 1} = \frac{a}{b-a} < \infty.
\]

\textbf{Proof.} Since
\[
 U^X_N (a,b) = U^{X/a}_N (1,b/a),
\]
it suffices to consider the case where \( a = 1 \) and \( b > 1 \). Let \( \tau_n \) be the stopping times defined in Eq. (21.41) with \( a = 1 \) and \( b > 1 \), i.e.
\[
 \tau_0 = 0, \quad \tau_1 = \inf \{ n \geq \tau_0 : X_n \leq 1 \}
\]
\[
 \tau_2 = \inf \{ n \geq \tau_1 : X_n \geq b \}, \quad \tau_3 := \inf \{ n \geq \tau_2 : X_n \leq 1 \}
\]
\[
 \vdots
\]
\[
 \tau_{2k} = \inf \{ n \geq \tau_{2k-1} : X_n \geq b \}, \quad \tau_{2k+1} := \inf \{ n \geq \tau_{2k} : X_n \leq 1 \},
\]
\[
 \vdots
\]
see Figure 21.2

Let \( k \geq 1 \) and use the switching Lemma 21.62 repeatedly to define a new positive supermartingale \( Y_n = Y_n^{(k)} \) (see Exercise 21.13 below) as follows,

\[
Y_n^{(k)} = 1_{n < \tau_1} + 1_{\tau_1 \leq n < \tau_2} X_n + b_{1_{\tau_2 \leq n < \tau_3}} + b_{X_n, 1_{\tau_3 \leq n < \tau_4}} + b_{1_{\tau_4 \leq n < \tau_5}} + b_{X_n, 1_{\tau_5 \leq n < \tau_6}} + \ldots \]

\[= b^{k-1}1_{\tau_{2k-2} \leq n < \tau_{2k-1}} + b^{k-1}1_{\tau_{2k-1} \leq n < \tau_{2k}} + b^{k+1}1_{\tau_{2k} \leq n}. \quad (21.52)\]

Since \( E[Y_n|B_0] \leq Y_0 \) a.s., \( Y_n \geq b^k1_{\tau_{2k} \leq n} \), and

\[Y_0 = 1_{0 < \tau_1} + 1_{\tau_1 = 0} X_0 = 1_{X_0 > 0} + 1_{X_0 \leq 0} X_0 = 1 \land X_0,\]

we may infer that

\[b^k P(\tau_{2k} \leq n|B_0) = E[b^k1_{\tau_{2k} \leq n}|B_0] \leq E[Y_n|B_0] \leq 1 \land X_0 \] a.s.

Using cMCT, we may now let \( n \to \infty \) to conclude

\[P(U^X(1,b) \geq k|B_0) \leq P(\tau_{2k} < \infty|B_0) \leq \frac{1}{b^k} (1 \land X_0) \] a.s.

which is Eq. (21.51). Using cDCT, we may let \( k \to \infty \) in this equation to discover

\[P(U^X(1,b) = \infty|B_0) = 0 \] a.s. and in particular, \( U^X(1,b) < \infty \) a.s. In fact we have

\[E[U^X(1,b)] = \sum_{k=1}^{\infty} P(U^X(1,b) \geq k) \leq \sum_{k=1}^{\infty} \frac{1}{b^k} (1 \land X_0) \]

\[= \frac{1}{b} \left( \frac{1}{1-1/b} \right) \leq \frac{1}{b-1} < \infty.\]

**Exercise 21.13.** In this exercise you are asked to fill in the details showing \( Y_n \) in Eq. (21.52) is still a supermartingale. To do this, define \( Y_n^{(k)} \) via Eq. (21.52) and then show (making use of the switching Lemma 21.62 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 21.62

\[Y_n^{(1)} = 1_{0 \leq n < \tau_1} U_n + 1_{\tau_1 \leq n < \tau_2} V_n = 1_{0 \leq n < \tau_1}\]

is also a supermartingale.

---

**21.7 *Supermartingale inequalities* 333**

**Corollary 21.65 (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

\[E[X_\infty|B_n] \leq X_n \text{ for all } n \in \mathbb{N}. \quad (21.53)\]

In particular,

\[E X_\infty \leq E X_n \leq E X_0 \text{ for all } n < \infty. \quad (21.54)\]

**Proof.** The set,

\[\{ U^X(1,a,b) < \infty : a,b \in \mathbb{Q} \text{ with } a < b \} \]

has full measure \( (P(\Omega_0) = 1) \) by Dubin’s upcrossing inequality in Theorem 21.64. So by Lemma 21.55 for \( \omega \in \Omega_0 \) we have \( X_\infty(\omega) := \lim_{n \to \infty} X_n(\omega) \) exists a.s. in \( [0,\infty] \). For definiteness, let \( X_\infty = 0 \) on \( \Omega_0 \). Equation (21.53) is now a consequence of cFatou;

\[E[X_\infty|B_n] = \lim_{m \to \infty} E[X_m|B_n] \leq \liminf_{m \to \infty} E[X_m|B_n] \leq \liminf_{m \to \infty} X_m = X_n \text{ a.s.}\]

The supermartingale property guarantees that \( E X_n \leq E X_0 \) for all \( n < \infty \) while taking expectations of Eq. (21.53) implies \( E X_\infty \leq E X_n \).  

---

**Theorem 21.66 (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 21.65), and \( \sigma \) and \( \tau \) are arbitrary stopping times. Then \( X_\tau := \lim_{n \to \infty} X_n \) is a positive \( \{B_n\}_{n=0}^\infty \) – supermartingale, \( X_\infty = \lim_{n \to \infty} X_\tau \land \sigma \), and

\[E[X_\tau|B_n] \leq X_{\sigma \land \tau} \text{ a.s.} \quad (21.55)\]

Moreover, if \( E X_0 < \infty \), then \( E X_\tau = E[X_\infty] < \infty \).  

**Proof.** We already know that \( X_\tau \) is a positive supermartingale by optional stopping Theorem 21.39. Hence an application of Corollary 21.65 implies that \( \lim_{n \to \infty} X_\tau = \lim_{n \to \infty} X_\tau \land \sigma \) is convergent and

\[E \left[ \lim_{n \to \infty} X_\tau|B_m \right] \leq X_m = X_\tau \land m \text{ for all } m < \infty. \quad (21.56)\]

On the set \( \{\tau < \infty\} \), \( \lim_{n \to \infty} X_\tau \land \sigma = X_\tau \) and on the set \( \{\tau = \infty\} \), \( \lim_{n \to \infty} X_\tau \land \sigma = \lim_{n \to \infty} X_n = X_\infty = X_\tau \) a.s. Therefore it follows that \( \lim_{n \to \infty} X_\tau = X_\tau \) and Eq. (21.56) may be expressed as

\[E[X_\tau|B_m] \leq X_\tau \land m \text{ for all } m < \infty. \quad (21.57)\]

An application of Lemma 21.31 now implies

\[E[X_\tau|B_\sigma] = \sum_{m \leq \infty} 1_{\sigma = m} E[X_\tau|B_m] \leq \sum_{m \leq \infty} 1_{\sigma = m} X_\tau \land m = X_\tau \land \sigma \text{ a.s.}\]

9 If \( E X_0 < \infty \), this may also be deduced by applying Corollary 21.56 to \( \{-X_n\}_{n=0}^\infty \).
21.8 Martingale Closure and Regularity Results

We are now going to give a couple of theorems which have already been alluded to in Exercises 16.6, 17.9 and 17.10.

**Theorem 21.67.** 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 21.56. Then the following are equivalent:

1. There exists \( X \in L^1(\Omega, \mathcal{B}, \mathbb{P}) \) 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 \( L^1(\Omega, \mathcal{B}, \mathbb{P}) \).

Moreover, if any of the above equivalent conditions hold we may take \( X = M_{\infty} \), i.e. \( M_n = \mathbb{E}[M_{\infty}|\mathcal{B}_n] \) as \( n \to \infty \), and thus, \( M_n = \mathbb{E}[M_{\infty}|\mathcal{B}_n] \) a.s.

**Definition 21.68.** A martingale satisfying any and all of the equivalent statements in Theorem 21.67 is said to be regular.

**Exercise 21.14 (Rademacher’s theorem).** Let \( \Omega := (0,1] \), \( \mathcal{B} := \mathcal{B}_{(0,1]} \), \( \mathbb{P} = m \) be Lebesgue measure, and \( f \in L^1(\mathbb{P}) \). To each partition \( \Pi := \{0 = x_0 < x_1 < x_2 < \cdots < x_n = 1\} \) of \( (0,1] \) we let \( f_{\Pi} := \sigma(J_i := (x_{i-1}, x_i] : 1 \leq i \leq n) \).

1. Show \( \mathbb{E}[f_{\Pi}|\mathcal{B}_\Pi](x) = \sum_{i=1}^{n} \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} f(s) \, ds \cdot 1_{(x_{i-1}, x_i]}(x) \) for a.e. \( x \in \Omega \).
2. For \( f \in C([0,1], \mathbb{R}) \), let \( f_{\Pi}(x) := \sum_{i=1}^{n} \frac{\Delta_i f}{\Delta_i} 1_{J_i}(x) \)

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

\( f_{\Pi} = \mathbb{E}[f_{\Pi'}|\mathcal{B}_{\Pi'}] \) a.s.

3. Show for any \( a, b \in \Pi \) with \( a < b \) that

\[
\frac{f(b) - f(a)}{b-a} = \frac{1}{b-a} \int_{a}^{b} f_{\Pi}(x) \, dx.
\]

**Hint:** consider the partition \( \Pi_0 := \{0 < a < b < 1\} \).

Now let \( \mathcal{B}_n := \mathcal{B}_{\Pi_n} \), and where \( \Pi_n := \{ k/2^n \}_{k=0}^{2^n} \) observe your have now shown \( g_n := f_{\Pi_n} \) is a martingale.

4. Let us now further suppose that \( |f(y) - f(x)| \leq K |y - x| \) for all \( x, y \in [0,1] \), i.e. \( f \) is Lipschitz. From Eq. (21.58) it follows that \( |g_n| := |f_{\Pi_n}| \leq K \) so that \( \{g_n\}_{n=1}^{\infty} \) is a bounded martingale. Use this along with Eq. (21.59) and Theorem 21.67 to conclude there exists \( g \in L^\infty(\mathbb{P}) \) such that

\[
f(b) - f(a) = \int_{a}^{b} g(x) \, dx \text{ for all } 0 \leq a < b \leq 1.
\]

[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 21.69.** 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 21.56. So, because of Theorem 15.51 to finish the proof it suffices to show \( \{M_n\}_{n=0}^{\infty} \) is uniformly integrable. But by Doob’s inequality, Corollary 21.48 and the MCT, we find

\[
\mathbb{E}\left[ \sup_k |M_k|^p \right] \leq \left( \frac{P}{p-1} \right)^P \sup_k \mathbb{E}[|M_k|^p] < \infty.
\]

As \( |M_n|^p \leq \sup_k |M_k|^p \in L^1(\mathbb{P}) \) for all \( n \in \mathbb{N} \), it follows by Example 15.39 and Exercise 15.5 that \( \{M_n\}_{n=0}^{\infty} \) is uniformly integrable.

**Theorem 21.70 (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 \( M_{\infty} \in L^1(\mathbb{P}) \),

\[
M_{\tau} = \mathbb{E}[M_{\infty}|\mathcal{F}_{\tau}] , \quad \mathbb{E}[M_{\tau}] \leq \mathbb{E}[M_{\infty}] < \infty
\]

and

\[
\mathbb{E}[M_{\tau}|\mathcal{F}_{\sigma}] = M_{\sigma \wedge \tau} \text{ a.s.}
\]

**Proof.** By Theorem 21.67, \( M_{\infty} \in L^1(\Omega, \mathcal{B}, \mathbb{P}) \) and \( M_n := \mathbb{E}_{\mathcal{F}_n} M_{\infty} \) a.s. for all \( n \leq \infty \). By Lemma 21.31.
\[ \mathbb{E}_B M_\infty = \sum_{n \leq \infty} \mathbb{1}_{\tau=n} \mathbb{E}_B M_n = \sum_{n \leq \infty} \mathbb{1}_{\tau=n} M_n = M_\tau. \]

Hence we have \( |M_\tau| = |\mathbb{E}_B M_\infty| \leq \mathbb{E}_B |M_\infty| \) a.s. and \( \mathbb{E} |M_\tau| \leq \mathbb{E} |M_\infty| < \infty. \)
An application of Theorem 21.32 now concludes the proof:
\[ \mathbb{E}_B M_\tau = \sum_{n \leq \infty} \mathbb{1}_{\tau=n} \mathbb{E}_B M_n = \sum_{n \leq \infty} \mathbb{1}_{\tau=n} M_n^\sigma = M_{\sigma \land \tau}. \]

\[
\text{Definition 21.71. Let } M = \{M_n\}_{n=0}^\infty \text{ be a martingale. We say that } \tau \text{ is a regular stopping time for } M \text{ if } M_\tau \text{ is a regular martingale.}
\]

\text{Example 21.72. Every bounded martingale is regular. More generally if } \tau \text{ is a stopping time such that } M^\tau \text{ is bounded, then } \tau \text{ is a regular stopping time for } M.

\text{Remark 21.73. If } \tau \text{ is regular for } M, \text{ then } \lim_{n \to \infty} M^\tau_n := M^\tau_\infty \text{ exists a.s. and in } L^1(P) \text{ and hence}
\[ M^\tau_n = \mathbb{E}[M^\tau_\infty | B_n] \quad \text{for all } n \leq \infty \quad (21.62) \]

and
\[ \lim_{n \to \infty} M^\tau_n = M^\tau_\infty \text{ a.s. on } \{\tau = \infty\}. \quad (21.63) \]

\text{Definition 21.74. If } \tau \text{ is a regular stopping time for } M, \text{ then we define } M_\tau \text{ as,}
\[ M_\tau := M^\tau_\infty = \lim_{n \to \infty} M^\tau_n = \begin{cases} M_\tau & \text{if } \tau < \infty \\ M_\infty & \text{if } \tau = \infty \end{cases}. \quad (21.64) \]

From Eq. (21.62) and the definition of \( M_\tau \) we have
\[ M_{\tau \land n} = M^\tau_n = \mathbb{E}[M^\tau_\infty | B_n] = \mathbb{E}[M_\tau | B_n] \text{ for all } n \leq \infty \quad (21.65) \]

and further note that
\[ \mathbb{E}|M_\tau| = \lim_{n \to \infty} \mathbb{E}|M^\tau_n| \leq \sup_n \mathbb{E}|M^\tau_n| < \infty. \]

\text{Theorem 21.75. Suppose } M = \{M_n\}_{n=0}^\infty \text{ is a martingale and } \sigma, \tau, \text{ are stopping times such that } \tau \text{ is a regular stopping time for } M. \text{ Then}
\[ \mathbb{E}_B M_\tau = M_{\tau \land \sigma}. \quad (21.66) \]

If we further assume that \( \sigma \leq \tau \) a.s. then
\[ M^\sigma_n = \mathbb{E}_B [M_\tau | B_n] \quad (21.67) \]

and \( \sigma \) is also regular for \( M. \)
Moreover for any \( a > 0 \),
\[
\sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \geq a, \ \tau \leq n] = 0.
\]
and the latter term goes to zero as \( a \to \infty \) by assumption 2. Hence we have shown,
\[
\lim_{a \to \infty} \sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \geq a] = 0
\]
as desired.

Now to prove the last assertion. If \( C := \sup_n \mathbb{E}[|M_n|] < \infty \), the (by Corollary 21.56) \( M_\infty := \lim_{n \to \infty} M_n \) a.s. and \( \mathbb{E}|M_\infty| < \infty \). Therefore,
\[
\mathbb{E}[|M_\tau| : \tau < \infty] \leq \mathbb{E}[|M_\tau|] = \mathbb{E}\left[\lim_{n \to \infty} |M_{\tau \wedge n}|\right] \leq \lim_{n \to \infty} \inf \mathbb{E}|M_{\tau \wedge n}| \leq \lim_{n \to \infty} \mathbb{E}|M_n| < \infty
\]
wherein we have used Fatou’s lemma, the optional sampling theorem to conclude \( M_{\tau \wedge n} = \mathbb{E}_{\mathbb{P}_{\tau \wedge n}} M_n \), Jensen to conclude \( |M_{\tau \wedge n}| \leq \mathbb{E}_{\mathbb{P}_{\tau \wedge n}} |M_n| \), and the tower property of conditional expectation to conclude \( \mathbb{E}|M_{\tau \wedge n}| \leq \mathbb{E}|M_n| \).

\begin{corollary}
Suppose that \( M \) is an \( L^1 \) - bounded martingale and \( J \in \mathcal{B}_\mathbb{F} \) is a bounded set, then \( \tau := \inf\{n : M_n \notin J\} \) is a regular stopping time for \( M \).
\end{corollary}

\begin{proof}
According to Proposition 21.76 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 \( J \subseteq [-A,A] \), since \( M_n 1_{n<\tau} \in J \) we have \( |M_n 1_{n<\tau}| \leq A \) which is sufficient to complete the proof.
\end{proof}

\section{Backwards (Reverse) Submartingales}

In this section we will consider submartingales indexed by \( \mathbb{Z}_- := \{\ldots, n, -n+1, \ldots, -2, -1, 0\} \). So again we assume that we have an increasing filtration, \( \{\mathcal{B}_n : n \leq 0\} \), i.e. \( \cdots \subset \mathcal{B}_{-2} \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \subset \mathcal{B} \). As usual, we say an adapted process \( \{X_n\}_{n \leq 0} \) is a submartingale (martingale) provided \( \mathbb{E}[X_n - X_m | \mathcal{B}_n] \geq 0 \) (\( = 0 \)) for all \( m \leq n \). Observe that \( \mathbb{E}X_n \geq \mathbb{E}X_m \) for \( n \geq m \), so that \( \mathbb{E}X_n \) decreases as \( n \) increases. Also observe that \( \{X_{-n}, X_{-n+1}, \ldots, X_{-1}, 0\} \) is a “finite string” submartingale relative to the filtration, \( \mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \cdots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \).

\begin{remark}
For all \( n \in \mathbb{Z}_- \) we have \( X_n \leq \mathbb{E}[X_0 | \mathcal{B}_n] \) and therefore by the same argument in Lemma 21.52 we have \( X_0^+ \leq \mathbb{E}[X_0^+ | \mathcal{B}_0] \). In particular we conclude that \( \{X_0^+\}_{n \in \mathbb{Z}_-} \) is uniformly integrable. In general, backwards submartingales are even better behaved than forward submartingales. To see this even more clearly notice if \( \{M_n\}_{n \leq 0} \) is a backwards martingale, then \( \mathbb{E}[M_n | \mathcal{B}_m] = M_{m \wedge n} \) for all \( m, n \leq 0 \). 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 \( M_0 \in L^1(\Omega, \mathcal{B}, \mathbb{P}) \). We have seen in Example 21.7 that this need not be the case for forward martingales.
\end{remark}

\begin{theorem}
(Backwards (or reverse) submartingale convergence). Let \( \{\mathcal{B}_n : n \leq 0\} \) be a reverse filtration, \( \{X_n\}_{n \leq 0} \) is a backwards submartingale. Then \( X_{\infty} = \lim_{n \to -\infty} X_n \) exists a.s. in \( \{-\infty\} \cup \mathbb{R} \) and \( X_{\infty}^+ \in L^1(\Omega, \mathcal{B}, \mathbb{P}) \). If we further assume that
\end{theorem}
\begin{enumerate}
\item Since \( X_n = X_n^+ - X_n^- \), it follows from Remark 21.78 that
\end{enumerate}
\[ C := \lim_{n \to -\infty} \mathbb{E}X_n = \inf_{n \leq 0} \mathbb{E}X_n > -\infty, \]  

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) \( X_{-\infty} \in L^1(\Omega, B, P) \), and 4) \( \lim_{n \to -\infty} \mathbb{E}|X_n - X_{-\infty}| = 0 \).

**Proof.** The number of downcrossings of \( (X_0, X_{-1}, \ldots, X_{(n-1)}, X_{-n}) \) across \([a, b]\), (denoted by \( D_n(a, b) \)) is equal to the number of upcrossings, \((X_{-1}, X_{-n}, \ldots, X_{-1}, X_0)\) across \([a, b]\), see Figure 21.5. Since \((X_{-n}, X_{(n-1)}, \ldots, X_{-1}, X_0)\) is a \( \mathcal{B}_{-n} \subset \mathcal{B}_{-n-1} \subset \cdots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \) submartingale, we may apply Doob’s upcrossing inequality (Theorem 21.53) to find:

\[
(b - a) \mathbb{E}[D_n(a, b)] \leq \mathbb{E}(X_0 - a) - \mathbb{E}(X_{-n} - a) + \leq \mathbb{E}(X_0 - a) < \infty. \tag{21.69}
\]

Letting \( D_n(a, b) := \lim_{n \to -\infty} D_n(a, b) \) be the total number of downcrossing of \((X_0, X_{-1}, \ldots, X_{-n}, \ldots)\), using the MCT to pass to the limit in Eq. (21.69), we have

\[
(b - a) \mathbb{E}[D_n(a, b)] \leq \mathbb{E}(X_0 - a) < \infty.
\]

In particular it follows that \( D_n(a, b) < \infty \) a.s. for all \( a < b \).

As in the proof of Corollary 21.56 (make use of the obvious downcrossing analogue of Lemma 21.55), it follows that \( X_{-\infty} := \lim_{n \to -\infty} X_n \) exists in \( \mathbb{R} \) 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 21.17 we should define

\[ A_n = \sum_{k \leq n} \mathbb{E}[\Delta_k X|B_{k-1}] \text{ where } \Delta_k X := X_k - X_{k-1}. \]

This is a well defined increasing predictable process since the submartingale property implies \( \mathbb{E}[\Delta_k X|B_{k-1}] \geq 0 \). 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 \( \sup_{n \leq 0} \mathbb{E}|X_n| < \infty \).

\[ E[A_0] = \sum_{k \leq 0} \mathbb{E}[E[\Delta_k X|B_{k-1}]] = \sum_{k \leq 0} \mathbb{E}[\Delta_k X]
\]

\[
= \lim_{N \to \infty} (\mathbb{E}X_0 - \mathbb{E}X_N) = \mathbb{E}X_0 - \inf_{n \leq 0} \mathbb{E}X_n = \mathbb{E}X_0 - C < \infty.
\]

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

\[ \mathbb{E}[\Delta_n M|B_{n-1}] = \mathbb{E}[\Delta_n X - \Delta_n A|B_{n-1}] = \mathbb{E}[\Delta_n X|B_{n-1}] - \Delta_n A = 0 \text{ a.s.} \]

Thus \( M \) is a martingale and therefore, \( M_n = \mathbb{E}[M_0|B_n] \) with \( M_0 = X_0 - A_0 \in L^1(P) \). An application of Proposition 21.8 implies \( \{M_n\}_{n \leq 0} \) is uniformly integrable and hence \( X_n = M_n + A_n \) is uniformly integrable as well. (See Remark 21.80 for an alternate proof of the uniform integrability of \( X \).) Therefore \( X_{-\infty} \in L^1(\Omega, B, P) \) and \( X_n \to X_{-\infty} \) in \( L^1(\Omega, B, P) \) as \( n \to \infty \).

To finish the proof we must show without assuming \( C > -\infty \) that \( X_{-\infty}^+ \in L^1(\Omega, B, P) \) which will then also imply \( P(X_{-\infty} = 0) = 0 \). 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 \( \inf \mathbb{E}X_n^+ \geq 0 \) a.s. \( \infty \), it follows by what we have just proved that \( X_{-\infty}^+ \in L^1(\Omega, B, P) \).

**Remark 21.80 (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;

\[
\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - C = K < \infty, \tag{21.70}
\]

which shows that \( \{X_n\}_{n=1}^\infty \) is \( L^1 \) - bounded. For uniform integrability we will use the following identity;

\[
\mathbb{E}[|X| : |X| \geq \lambda] = \mathbb{E}[X : X \geq \lambda] - \mathbb{E}[X : X \leq -\lambda] = \mathbb{E}[X : X \geq -\lambda] + (\mathbb{E}X - \mathbb{E}[X : X > -\lambda]) = \mathbb{E}[X : X \geq \lambda] + \mathbb{E}[X : X > -\lambda] - \mathbb{E}X.
\]

Taking \( X = X_n \) and \( k \geq n \), we find

\[
\mathbb{E}[|X_n| : |X_n| \geq \lambda] = \mathbb{E}[X_n : X_n \geq \lambda] + \mathbb{E}[X_n : X_n > -\lambda] - \mathbb{E}X_n \\
\leq \mathbb{E}[X_k : X_n \geq \lambda] + \mathbb{E}[X : X_n > -\lambda] - \mathbb{E}X_k + (\mathbb{E}X_k - \mathbb{E}X_n) \\
= \mathbb{E}[X_k : X_n \geq \lambda] + (\mathbb{E}X_k - \mathbb{E}X_n).
\]

Given \( \varepsilon > 0 \) we may choose \( k = k_\varepsilon < 0 \) such that if \( n \leq k_\varepsilon \), \( 0 \leq \mathbb{E}X_k - \mathbb{E}X_n \leq \varepsilon \) and hence
\[
\limsup_{\lambda \to \infty} \sup_{n \leq k} \mathbb{E} \left[ |X_n| : |X_n| \geq \lambda \right] \leq \limsup_{\lambda \to \infty} \mathbb{E} \left[ |X_k| : |X_k| \geq \lambda \right] + \varepsilon \leq \varepsilon
\]

wherein we have used Eq. (21.70). Chebyshev’s inequality to conclude
\[ P \left( |X_k| \geq \lambda \right) \leq K/\lambda \] and then the uniform integrability of the singleton set, \( \{|X_k|\} \subset L^1(\Omega, \mathcal{F}, P) \). From this it now easily follows that \( \{X_n\}_{n \leq 0} \) is a uniformly integrable.

**Corollary 21.81.** Suppose \( 1 \leq p < \infty \) and \( X_n = M_n \) in Theorem 21.79, 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 \( L^p \). Moreover \( M_{-\infty} = \mathbb{E} [M_0 | \mathcal{B}_{-\infty}] \), where \( \mathcal{B}_{-\infty} = \cap_{n \leq 0} \mathcal{B}_n \).

**Proof.** Since \( M_n = \mathbb{E} [M_0 | \mathcal{B}_n] \) for all \( n \), it follows by Jensen that \( |M_n|^p \leq \mathbb{E} [|M_0|^p | \mathcal{B}_n] \) for all \( n \). By Proposition 21.8 \( \{ \mathbb{E} [|M_0|^p | \mathcal{B}_n] \}_{n \leq 0} \) is uniformly integrable and so is \( \{ |M_n|^p \}_{n \leq 0} \). By Theorem 21.79 \( M_n \Rightarrow M_{-\infty} \) a.s. Hence we may now apply Theorem 15.5.1 to see that \( M_n \Rightarrow M_{-\infty} \) in \( L^p \).

**Example 21.82 (Kolmogorov’s SLLN).** In this example we are going to give another proof of the strong law of large numbers in Theorem 19.10 also see Theorem 23.31 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 \( S_0 = 0, S_n := X_1 + \cdots + X_n \) and \( \mathcal{B}_n = \sigma (S_n, S_{n+1}, S_{n+2}, \ldots) \) so that \( S_n \) is \( \mathcal{B}_n \)-measurable for all \( n \geq 0 \).

The first three items below give a solution to Exercise 17.8.

1. For any permutation \( \sigma \) of the set \( \{1, 2, \ldots, n\} \),
\[
(X_1, \ldots, X_n, S_n, S_{n+1}, S_{n+2}, \ldots) \overset{d}{=} (X_{\sigma 1}, \ldots, X_{\sigma n}, S_n, S_{n+1}, S_{n+2}, \ldots)
\]
and in particular
\[
(X_j, S_n, S_{n+1}, S_{n+2}, \ldots) \overset{d}{=} (X_1, S_n, S_{n+1}, S_{n+2}, \ldots) \quad \text{for all } j \leq n.
\]
2. By Exercise 17.7 we may conclude that
\[
\mathbb{E} [X_j | \mathcal{B}_{-n}] = \mathbb{E} [X_1 | \mathcal{B}_{-n}] \quad \text{a.s. for all } j \leq n.
\]
3. Summing Eq. (21.71) over \( j = 1, 2, \ldots, n \) gives,
\[
S_n = \mathbb{E} [S_n | S_n, S_{n+1}, S_{n+2}, \ldots] = n \mathbb{E} [X_1 | S_n, S_{n+1}, S_{n+2}, \ldots]
\]
from which it follows that
\[
M_{-n} := \frac{S_n}{n} := \mathbb{E} [X_1 | S_n, S_{n+1}, S_{n+2}, \ldots] \quad (21.72)
\]
and hence \( \{ M_{-n} = \frac{1}{n} S_n \} \) is a backwards martingale.
4. By Theorem 21.79 we know;
\[
\lim_{n \to -\infty} \frac{S_n}{n} = \lim_{n \to -\infty} M_{-n} =: M_{-\infty} \text{ exists a.s. and in } L^1 \text{.}
\]
5. Since \( M_{-\infty} = \lim_{n \to -\infty} \frac{S_n}{n} \) is a \( \{ \sigma (X_1, \ldots, X_n) \}_{n=1}^\infty \) -tail random variable it follows by Corollary 13.57 (basically by Kolmogorov’s zero one law of Proposition 13.56) 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 \( L^1 \) 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 21.15.** 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 16.5]
21.10 Some More Martingale Exercises

(The next four problems were taken directly from http://math.nyu.edu/~sheff/martingalenote.pdf.)

Exercise 21.16. Suppose Harriet has 7 dollars. Her plan is to make one dollar bets on a 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 21.17. Consider a contract that at time $N$ will be worth either 100 or 0. Let $S_n$ be its price at time $0 \leq n \leq N$. 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 21.18. 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, $T \leq N$.

Exercise 21.19. 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 21.20. 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 $\lambda > 0$,

$$P \left( \max_{1 \leq m \leq n} M_m \geq \lambda \right) \leq \frac{E[M_n^2]}{E[M_n^2] + \lambda^2}.$$ 

Hints: First show that for any $c > 0$ that $\{X_n := (M_n + c)^2\}_{n=0}^\infty$ is an submartingale and then observe,

$$\left\{ \max_{1 \leq m \leq n} M_m \geq \lambda \right\} \subset \left\{ \max_{1 \leq m \leq n} X_m \geq (\lambda + c)^2 \right\}.$$ 

Now use Doob’ Maximal inequality (Proposition 21.43) to estimate the probability of the last set and then choose $c$ so as to optimize the resulting estimate you get for $P(\max_{1 \leq m \leq n} M_m \geq \lambda)$. (Notice that this result applies to $-M_n$ as well so it also holds that; 

$$P \left( \min_{1 \leq m \leq n} M_m \leq -\lambda \right) \leq \frac{E[M_n^2]}{E[M_n^2] + \lambda^2} \text{ for all } \lambda > 0.$$ 

Exercise 21.21. Let $\{Z_n\}_{n=1}^\infty$ be independent random variables, $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$, and $f_n(\lambda) := \mathbb{E}[e^{i\lambda Z_n}]$. Suppose $\mathbb{E}e^{i\lambda Z_n} = \prod_{n=1}^\infty f_n(\lambda)$ converges to a continuous function, $F(\lambda)$, as $N \to \infty$. Show for each $\lambda \in \mathbb{R}$ that

$$P \left( \lim_{n \to \infty} e^{i\lambda S_n} \text{ exists} \right) = 1.$$ 

Hints:

1. Show it is enough to find an $\varepsilon > 0$ such that Eq. (21.73) holds for $|\lambda| \leq \varepsilon$.
2. Choose $\varepsilon > 0$ such that $|F(\lambda) - 1| < 1/2$ for $|\lambda| \leq \varepsilon$. For $|\lambda| \leq \varepsilon$, show $M_n(\lambda) := e^{i\lambda S_n}$ is a bounded complex martingale relative to the filtration, $B_n = \sigma(Z_1, \ldots, Z_n)$.

Lemma 21.83 (Protter [36, See the lemma on p. 22.]). Let $\{x_n\}_{n=1}^\infty \subset \mathbb{R}$ such that $\{e^{itx_n}\}_{n=1}^\infty$ 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 $t \in \mathbb{R}$, $\lim_{n \to \infty} e^{it U x_n}$ exists a.s. Thus if $n_k$ and $m_k$ are any increasing sequences we have

$$\lim_{k \to \infty} e^{it U x_{n_k}} = \lim_{n \to \infty} e^{it U x_n} = \lim_{k \to \infty} e^{it U x_{m_k}} \text{ a.s.}$$

and therefore,

$$e^{it(U x_{n_k} - U x_{m_k})} = e^{it U x_{n_k}} e^{-it U x_{m_k}} \to 1 \text{ a.s. as } k \to \infty.$$ 

Hence by DCT it follows that

$$\mathbb{E} \left[ e^{it(U x_{n_k} - U x_{m_k})} \right] \to 1 \text{ as } k \to \infty$$

and therefore

$$(x_{n_k} - x_{m_k}) : U = U x_{n_k} - U x_{m_k} \to 0$$

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 21.22 (Continuation of Exercise 21.21 – See Doob [8, Chapter VII.5]). Let $(Z_n)_{n=1}^\infty$ be independent random variables. Use Exercise 21.21 an Lemma 21.83 to prove the series, $\sum_{n=1}^\infty Z_n$, converges in $\mathbb{R}$ a.s. iff $\prod_{n=1}^\infty f_n(\lambda)$ 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.

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.
Exercise 21.26. Let \( P \) for all \( \tau \) to conclude that \( P \) is the (fair) simple random walk on \( Z \). For any \( a \in Z \), let
\[
\sigma_a := \inf \{ n : S_n = a \} .
\]

Exercise 21.23. For \( a < 0 < b \) with \( a, b \in Z \), 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 21.13 and the fact that \( |S_n - S_{n-1}| = |Z| = 1 \) for all \( n \).

Exercise 21.24. In this exercise, you are asked to use the central limit Theorem 13.39 to prove again that \( P(\tau = \infty) = 0 \). **Hints:** Use the central limit theorem to show
\[
\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-x^2/2} \, dx \geq f(0) \, P(\tau = \infty)
\]
for all \( f \in C^3(\mathbb{R} \to [0, \infty)) \) with \( M : = \sup_{x \in \mathbb{R}} |f^{(3)}(x)| < \infty \). Use this inequality to conclude that \( P(\tau = \infty) = 0 \).

Exercise 21.25. Show
\[
P(\sigma_b < \sigma_a) = \frac{|a|}{b + |a|}
\]
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 13.61 where it is shown \( b \) is visited infinitely often.) **Hint:** Using properties of martingales and Exercise 21.23 compute \( \lim_{n \to \infty} \mathbb{E}[\sigma_a \wedge \sigma_b] \) in two different ways.

Exercise 21.26. 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 21.24 above to conclude \( N_n := S_n^2 - n \) is a martingale.
2. Now show
\[
0 = \mathbb{E}N_0 = \mathbb{E}N_{\tau \wedge n} = \mathbb{E}S_{\tau \wedge n}^2 - \mathbb{E}[\tau \wedge n] .
\]
3. Now use DCT and MCT along with Exercise 21.25 to compute the limit as \( n \to \infty \) in Eq. (21.76) to find
\[
\mathbb{E}[\sigma_a \wedge \sigma_b] = \mathbb{E}[\tau] = |a|/b .
\]
4. By considering the limit, \( a \to -\infty \) in Eq. (21.77), show \( \mathbb{E}[\sigma_b] = \infty \).

For the next group of exercise we are now going to suppose that \( P(Z_n = 1) = p > 1/2 \) and \( P(Z_n = -1) = q = 1 - p < 1/2 \). As before let \( B_n = \sigma(1, \ldots, Z_n) \), \( S_0 = 0 \) and \( S_n = Z_1 + \cdots + Z_n \) for \( n \in \mathbb{N} \). Let us review the method above and what you did in Exercise 20.15 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,
\[
\varphi(S_n) = \mathbb{E}_{\varphi}(\varphi(S_{n+1}) = \varphi(S_n + 1) p + \varphi(S_n - 1) q,
\]
and this then leads us to try to solve the following difference equation for \( \varphi \);
\[
\varphi(x) = p \varphi(x + 1) + q \varphi(x - 1) \text{ for all } x \in \mathbb{Z} .
\]
(21.78)

Similar to the theory of second order ODE’s this equation has two linearly independent solutions which could be found by solving Eq. (21.78) 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
\[
\lambda = \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p}
\]
\[
= \frac{1 \pm \sqrt{4p^2 - 4p + 1}}{2p} = \frac{1 \pm \sqrt{(2p - 1)^2}}{2p} = \{1, (1 - p)/p\} = \{1, q/p\}.
\]
The most general solution to Eq. (21.78) is then given by
\[
\varphi(x) = A + B (q/p)^x .
\]

Below we will take \( A = 0 \) and \( B = 1 \). As before let \( \sigma_a = \inf \{ n : S_n = a \} \).

Exercise 21.27. Let \( a < 0 < b \) and \( \tau := \sigma_a \wedge \sigma_b \).

1. Apply the method in Exercise 21.23 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 21.14]
2. Now use the method in Exercise 21.25 to show
\[
P(\sigma_a < \sigma_b) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a} .
\]
(21.79)
3. By letting \( a \to -\infty \) in Eq. (21.79), conclude \( P(\sigma_k = \infty) = 0 \).
4. By letting \( b \to \infty \) in Eq. (21.79), conclude \( P(\sigma_n < \infty) = (q/p)^{\lceil a \rceil} \).

Exercise 21.28. 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 21.4 or Exercise 21.3)

Exercise 21.29. Using exercise 21.28 show
\[
E(\sigma_a \land \sigma_b) = \left( \frac{b[1 - (q/p)^a] + a[(q/p)^b - 1]}{(q/p)^b - (q/p)^a} \right)(p - q)^{-1}.
\]
By considering the limit of this equation as \( a \to -\infty \), show
\[
E[\sigma_a] = \frac{b}{p - q}
\]
and by considering the limit as \( b \to \infty \), show \( E[\sigma_a] = \infty \).

21.11 Appendix: Some Alternate Proofs

This section may be safely omitted (for now).

Proof. Alternate proof of Theorem 21.40 Let \( A \in \mathcal{B}_\sigma \). Then
\[
E[X_t - X_\sigma : A] = E \left[ \sum_{k=0}^{N-1} 1_{\{\sigma \leq k < \tau\}} \Delta_k X_k : A \right]
\]
\[
= E \left[ \sum_{k=1}^{N} \Delta_k X : A \cap \{\sigma \leq k < \tau\} \right].
\]
Since \( A \in \mathcal{B}_\sigma \), \( A \cap \{\sigma \leq k\} \in \mathcal{B}_k \) and since \( \{k < \tau\} = \{\tau \leq k\} \in \mathcal{B}_k \), it follows that \( A \cap \{\sigma \leq k < \tau\} \in \mathcal{B}_k \). Hence we know that
\[
E[\Delta_{k+1}X : A \cap \{\sigma \leq k < \tau\}] \leq 0 \text{ respectively.}
\]
and hence that
\[
E[X_t - X_\sigma : A] \leq 0 \text{ respectively.}
\]
Since this true for all \( A \in \mathcal{B}_\sigma \), Eq. (21.23) follows.

Lemma 21.84. Suppose \((\Omega, \mathcal{B}, \{\mathcal{B}_n\}_n=P)\) is a filtered probability space, \(1 \leq p < \infty\), and let \( \mathcal{B}_\infty := \bigvee_{n=1}^\infty \mathcal{B}_n := \sigma(\bigcup_{n=1}^\infty \mathcal{B}_n) \). Then \( \bigcup_{n=1}^\infty L^p(\Omega, \mathcal{B}_n, P) \) 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 \( \mathcal{B}_\infty = L^p(\Omega, \mathcal{B}_\infty, P) \). Further let \( A \) be the algebra of functions consisting of those \( f \in \bigcup_{n=1}^\infty M_n \) such that \( f \) is bounded. As a consequence of the density Theorem 15.27, we know that \( A \) and hence \( \bigcup_{n=1}^\infty M_n \) is dense in \( \mathcal{B}_\infty = L^p(\Omega, \mathcal{B}_\infty, P) \). This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 15.27 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 \( \lim_{n \to \infty} \|f - \varphi_n\|_{L^p(\Omega)} = 0 \). A routine check shows \( \mathbb{H} \) is a subspace of the bounded \( \mathcal{B}_\infty \) measurable \( \mathbb{R} \) valued functions on \( \Omega \), \( 1 \in \mathbb{A} \subset \mathbb{H} \) 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, \( \lim_{n \to \infty} \|f - f_n\|_{L^p(\Omega)} = 0 \). We may now choose \( \varphi_n \in \mathbb{A} \) such that \( \|f - \varphi_n\|_{L^p(\Omega)} \leq \frac{1}{n} \) then
\[
\lim_{n \to \infty} \sup \|f - \varphi_n\|_{L^p(\Omega)} \leq \lim_{n \to \infty} \sup \|f - f_n\|_{L^p(\Omega)} + \lim_{n \to \infty} \sup \|f_n - \varphi_n\|_{L^p(\Omega)} = 0,
\]
which implies \( f \in \mathbb{H} \).

An application of Dynkin’s Multiplicative System Theorem 11.20 now shows \( \mathbb{H} \) contains all bounded \( \sigma(\mathbb{A}) = \mathcal{B}_\infty \) measurable functions on \( \Omega \). Since for any \( f \in L^p(\Omega, \mathcal{B}_\infty, P) \), \( \mathbb{1}_{\{f \leq n\}} \in \mathbb{H} \) there exists \( \varphi_n \in \mathbb{A} \) such that \( \|f - \varphi_n\|_p \leq n^{-1} \). 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 21.85. Suppose \((\Omega, \mathcal{B}, \{\mathcal{B}_n\}_n=P)\) is a filtered probability space, \(1 \leq p < \infty\), and let \( \mathcal{B}_\infty := \bigvee_{n=1}^\infty \mathcal{B}_n := \sigma(\bigcup_{n=1}^\infty \mathcal{B}_n) \). Then for every \( X \in L^p(\Omega, \mathcal{B}_\infty, P) \), \( X_n = E[X|\mathcal{B}_n] \) is a martingale and \( X_n \to X_\infty := E[X|\mathcal{B}_\infty] \) in \( L^p(\Omega, \mathcal{B}_\infty, P) \) as \( n \to \infty \).

Proof. We have already seen in Example 21.6 that \( X_n = E[X|\mathcal{B}_n] \) is always a martingale. Since conditional expectation is a contraction on \( L^p \) it follows that \( E[X_n]^p \leq E[X]^p < \infty \) for all \( n \in \mathbb{N} \cup \{\infty\} \). So to finish the proof we need to show \( X_n \to X_\infty \) in \( L^p(\Omega, \mathcal{B}_\infty, P) \) 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 \( X \in \bigcup_{n=1}^\infty M_n \), then \( X_n = X \) for all sufficiently large \( n \) and for \( n = \infty \). Now suppose that \( X \in M_\infty \) and \( Y \in \bigcup_{n=1}^\infty M_n \). Then

\[\text{It is at this point that the proof would break down if} p = \infty.\]
Using the density Lemma 21.84 we may choose $X \in \mathbb{B}_p$ such that $\mathbb{E}|X| < \infty$.

Suppose $X$ is a submartingale. Then it follows that $\mathbb{E}|X| < \infty$ as desired.

We are now ready to prove the converse of Theorem 21.85.

**Theorem 21.86.** Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=1}^\infty, P)$ is a filtered probability space, $1 \leq p < \infty$, $\mathbb{B}_\infty := \bigvee_{n=1}^\infty \mathbb{B}_n := \sigma(\bigcup_{n=1}^\infty \mathbb{B}_n)$, and $\mathbb{X}_\infty := \mathbb{E}[X|\mathbb{B}_\infty] \subset L^p(\Omega, \mathcal{B}_\infty, P)$ is a martingale. Further assume that $\mathbb{X}_\infty$ is uniformly integrable if $p = 1$. Then there exists $\mathbb{X}_\infty \in L^p(\Omega, \mathcal{B}_\infty, P)$ such that $\mathbb{X}_\infty := \mathbb{E}[X|\mathbb{B}_\infty]$. Moreover by Theorem 21.85 we know that $\mathbb{X}_\infty \rightarrow \mathbb{X}_\infty$ in $L^p(\Omega, \mathcal{B}_\infty, P)$ as $n \rightarrow \infty$.

**Proof.** By Theorems 16.21 and 16.23 exists $\mathbb{X}_\infty \in L^p(\Omega, \mathcal{B}_\infty, P)$ and a subsequence, $Y_k = X_{n_k}$ such that

$$\lim_{k \rightarrow \infty} \mathbb{E}[Y_k|\mathcal{B}_n] = \mathbb{E}[X_n|\mathcal{B}_n] \text{ for all } h \in L^q(\Omega, \mathcal{B}_\infty, P)$$

where $q := p/(p-1)$. Using the martingale property, if $h \in \mathbb{B}_n$, for some $n$, it follows that $\mathbb{E}[Y_k|\mathcal{B}_n] = \mathbb{X}_n|\mathcal{B}_n$ for all large $k$ and therefore that

$$\mathbb{E}[X_n|h] = \mathbb{E}[X|h] \text{ for all } h \in \mathbb{B}_n.$$ 

This implies that $\mathbb{X} = \mathbb{E}[X|\mathcal{B}_\infty]$ as desired.

**Theorem 21.87 (Almost sure convergence).** Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=1}^\infty, P)$ is a filtered probability space, $1 \leq p < \infty$, and let $\mathbb{B}_\infty := \bigvee_{n=1}^\infty \mathbb{B}_n := \sigma(\bigcup_{n=1}^\infty \mathbb{B}_n)$. Then for every $X \in L^1(\Omega, \mathcal{B}, P)$, the martingale, $\mathbb{X}_n = \mathbb{E}[X|\mathcal{B}_n]$, converges almost surely to $\mathbb{X}_\infty := \mathbb{E}[X|\mathcal{B}_\infty]$.

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

$$\limsup_{n \rightarrow \infty} (a_n + b_n) - \liminf_{n \rightarrow \infty} (a_n + b_n) \leq \limsup_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n - \left( \liminf_{n \rightarrow \infty} a_n + \liminf_{n \rightarrow \infty} b_n \right) = \limsup_{n \rightarrow \infty} a_n - \liminf_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n - \liminf_{n \rightarrow \infty} b_n.$$ (21.81)

**Proof.** Since

$$X_n = \mathbb{E}[X|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n] = \mathbb{E}[X_n|\mathcal{B}_n],$$

there is no loss in generality in assuming $X = X_\infty$. If $X \in M_n := L^1(\Omega, \mathcal{B}_\infty, P)$, then $X_n = X_\infty$ a.s. for all $m \geq n$ and hence $X_n \rightarrow X_\infty$ a.s. Therefore the theorem is valid for any $X$ in the dense (by Lemma 21.84) 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 $Y_j \rightarrow X \in L^1(\Omega, \mathcal{B}_\infty, P)$ and let $Y_{jn} := \mathbb{E}[Y_j|\mathcal{B}_n]$ and $X_n := \mathbb{E}[X|\mathcal{B}_n]$. We know that $Y_{jn} \rightarrow Y_{jn}$ a.s. for each $j \in \mathbb{N}$ and our goal is to show $X_n \rightarrow X_\infty$ a.s. By Doob’s inequality in Corollary 21.48 and the $L^1$ - contraction property of conditional expectation we know that

$$P \left( \sup_{n} |X_n| \geq a \right) \leq \frac{1}{a} E[|X|] \leq \frac{1}{a} E[|X|]$$

and so passing to the limit as $N \rightarrow \infty$ we learn that

$$P \left( \sup_{n} |X_n| \geq a \right) \leq \frac{1}{a} E[|X|] \text{ for all } a > 0.$$ (21.82)

Letting $a \uparrow \infty$ then shows $P \left( \sup_{n} |X_n| = \infty \right) = 0$ and hence $\mathbb{X}_n \rightarrow \mathbb{X}_\infty$ a.s. Hence we may use Eq. (21.81) with $a_n = X_n - Y_{jn}$ and $b_n := Y_{jn}$ to find

$$D = \limsup_{n \rightarrow \infty} X_n - \liminf_{n \rightarrow \infty} X_n \leq \limsup_{n \rightarrow \infty} a_n - \liminf_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n - \liminf_{n \rightarrow \infty} b_n \leq 2 \sup_{n} |a_n| \leq 2 \sup_{n} |X_n - Y_{jn}|.$$ 

wherein we have used $\limsup_{n \rightarrow \infty} b_n - \liminf_{n \rightarrow \infty} b_n \geq 0$ a.s. since $Y_{jn} \rightarrow Y_{jn}$ a.s.

We now apply Doob’s inequality one more time, i.e. use Eq. (21.82) with $X_n$ being replaced by $X_n - Y_{jn}$ and $X$ by $X - Y_j$, to conclude,

$$P(D \geq a) \leq P \left( \sup_{n} |X_n - Y_{jn}| \geq \frac{a}{2} \right) \leq \frac{2}{a} E[|X - Y_j|] \rightarrow 0 \text{ as } j \rightarrow \infty.$$
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 \cite{Stroock} Corollary 5.2.7. Let \( \mathbb{H} \) denote those \( X \in L^1(\Omega, \mathcal{B}, P) \) 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 \( \bigcup_{n=1}^\infty M_n \). 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 \( X(k)_n \in \mathbb{H} \) with \( X(k)_n \to X \) in \( L^1(P) \) and let \( X(k)_n := \mathbb{E}[X(k)|\mathcal{B}_n] \). Then by the maximal inequality in Eq. (21.82),

\[
P \left( \sup_{n \geq N} |X_n - X(k)_n| \geq a \right) \leq \frac{1}{a} \mathbb{E} |X - X(k)| \text{ for all } a > 0 \text{ and } k \in \mathbb{N}.
\]

Therefore,

\[
P \left( \sup_{n \geq N} |X - X_n| \geq 3a \right)
\leq P \left( |X - X(k)| \geq a \right) + P \left( \sup_{n \geq N} |X(k)_n - X(k)| \geq a \right)
\leq \frac{2}{a} \mathbb{E} |X - X(k)| + P \left( \sup_{n \geq N} |X(k)_n - X(k)| \geq a \right)
\] and hence

\[
\limsup_{N \to \infty} P \left( \sup_{n \geq N} |X - X_n| \geq 3a \right) \leq \frac{2}{a} \mathbb{E} |X - X(k)| \to 0 \text{ as } k \to \infty.
\]

Thus we have shown

\[
\limsup_{N \to \infty} P \left( \sup_{n \geq N} |X - X_n| \geq 3a \right) = 0 \text{ for all } a > 0.
\]

Since

\[
\left\{ \limsup_{n \to \infty} |X - X_n| \geq 3a \right\} \subset \left\{ \sup_{n \geq N} |X - X_n| \geq 3a \right\} \text{ for all } N,
\]

it follows that

\[
P \left( \limsup_{n \to \infty} |X - X_n| \geq 3a \right) = 0 \text{ for all } a > 0
\]

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.)
Some Martingale Examples and Applications

**Exercise 22.1.** 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 Ruined. Show

$$P(\text{Ruin}) = P(R) = e^{-2(c-\mu)S_0/\sigma^2}.$$ 

**Outline:**

1. Show that $\lambda = -2(c-\mu)/\sigma^2 < 0$ satisfies, $E\left[e^{\lambda \xi_n}\right] = 1$.
2. With this $\lambda$ show

$$Y_n := \exp (\lambda S_n) = e^{\lambda S_0} \prod_{j=1}^n e^{\lambda \xi_j} \quad (22.1)$$

is a non-negative $B_n = \sigma (Z_1, \ldots, Z_n)$ - 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 $E Y_\tau \leq e^{\lambda S_0}$.
4. Finally conclude that

$$P(R) \leq E[Y_\tau : \tau < \infty] \leq E Y_\tau \leq e^{\lambda S_0} = e^{-2(c-\mu)S_0/\sigma^2}.$$ 

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 likelihood that $S_n$ makes a sufficiently “large deviation” from the “typical” value of $n(c-\mu)$ in order for the company to go bankrupt.

**22.1 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 $S_n := Z_1 + \cdots + Z_n$ 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 22.1.** A real valued random variable, $Z$, is said to be exponentially integrable if $M(\theta) := E\left[e^{\theta Z}\right] < \infty$ for all $\theta \in \mathbb{R}$. The function $M$ is called the moment generating function of $Z$ and we further let $\psi(\theta) := \ln M(\theta) = \ln E\left[e^{\theta Z}\right]$ be the log-moment generating function of $Z$.

**Theorem 22.2 (Large Deviation Upper Bound).** Let, for $n \in \mathbb{N}$, $S_n := Z_1 + \cdots + Z_n$ where $\{Z_n\}_{n=1}^\infty$ be i.i.d. exponentially integrable random variables such that $EZ = 0$ where $Z \overset{d}{=} Z_n$. Then for all $\ell > 0$,

$$P(S_n \geq n\ell) \leq e^{-nI(\ell)} \quad (22.2)$$

where $I(\ell)$ is the “Legendre transformation” of the log-moment generating function, $\psi(\theta) = \ln E\left[e^{\theta Z}\right]$, defined by

$$I(\ell) = \sup_{\theta \geq 0} (\theta \ell - \psi(\theta)) = \sup_{\theta \in \mathbb{R}} (\theta \ell - \psi(\theta)) \geq 0. \quad (22.3)$$

In particular,

$$\limsup_{n \to \infty} \ln P(S_n \geq n\ell) \leq -I(\ell) \text{ for all } \ell > 0. \quad (22.4)$$

**Proof.** Let $\{Z_n\}_{n=1}^\infty$ be i.i.d. exponentially integrable random variables such that $EZ = 0$ where $Z \overset{d}{=} Z_n$. Then for $\ell > 0$ we have for any $\theta \geq 0$ that

$$P(S_n \geq n\ell) = P(e^{\theta S_n} \geq e^{\theta n\ell})$$

$$\leq e^{-\theta n\ell E\left[e^{\theta S_n}\right]} = (e^{-\theta \ell E\left[e^{\theta Z}\right]})^n = \left(e^{-\theta \ell - \psi(\theta)}\right)^n$$

$$= \exp (-n (\theta \ell - \psi(\theta))).$$

Minimizing the far right member of this inequality over $\theta \geq 0$ gives the upper bound in Eq. (22.2), where $I(\ell)$ is given as in the first equality in Eq. (22.3).

To prove the second equality in Eq. (22.3), 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,

$$e^{\theta x} \geq \theta x + 1.$$
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. (22.4) easily follows from Eq. (22.2).

**Theorem 22.3 (Large Deviation Lower Bound).** If there is a maximizer, $\theta_0$, for the the function $\theta \rightarrow \theta \ell - \psi(\theta)$, then
\[
\lim_{n \to \infty} \frac{1}{n} \ln P(S_n \geq n\ell) = -I(\ell) = \theta_0 \ell - \psi(\theta_0).
\]
Combining this result with Eq. (22.4) 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 function $\theta \rightarrow \theta \ell - \psi(\theta)$, then
\[
0 = \ell - \psi'(\theta_0) = \ell - \frac{M'(\theta_0)}{M(\theta_0)} = \ell - \frac{\mathbb{E}[Z e^{\theta_0 Z}]}{M(\theta_0)}.
\]
Thus if $W$ is a random variable with law determined by
\[
\mathbb{E}[f(W)] = M(\theta_0)^{-1} \mathbb{E}[f(Z) e^{\theta_0 Z}]
\]
for all non-negative functions $f : \mathbb{R} \to [0, \infty]$ then $\mathbb{E}[W] = \ell$.

Suppose that $\left\{W_n\right\}_{n=1}^\infty$ has been chosen to be a sequence of i.i.d. random variables such that $W_n \overset{d}{=} W$ for all $n$. Then, for all non-negative functions $f : \mathbb{R}^n \to [0, \infty]$ we have
\[
\mathbb{E}[f(W_1, \ldots, W_n)] = M(\theta_0)^{-n} \mathbb{E}\left[f(Z_1, \ldots, Z_n) \prod_{i=1}^n e^{\theta_0 Z_i}\right]
\]
\[
= M(\theta_0)^{-n} \mathbb{E}\left[f(Z_1, \ldots, Z_n) e^{\theta_0 S_n}\right].
\]
This is easily verified by showing the right side of this equation gives the correct expectations when $f$ is a product function. Replacing $f(z_1, \ldots, z_n)$ by $M(\theta_0)^n e^{-\theta_0 (z_1 + \cdots + z_n)} f(z_1, \ldots, z_n)$ in the previous equation then shows
\[
\mathbb{E}[f(Z_1, \ldots, Z_n)] = M(\theta_0)^n \mathbb{E}[f(W_1, \ldots, W_n) e^{-\theta_0 T_n}] \quad (22.6)
\]
where $T_n := W_1 + \cdots + W_n$.

Taking $\delta > 0$ and $f(z_1, \ldots, z_n) = 1_{z_1 + \cdots + z_n \geq n\ell}$ in Eq. (22.6) shows
\[
P(S_n \geq n\ell) = M(\theta_0)^n \mathbb{E}[e^{-\theta_0 T_n} : n\ell \leq T_n]
\]
\[
\geq M(\theta_0)^n e^{-\theta_0 \ell} : n\ell \leq T_n \leq n(\ell + \delta)
\]
\[
\geq M(\theta_0)^n e^{-\theta_0 (\ell + \delta)} P[n\ell \leq T_n \leq n(\ell + \delta)]
\]
\[
= e^{-n\ell(\ell + \delta)} P[n\ell \leq T_n \leq n(\ell + \delta)].
\]
Taking logarithms of this equation, then dividing by $n$, then letting $n \to \infty$ we learn
\[
\lim_{n \to \infty} \frac{1}{n} \ln P(S_n \geq n\ell) \geq -I(\ell) - \theta_0 \ell \delta + \lim_{n \to \infty} \frac{1}{n} \ln P[n\ell \leq T_n \leq n(\ell + \delta)]
\]
\[
= -I(\ell) - \theta_0 \ell \delta + 0 \quad (22.7)
\]
wherein have used the central limit theorem to argue that
\[
P[n\ell \leq T_n \leq n(\ell + \delta)] = P[0 \leq T_n - n\ell \leq n\delta]
\]
\[
= P\left[0 \leq \frac{T_n - n\ell}{\sqrt{n}} \leq \sqrt{n}\delta\right] \to \frac{1}{2} \text{ as } n \to \infty.
\]
Equation (22.5) now follows from Eq. (22.7) as $\delta > 0$ was arbitrary.

**Example 22.4.** Suppose that $Z \overset{d}{=} N(0, \sigma^2) \overset{d}{=} \sigma N$ where $N \overset{d}{=} N(0, 1)$, then
\[
M(\theta) = \mathbb{E}[e^{\theta Z}] = \mathbb{E}[e^{\theta \sigma N}] = \exp\left(\frac{1}{2} (\sigma \theta)^2\right)
\]
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
\[
I(\ell) = \theta_0 \ell - \psi(\theta_0) = \frac{\ell^2}{\sigma^2} - \frac{1}{2} \sigma^2 \left(\frac{\ell}{\sigma^2}\right)^2 = \frac{1}{2} \ell^2.
\]
In this Gaussian case we actually know that $S_n \overset{d}{=} N(0, n\sigma^2)$ and therefore by Mill’s ratio (see Lemma 10.71),
\[
P(S_n \geq n\ell) = P\left(\sqrt{n}\sigma N \geq n\ell\right) = P\left(N \geq \sqrt{n}\ell\sigma\right)
\]
\[
\sim \frac{1}{\sqrt{2\pi n}\sigma} e^{-n\frac{\ell^2}{2\sigma^2}} = \frac{\sigma}{\sqrt{2\pi n}\ell} e^{-nI(\ell)} \text{ as } n \to \infty.
\]
Remark 22.5. The technique used in the proof of Theorem 22.3 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

\[ E[f(W)] = \frac{E[f(Z)\rho(Z)]}{E[\rho(Z)]} \]

where \( \rho \) is some positive function. Under this change of measure the analogue of Eq. (22.6) is

\[ E[f(Z_1, \ldots, Z_n)] = (E[\rho(Z)])^n \cdot E\left[f(W_1, \ldots, W_n) \frac{1}{\rho(W_i)}\right]. \]

However to make this change of variable easy to deal with in the setting at hand we would like to further have

\[ \prod_{j=1}^{n} \frac{1}{\rho(W_j)} = f_n(T_n) = f_n(W_1 + \cdots + W_n) \]

for some function \( f_n \). Equivalently we would like, for some function \( g_n \), that

\[ \prod_{j=1}^{n} \rho(w_j) = g_n(w_1 + \cdots + w_n) \]

for all \( w_i \). Taking logarithms of this equation and differentiating in the \( w_j \) and \( w_k \) variables shows,

\[ \frac{\rho'(w_j)}{\rho(w_j)} = (\ln g_n)'(w_1 + \cdots + w_n) = \frac{\rho'(w_k)}{\rho(w_k)}. \]

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.

## 22.2 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 20.4. 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 urn at time \( n \). Then we have

\[ P((r_{n+1}, g_{n+1}) = (r + c, g) \mid (r_n, g_n) = (r, g)) = \frac{r}{r + g} \quad \text{and} \]

\[ P((r_{n+1}, g_{n+1}) = (r, g + c) \mid (r_n, g_n) = (r, g)) = \frac{g}{r + g}. \]

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

\[ B_n := \sigma((r_k, g_k) : k \leq n) = \sigma(X_k : k \leq n). \]

Indeed,

\[ E[X_{n+1} \mid B_n] = E[X_{n+1} \mid X_n] = \frac{g_n}{r_n + g_n} + \frac{g_n + c}{r_n + g_n + c} = X_n. \]

Since \( X_n \geq 0 \) and \( E X_n = E X_0 < \infty \) for all \( n \) it follows by Corollary 21.56 that \( X_\infty := \lim_{n \to \infty} X_n \) exists a.s. The distribution of \( X_\infty \) is described in the next theorem.

**Theorem 22.6.** Let \( \gamma := g/c \) and \( \rho := r/c \) and \( \mu := \text{Law}_P(X_\infty) \). Then \( \mu \) is the beta distribution on \([0,1]\) with parameters, \( \gamma, \rho, \) i.e.

\[ d\mu(x) = \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} x^{\gamma-1} (1-x)^{\rho-1} \, dx \quad \text{for} \quad x \in [0,1]. \quad (22.8) \]

**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

\[ \frac{g \cdot (g + c) \cdots (g + (n - 1)c) \cdot r \cdot (r + c) \cdots (r + (n - 1)c)}{(r + g) \cdot (r + g + c) \cdots (r + g + (n - 1)c)} \cdot \frac{r}{r + g} \cdot \frac{g + c}{r + g + c} \cdots \frac{g + 2c}{r + g + 2c} \cdot \frac{r}{r + g + 3c} \cdot \frac{r + c}{r + g + 4c}. \]
Before going to the general case let us warm up with the special case, 

On the set, 

\[ \left\{ \begin{align*} &\text{for any } n, \\ &\text{hence we may conclude that} \\ &\text{and hence it follows that for any } f \in C([0,1]), \end{align*} \right. \]

\[
P(\text{Draw } m - \text{greens}) = \left( \frac{n}{m} \right) \frac{1 \cdot 2 \cdot \cdots \cdot m \cdot 1 \cdot 2 \cdot \cdots \cdot (n-m)}{2 \cdot 3 \cdots \cdot (n+1)} = \frac{1}{n+1}. \]

On the set, \{Draw m – greens\}, we have \( X_n = \frac{1+m}{2n+m} \) and hence it follows that 

\[
E[f(X_n)] = \sum_{n=0}^{\infty} f \left( \frac{m+1}{n+2} \right) P(\text{Draw } m - \text{greens})
\]

Therefore 

\[
E[f(X)] = \lim_{n \to \infty} E[f(X_n)] = \int_0^1 f(x) \, dx \tag{22.10}
\]

and hence we may conclude that \( X_\infty \) has the uniform distribution on \([0,1]\).

For the general case, recall from Example 10.50 that \( n! = \Gamma(n+1) \), \( \Gamma(t+1) = t \Gamma(t) \), and therefore for \( m \in \mathbb{N}, \)

\[
\Gamma(x+m) = (x+m-1)(x+m-2)\ldots(x+1)x \Gamma(x). \tag{22.11}
\]

Also recall Stirling’s formula in Eq. 10.61 (also see Theorem 10.72) that

\[
\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} \left[ 1 + O \left( x^{-1} \right) \right] \tag{22.12}
\]

where \( r(x) \to 0 \) as \( x \to \infty \). To finish the proof we will follow the strategy of the proof of Eq. (22.10) using Stirling’s formula to estimate the expression for \( P(\text{Draw } m - \text{greens}) \) in Eq. (22.9).

On the set, \{Draw m – 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. (22.9) as

\[
P(\text{Draw } m - \text{greens}) = \left( \frac{n}{m} \right) \frac{\Gamma(\gamma + m)}{\Gamma(\gamma)} \cdot \frac{\Gamma(\rho + n - m)}{\Gamma(\rho)}
\]

\[
= \frac{\Gamma(\rho + \gamma + n)}{\Gamma(\rho)} \cdot \frac{\Gamma(n+1)}{\Gamma(n-m+1)} \cdot \frac{\Gamma(\gamma + m)}{\Gamma(\rho + n - m)} \tag{22.13}
\]

Now by Stirling’s formula,

\[
\frac{\Gamma(\gamma + m)}{\Gamma(m+1)} = \frac{(\gamma + m)^{\gamma + m - 1/2} e^{-(\gamma + m)} [1 + r (\gamma + m)]}{(1 + m)^{m+1 - 1/2} e^{-(m+1)} [1 + r (1 + m)]}
\]

\[
= (\gamma + m)^{-1} \frac{\gamma + m}{m+1} \cdot e^{-(\gamma + m)} \left[ 1 + r (\gamma + m) \right]
\]

\[
\approx \frac{(\gamma + m)^{-1} \cdot \left( \frac{1 + \gamma/m}{1 + 1/m} \right)^{m+1/2} e^{-(\gamma + m)} \left[ 1 + r (\gamma + m) \right]}{1 + r (m+1)}
\]

We will keep \( m \) fairly large, so that

\[
\left( \frac{1 + \gamma/m}{1 + 1/m} \right)^{m+1/2} = \exp \left( (m+1/2) \ln \left( \frac{1 + \gamma/m}{1 + 1/m} \right) \right)
\]

\[
\approx \exp \left( (m+1/2) (\gamma/m - 1/m) \right) \approx \gamma^{-1}.
\]

Hence we have

\[
\frac{\Gamma(\gamma + m)}{\Gamma(\rho + m - n)} \approx (\gamma + m)^{-1}.
\]

Similarly, keeping \( n - m \) fairly large, we also have

\[
\frac{\Gamma(\rho + n - m)}{\Gamma(n-1)} \approx (\rho + n - m)^{\rho - 1}
\]

and

\[
\frac{\Gamma(\rho + \gamma + n)}{\Gamma(n+1)} \approx (\rho + \gamma + n)^{\rho + \gamma - 1}.
\]

Combining these estimates with Eq. (22.13) gives,
\[ P(\text{Draw } m - \text{greens}) \]
\[ \leq \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} \cdot (\gamma + m)^{\gamma - 1} \cdot (\rho + n - m)^{\rho - 1} \]
\[ = \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} \cdot \left(\frac{\gamma + m}{\rho + \gamma + m}\right)^{\gamma - 1} \cdot \left(\frac{\rho + n - m}{\rho + \gamma + n}\right)^{\rho - 1} \]
\[ = \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} \cdot (x_m)^{\gamma - 1} \cdot (1 - x_m)^{\rho - 1} \Delta_m x. \]

Therefore, for any \( f \in C([0, 1]) \), it follows that
\[ \mathbb{E}[f(X_\infty)] = \lim_{n \to \infty} \mathbb{E}[f(X_n)] \]
\[ = \lim_{n \to \infty} \sum_{m=0}^{n} f(x_m) \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} \cdot (x_m)^{\gamma - 1} \cdot (1 - x_m)^{\rho - 1} \Delta_m x \]
\[ = \int_{0}^{1} f(x) \frac{\Gamma(\rho + \gamma)}{\Gamma(\rho) \Gamma(\gamma)} x^{\gamma - 1} (1 - x)^{\rho - 1} \, dx. \]

\[ \square \]

### 22.3 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 \( n^{th} \) generation and \( \xi_1^{n+1}, \ldots, \xi_{Z_n}^{n+1} \) are the number of off-spring of the \( Z_n \) people of generation \( n \). Then
\[ Z_{n+1} = \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} \]
\[ = \sum_{k=1}^{Z_n} (\xi_1^{n+1} + \cdots + \xi_k^{n+1}) \mathbb{1}_{Z_n=k}. \tag{22.14} \]

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 22.1.

To understand \( Z_n \) a bit better observe that

The sample path in Figure 22.1 corresponds to
\[ \xi_1 = 1, \]
\[ Z_1 = \xi_1 Z_0 = \xi_1^1, \]
\[ Z_2 = \xi_2^1 + \cdots + \xi_3^1, \]
\[ Z_3 = \xi_3^3 + \cdots + \xi_2^2, \]
\[ \vdots \]

The sample path in Figure 22.1 corresponds to
\[ \xi_1 = 3, \]
\[ \xi_2 = 2, \xi_2 = 0, \xi_3 = 3, \]
\[ \xi_3 = \xi_1^3 + \xi_3^1 + \xi_3^3, \]
\[ \xi_4 = \xi_4 = \xi_4^4 = 0. \]

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 22.4.

Let \( \xi \overset{\text{d}}{=} \xi_r^n, p_k := P(\xi = k) \) be the off-spring distribution,
\[ \mu := \mathbb{E}\xi = \sum_{k=0}^{\infty} k \xi_k, \]
which we assume to be finite.

Let \( B_0 = \{\emptyset, \Omega\} \) and
\[ B_n := \sigma \{\xi_i^m : i \geq 1 \text{ and } 1 \leq m \leq n\}. \]
Notation 22.7. Given a bounded function \( f : S = \mathbb{N}_0 \rightarrow \mathbb{C} \), let \( Qf : S \rightarrow \mathbb{C} \) be defined by bounded or non-negative let
\[
Qf(0) := f(0) \quad \text{and} \quad Qf(k) := \mathbb{E}[f(Y_1 + \cdots + Y_k)] \quad \text{for all } k \geq 1.
\]
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
\[
\varphi(\lambda) := \mathbb{E}[\lambda^{Y_1}] = \sum_{k \geq 0} p_k \lambda^k
\]
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
\[
Qf(k) = \mathbb{E}[Y_1 + \cdots + Y_k] = k \cdot \mu \implies Qf = \mu f. \tag{22.16}
\]
If \( f(k) = \lambda^k \) for some \( |\lambda| \leq 1 \), then
\[
\left(Q\lambda^{(i)}\right)(k) = \mathbb{E}[\lambda^{Y_1+\cdots+Y_k}] = \varphi(\lambda)^k. \tag{22.17}
\]

Remark 22.8. Notice that
\[
Qf(k) = \sum_{l \in S} f(l) p_l^k
\]
where
\[
p_l^k := P(Y_1 + \cdots + Y_k = l) = \sum_{t_1 + \cdots + t_k = l} p_{t_1} \cdots p_{t_k}
\]
with the convention that \( p_0^0 = \delta_{0,n} \). For example

Exercise 22.2. Show that \( \{Z_n\}_{n=0}^{\infty} \) is a time homogeneous Markov process with one step transition kernel being \( Q \), i.e. show
\[
\mathbb{E}[f(Z_{n+1})|\mathcal{B}_n] = (Qf)(Z_n) \tag{22.18}
\]
for all bounded or non-negative functions \( f : S = \mathbb{N}_0 \rightarrow \mathbb{C} \). In particular verify that
\[
P(Z_n = j|Z_{n-1} = k) = p_j^k \quad \text{for all } j, k \in S \text{ and } n \geq 1
\]
and
\[
\mathbb{E}[\lambda^{Z_n}|\mathcal{B}_{n-1}] = \varphi(\lambda)^{Z_{n-1}} \quad \text{a.s.} \tag{22.19}
\]
for all \( \lambda \in \mathbb{C} \) with \( |\lambda| \leq 1 \).

Corollary 22.9. Continuing the notation used above, \( M_n := Z_n/\mu^n \) is a positive martingale and in particular
\[
\mathbb{E}Z_n = \mu^n < \infty \quad \text{for all } n \in \mathbb{N}_0. \tag{22.20}
\]

Proof. If \( f(n) = n \) for all \( n \), then \( Qf = \mu f \) by Eq. \([22.16]\) 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. \([22.20]\).

Theorem 22.10. 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 22.11. If \( \mu = 1 \) and \( P(\xi_i^n = 1) < 1 \)\(^1\) 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, \( \lim_{n \rightarrow \infty} Z_n =: Z_\infty \) 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
\[
\{Z_\infty = k\} = \{Z_n = k \text{ a.a. } n\} = \bigcup_{n=1}^{\infty} \{Z_n = k \text{ for all } n \geq N\},
\]
we have
\[
P(Z_\infty = k) = \lim_{N \rightarrow \infty} P(Z_n = k \text{ for all } n \geq N).
\]
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
\[
P(Z_n = k \text{ for all } n \geq N - 1) \leq P(\xi_1^n + \cdots + \xi_k^n = k \text{ for all } n \geq N) = |P(\xi_1^n + \cdots + \xi_k^n = k)| = 0,
\]
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 \( \mu = 1 \).
\[ P(\xi^n_0 + \cdots + \xi^n_k = 0) = \prod_{i=1}^{k} P(\xi^n_i = 0) = p^n_0 > 0. \]

which then implies \( P(\xi^n_0 + \cdots + \xi^n_k = k) < 1 \). Therefore we have shown \( P(Z_\infty = k) = 0 \) 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 \( \{Z_n\}_{n=0}^\infty \).]

\[ \text{Remark 22.12. By the way, the branching process,} \{Z_n\}_{n=0}^\infty \text{ with} \ \mu = 1 \text{ and} \ P(\xi = 1) < 1 \text{ gives a nice example of a non regular martingale. Indeed, if} \ Z \text{ were regular, we would have} \]

\[ Z_n = \mathbb{E}\left[ \lim_{m \to \infty} Z_m | \mathcal{B}_n \right] = \mathbb{E}[0|\mathcal{B}_n] = 0 \]

which is clearly false.

We now wish to consider the case where \( \mu := \mathbb{E}[\xi^m] > 1 \). Let \( \varphi(\lambda) \) be as in Eq. 22.15 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

\[ \varphi'(s) = \sum_{k \geq 0} kp_k s^{k-1} \text{ and } \varphi''(s) = \sum_{k \geq 0} k(k-1)p_k s^{k-2} \geq 0 \]

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 \( \varphi'(1) = \mu \).

\[ \text{Lemma 22.13. If} \ \mu = \varphi'(1) > 1 \text{, there exists a unique} \ \rho < 1 \text{ so that} \ \varphi(\rho) = \rho. \]

\[ \text{Proof. See Figure 22.2 below.} \]

\[ \text{Theorem 22.14 (See Durrett [12], p. 247-248.). If} \ \mu > 1 \text{, then} \]

\[ P(\text{Extinction}) = P\left( \left\{ \lim_{n \to \infty} Z_n = 0 \right\} \right) = P(\{Z_n = 0 \text{ for some} n\}) = \rho. \]

\[ \text{Exercise 22.3. In the notation used in this section (Section 22.3), show for all} \ n \in \mathbb{N} \text{ and} \ \lambda_i \in \mathbb{C} \text{ with} \ |\lambda_i| \leq 1 \text{ that} \]

\[ \mathbb{E}\left[ \prod_{j=1}^{n} \lambda_{i_j}^{Z_i} \right] = \varphi(\lambda_1 \varphi(\cdots \lambda_{n-2} \varphi(\lambda_{n-1} \varphi(\lambda_n)))) \].

\[ ^2\text{This argument is made precise with the aid of Exercise 22.4.} \]
For example, you should show,

\[ E \left[ \lambda_1 Z_1 \lambda_2 Z_2 \lambda_3 Z_3 \right] = \varphi(\lambda_1 \varphi(\lambda_2 \varphi(\lambda_3))) \]

and

\[ E \left[ \lambda_1 Z_1 \lambda_2 Z_2 \lambda_3 Z_3 \lambda_4 Z_4 \right] = \varphi(\lambda_1 \varphi(\lambda_2 \varphi(\lambda_3 \varphi(\lambda_4)))) . \]

**Exercise 22.4.** 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

\[
E \left[ f (Z_2, \ldots, Z_n) \mid Z_1 = k \right] = E \left[ f \left( \sum_{l=1}^k (Z_{l1}, \ldots, Z_{ln-1}) \right) \right] \tag{22.21}
\]

where \( \{ Z_{l1} \}_{n=0}^\infty \) for \( 1 \leq l \leq k \) are i.i.d. Galton-Watson Branching processes such that \( \{ Z_{l1} \}_{n=0}^\infty \overset{d}{=} \{ Z_n \}_{n=0}^\infty \) for each \( l \).

**Suggestion:** it suffices to prove Eq. (22.21) for \( f \) of the form,

\[
f (k_2, \ldots, k_n) = \prod_{j=2}^n \lambda_j^{k_j} . \tag{22.22}
\]

### 22.4 Kakutani’s Theorem

For broad generalizations of the results in this section, see [21 Chapter IV.] or [22].

**Proposition 22.15.** 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 \( d\nu_a = pd\mu \) 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 \( gd\mu \leq dv \), (i.e. \( \int_B gd\mu \leq \nu(B) \) for all \( B \in \mathcal{M} \)), then \( g \leq \rho, \mu - \text{a.e.} \)

**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 10.24 \( g \leq \rho. \) \( \blacksquare \)

**Example 22.16.** This example generalizes Example 21.9. Suppose \((\Omega, \mathcal{B}, \{B_n\}_{n=0}^\infty, P)\) is a filtered probability space and \( Q \) is any another probability measure on \((\Omega, \mathcal{B})\). By the Raydon-Nikodym Theorem 18.8 for each \( n \in \mathbb{N} \) we may write

\[
dQ|_{B_n} = X_n dP|_{B_n} + dR_n \tag{22.23}
\]

where \( R_n \) is a measure on \((\Omega, \mathcal{B}_n)\) which is singular relative to \( P|_{B_n} \) 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 \( X : = \{X_n\}_{n=1}^\infty \) is a positive supermartingale. To verify this assertion, for \( B \in \mathcal{B}_n \) and \( n \leq m \leq \infty \), we have

\[
Q(B) = E [X_m : B] + R_m (B) \geq E [X_m : B] = E [E_{B_n} (X_m) : B]
\]

from which it follows that \( E_{B_n} (X_m) : dP|_{B_n} \leq dQ|_{B_n} \). So according to Proposition 22.15

\[
E_{B_n} (X_m) \leq X_n (P - \text{a.s.}) \text{ for all } n \leq m \leq \infty. \tag{22.24}
\]
Proposition 22.17. Keeping the assumptions and notation used in Example 22.16, then \( \lim_{n \to \infty} X_n = X_\infty \) a.s. and in particular the Lebesgue decomposition of \( Q|_{B_\infty} \) relative to \( P|_{B_\infty} \) may be written as

\[
dQ|_{B_\infty} = \left( \lim_{n \to \infty} X_n \right) \cdot dP|_{B_\infty} + dR_\infty.
\]

Proof. By Example 22.16, we know that \( \{X_n\}_{n \leq \infty} \) is a positive supermartingale and by letting \( m = \infty \) in Eq. (22.24), we know

\[
E_{B_n} X_\infty \leq X_n \text{ a.s.}
\]

By the supermartingale convergence Corollary 21.56 or by the submartingale convergence Corollary 21.56 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. (22.23) holds for all \( n = \infty \).

From the regular martingale convergence Theorem 21.67 we also know that \( \lim_{n \to \infty} E_{B_n} X_\infty = X_\infty \) a.s. as well. So passing to the limit in Eq. (22.26) implies \( X_\infty \leq Y \) a.s. To prove the reverse inequality, \( Y \leq X_\infty \) a.s., let \( B \in B_m \) and \( n \geq m \). Then

\[
Q(B) = E[X_n : B] + R_n(B) \geq E[X_n : B]
\]

and so by Fatou’s lemma,

\[
E[Y : B] = E\left[\liminf_{n \to \infty} X_n : B\right] \leq \liminf_{n \to \infty} E[X_n : B] \leq Q(B).
\]

Since \( m \in \mathbb{N} \) was arbitrary, we have proved \( E[Y : B] \leq Q(B) \) for all \( B \) in the algebra, \( A := \cup_{m \in \mathbb{N}} B_m \). As a consequence of the regularity Theorem 7.7 or of the monotone class Lemma 7.6, or of Theorem 6.20, it follows that \( E[Y : B] \leq Q(B) \) for all \( B \in \sigma(A) = B_\infty \). An application of Proposition 22.15 then implies \( Y \leq X_\infty \) a.s. \( \blacksquare \)

Theorem 22.18. \((\Omega, \mathcal{B}, \{B_n\}_{n=0}^{\infty}, P)\) be a filtered probability space and \( Q \) be a probability measure on \((\Omega, \mathcal{B})\) such that \( Q|_{B_n} \ll P|_{B_n} \) for all \( n \in \mathbb{N} \). Let \( M_n := dQ|_{B_n} \) be a version of the Raydon-Nikodym derivative of \( Q|_{B_n} \) relative to \( P|_{B_n} \) see Theorem 18.8. Recall from Example 21.9 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. \( Q|_{B_\infty} \ll P|_{B_\infty} \).
2. \( E_P M_\infty = 1 \).
3. \( M_n \to M_\infty \) in \( L^1(\Omega, \mathcal{B}, P) \).
4. \( \{M_n\}_{n=1}^{\infty} \) is uniformly integrable.
5. \( \sum_{n=1}^{\infty} (1 - E(\sqrt{X_n})) < \infty \).

Moreover, if any one, and hence all of the above statements, fails to hold, then \( P(M_\infty = 0) = 1 \).
Proof. If \( a_n := \mathbb{E}\left(\sqrt{X_n}\right) \), then \( 0 < a_n \) and \( a_n^2 \leq \mathbb{E}X_n = 1 \) with equality iff \( X_n = 1 \) a.s. So Remark \ref{rem:22.19} 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 \ref{thm:22.18} 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 parts of the argument, let \( N_0 = 1 \) and \( N_n \) be the martingale (again see Example \ref{ex:21.11}) defined by

\[
N_n := \prod_{k=1}^{n} \frac{\sqrt{X_k}}{a_k} = \frac{\sqrt{M_n}}{\prod_{k=1}^{n} a_k}. \tag{22.29}
\]

Further observe that, in all cases, \( N_\infty = \lim_{n \to \infty} N_n \) exists in \( [0, \infty) \) \( \mu \) a.s., see Corollary \ref{cor:21.50} or Corollary \ref{cor:21.65} 4. \( \implies \) 3. Since

\[
E\left[N_n^2\right] = \frac{EM_n}{\left(\prod_{k=1}^{n} a_k\right)^2} = \frac{1}{\left(\prod_{k=1}^{n} a_k\right)^2} \leq \frac{1}{\left(\lim_{k \to \infty} a_k\right)^2} < \infty,
\]

and hence \( \{N_n\}_{n=1}^\infty \) is bounded in \( L^2 \). Therefore, using

\[
M_n = \left(\prod_{k=1}^{n} a_k\right)^2 N_n^2 \leq N_n^2 \tag{22.30}
\]

and Doob’s inequality in Corollary \ref{cor:21.48} we find

\[
E\left[\sup_n M_n\right] = E\left[\sup_n N_n^2\right] \leq 4 \sup_n E\left[N_n^2\right] < \infty. \tag{22.31}
\]

Equation Eq. \ref{eq:22.31} certainly implies \( \{M_n\}_{n=1}^\infty \) is uniformly integrable, see Proposition \ref{prop:15.48}.

Not(4.) \( \implies \) \( P(M_\infty = 0) = 1 \). If

\[
\prod_{n=1}^{\infty} E\left(\sqrt{X_n}\right) = \lim_{n \to \infty} \prod_{k=1}^{n} a_k = 0,
\]

we may pass to the limit in Eq. \ref{eq:22.30} to find

\[
M_\infty = \lim_{n \to \infty} M_n = \lim_{n \to \infty} \left(\prod_{k=1}^{n} a_k\right)^2 N_n^2 = 0 \cdot \left(\lim_{n \to \infty} N_n\right)^2 = 0 \text{ a.s.}
\]

\[\blacksquare\]

Lemma 22.21. Given two probability measures, \( \mu \) and \( \nu \) on a measurable space, \( (\Omega, B) \), there exists \( a \) 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 \( (\Omega, B) \) such that \( \mu \ll \lambda \) and \( \nu \ll \lambda \). We will write \( \sqrt{\frac{d\mu}{d\lambda}} \cdot \frac{d\nu}{d\lambda} \) 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

\[
d\lambda_1 = \frac{d\lambda_1}{d\lambda} d\lambda,
\]

\[
d\mu = \frac{d\mu}{d\lambda_1} d\lambda_1 = \frac{d\mu}{d\lambda_1} d\lambda_1 d\lambda,
\]

\[
d\nu = \frac{d\nu}{d\lambda_1} d\lambda_1 = \frac{d\nu}{d\lambda_1} d\lambda_1 d\lambda.
\]

So

\[
\sqrt{\frac{d\mu}{d\lambda} \cdot \frac{d\nu}{d\lambda}} d\lambda = \sqrt{\frac{d\mu}{d\lambda_1} \cdot \frac{d\nu}{d\lambda_1}} d\lambda_1 d\lambda = \sqrt{\frac{d\mu}{d\lambda_1} \cdot \frac{d\nu}{d\lambda_1}} d\lambda = \sqrt{\frac{d\mu}{d\lambda_1} \cdot \frac{d\nu}{d\lambda_1}} d\lambda_1
\]

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

\[
\sqrt{\frac{d\mu}{d\lambda_1} \cdot \frac{d\nu}{d\lambda_2}} d\lambda_2 = \sqrt{\frac{d\mu}{d\lambda_1} \cdot \frac{d\nu}{d\lambda_1}} d\lambda_1
\]

and hence \( d\rho = \sqrt{\frac{d\mu}{d\lambda}} \cdot \frac{d\nu}{d\lambda} \) is well defined.

\[\blacksquare\]

Definition 22.22. Two probability measures, \( \mu \) and \( \nu \) on a measurable space, \( (\Omega, B) \) 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

\[
H(\mu, \nu) := \int_{\Omega} \sqrt{\frac{d\mu}{d\lambda}} \cdot \frac{d\nu}{d\lambda} = \int_{\Omega} \frac{d\mu}{d\lambda} \cdot \frac{d\nu}{d\lambda} d\lambda \tag{22.32}
\]

where \( \lambda \) is any measure (for example \( \lambda = \frac{1}{2} (\mu + \nu) \) would work) on \( (\Omega, B) \) such that there exists \( \frac{d\mu}{d\lambda} \) and \( \frac{d\nu}{d\lambda} \) in \( L^1 (\Omega, B, \lambda) \) such that \( d\mu = \frac{d\mu}{d\lambda} d\lambda \) and \( d\nu = \frac{d\nu}{d\lambda} d\lambda \). Lemma \ref{lem:22.21} guarantees that \( H(\mu, \nu) \) is well defined.
Proposition 22.23. The Hellinger integral, $H(\mu, \nu)$, of two probability measures, $\mu$ and $\nu$, is well defined. Moreover $H(\mu, \nu)$ satisfies:

1. $0 \leq H(\mu, \nu) \leq 1$,
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 \footnote{This statement and its proof may be safely omitted.}

$$H(\mu, \nu) = \inf \left\{ \sqrt{\frac{1}{n} \sum_{i=1}^{n} \sqrt{\mu(A_i) \nu(A_i)} : \Omega = \sum_{i=1}^{n} A_i \text{ and } n \in \mathbb{N}} \right\}. \quad (22.23)$$

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 $\frac{d\mu}{d\lambda} \cdot \frac{d\nu}{d\lambda} = 0$, $\lambda$-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 $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\lambda} - \lambda$-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, $\nu \ll \mu$, then

$$H(\mu, \nu) = \int_{\Omega} \sqrt{\frac{d\nu}{d\mu} \frac{d\mu}{d\nu} d\mu} = \int_{\Omega} \sqrt{\frac{d\nu}{d\mu}} d\mu.$$

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 \footnote{This statement and its proof may be safely omitted.} \ref{exercise:22.6}.

Exercise 22.5. Find a counter example to the statement that $H(\mu, \nu) > 0$ implies $\nu \ll \mu$.

Exercise 22.6. Prove Eq. \eqref{22.23}.

Corollary 22.24 (Kakutani \cite{25}). Let $\Omega = \mathbb{R}^n$, $Y_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, and $\mathcal{B} := \mathcal{B} = \sigma(\{Y_n : n \in \mathbb{N}\})$ 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 $(\Omega, \mathcal{B})$ 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 \ref{3.68} (take $\mu := P_\sigma(Y_1, Y_2, \ldots)^{-1}$). Let us further assume that $\nu_n \ll \mu_n$ for all $n$ so that

$$0 < H(\mu_n, \nu_n) = \int_{\mathbb{R}} \frac{d\nu_n}{d\mu_n} d\mu_n \leq 1.$$

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

$$\frac{d\nu}{d\mu} = \prod_{n=1}^{\infty} \frac{d\nu_n}{d\mu_n}(Y_n) \quad \mu\text{-a.s.} \quad (22.34)$$

and in all cases we have

$$H(\mu, \nu) = \prod_{n=1}^{\infty} H(\mu_n, \nu_n).$$

Proof. Let $P = \mu$, $Q = \nu$, $B_n := \sigma(Y_1, \ldots, Y_n)$, $X_n := \frac{d\nu_n}{d\mu_n}(Y_n)$, and

$$M_n := X_1 \cdots X_n = \frac{d\nu_1}{d\mu_1}(Y_1) \cdots \frac{d\nu_n}{d\mu_n}(Y_n).$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded measurable function, then

$$\mathbb{E}_\nu(f(Y_1, \ldots, Y_n)) = \int_{\mathbb{R}^n} f(y_1, \ldots, y_n) d\nu_1(y_1) \cdots d\nu_n(y_n)$$

$$= \int_{\mathbb{R}^n} f(y_1, \ldots, y_n) \frac{d\nu_1}{d\mu_1}(y_1) \cdots \frac{d\nu_n}{d\mu_n}(y_n) \ d\mu_1(y_1) \cdots d\mu_n(y_n)$$

$$= \mathbb{E}_\mu \left[ f(Y_1, \ldots, Y_n) \frac{d\nu_1}{d\mu_1}(Y_1) \cdots \frac{d\nu_n}{d\mu_n}(Y_n) \right]$$

$$= \mathbb{E}_\mu(f(Y_1, \ldots, Y_n) M_n)$$

from which it follows that

$$d\nu|_{\mathcal{B}_n} = M_n d\mu|_{\mathcal{B}_n}.$$ 

Hence by Theorem \ref{22.18} $M := \lim_{n \to \infty} M_n$ exists a.s. and the Lebesgue decomposition of $\nu$ is given by

$$d\nu = M d\mu + dR_{\infty}.$$
where $R_\infty \perp \mu$. Moreover $\nu \ll \mu$ iff $R_\infty = 0$ which happens iff $EM_\infty = 1$ and $\nu \perp \mu$ iff $R_\infty = \nu$ which happens iff $M_\infty = 0$. From Theorem 22.20,

$$E_\mu M_\infty = 1 \text{ iff } 0 < \sum_{n=1}^{\infty} E_\mu \left( \sqrt{X_n} \right) = \sum_{n=1}^{\infty} \int_\mathbb{R} \sqrt{d\nu} \cdot d\mu_n = \sum_{n=1}^{\infty} H(\mu_n, \nu_n)$$

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

$$\prod_{n=1}^{\infty} E_\mu \left( \sqrt{X_n} \right) = \prod_{n=1}^{\infty} H(\mu_n, \nu_n) = 0,$$

Theorem 22.20 implies $M_\infty = 0$, $\mu \text{ a.s.}$ in which case Theorem 22.18 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 $\nu \perp \nu$ 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

$$E_\mu \left[ \sqrt{M_n} - \sqrt{M_\infty} \right]^2 \leq E_\mu \left[ \sqrt{M_n} - \sqrt{M_\infty} \right] \cdot \sqrt{M_n + M_\infty}$$

$$= E_\mu \left[ |M_n - M_\infty| \right] \to 0 \text{ as } n \to \infty.$$

Since $d\nu = M_\infty d\mu$ in this case, it follows that

$$H(\mu, \nu) = E_\mu \left[ \sqrt{M_\infty} \right] = \lim_{n \to \infty} E_\mu \left[ \sqrt{M_n} \right] = \lim_{n \to \infty} \prod_{k=1}^{\infty} H(\mu_k, \nu_k) = \prod_{k=1}^{\infty} H(\mu_k, \nu_k).$$

Example 22.25. Suppose that $\nu_n = \delta_1$ for all $n$ and $\mu_n = (1 - p_n^2) \delta_0 + p_n^2 \delta_1$ with $p_n \in (0,1)$. Then $\nu_n \ll \mu_n$ with

$$\frac{d\nu_n}{d\mu_n} = \delta_{1}(p_n^{-2}) \quad \text{and} \quad H(\mu_n, \nu_n) = \int_\mathbb{R} \sqrt{1_{\{1\}} p_n^{-2}} d\mu_n = \sqrt{p_n^{-2} \cdot p_n} = p_n.$$

So in this case $\nu \ll \mu$ if $\sum_{n=1}^{\infty} (1 - p_n) < \infty$. Observe that $\mu$ is never absolutely continuous relative to $\nu$.

On the other hand; if we further assume in Corollary 22.24 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 want to make use of the following Gaussian integral formula;

$$\int_\mathbb{R} \exp \left( -\frac{a}{2} x^2 + bx \right) dx = \int_\mathbb{R} \exp \left( -\frac{a}{2} (x - \frac{b}{a})^2 + \frac{b^2}{2a} \right) dx$$

$$= e^{\frac{b^2}{2a}} \int_\mathbb{R} \exp \left( -\frac{a}{2} x^2 \right) dx = \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a}}$$

which is valid for all $a > 0$ and $b \in \mathbb{R}$.

Exercise 22.7 (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 $\sum_{n=1}^{\infty} a_n^2 < \infty$.

Exercise 22.8. Suppose $s, t > 0$, $\{a_n\} \subset \mathbb{R}$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi s}} e^{-x^2/2s} dx$ and $d\nu_n(x) = \frac{1}{\sqrt{2\pi t}} e^{-(x+a_n)^2/2t} dx$. Show $\mu \perp \nu$ if $s \neq t$.

Exercise 22.9. Suppose $\{t_n\} \subset (0, \infty)$, $d\mu_n(x) = \frac{1}{\sqrt{2\pi t_n}} e^{-x^2/2t_n} dx$ 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 $\mu \sim \nu$. 

Page: 356  job: prob  macro: svmonob.cls  date/time: 2-Nov-2018/7:31
(Weak) Convergence of Random Sums
Random Sums

As usual let \((\Omega, \mathcal{F}, P)\) be a probability space. The general theme of this chapter is to consider arrays of random variables, \(\{X^n_k\}_{k=1}^{\infty}\), for each \(n \in \mathbb{N}\). We are going to look for conditions under which \(\lim_{n \to \infty} \sum_{k=1}^{\infty} X^n_k\) exists almost surely or in \(L^p\) for some \(0 \leq p < \infty\). Typically we will start with a sequence of random variables, \(\{X_k\}_{k=1}^{\infty}\) and consider the convergence of

\[
S_n = \frac{X_1 + \cdots + X_n}{b_n} - a_n
\]

for appropriate choices of sequence of numbers, \(\{a_n\}\) and \(\{b_n\}\). This fits into our general scheme by taking \(X^n_k = X_k^p/b_n - a_n/n\).

23.1 Weak Laws of Large Numbers

**Theorem 23.1 (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 \(\sigma^2_n = \text{Var}(X_n)\). If there exists an increasing positive sequence, \(\{a_n\}\) and \(\mu \in \mathbb{R}\) such that

\[
\lim_{n \to \infty} \frac{1}{a_n} \sum_{j=1}^{n} \mu_j = \mu \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{a_n^2} \sum_{j=1}^{n} \sigma^2_j = 0,
\]

then \(\frac{S_n}{a_n} \to \mu\) in \(L^2(P)\) (and hence also in probability).

**Exercise 23.1.** Prove Theorem 23.1

**Example 23.2.** Suppose that \(\{X_k\}_{k=1}^{\infty} \subset L^2(P)\) are uncorrelated identically distributed random variables. Then

\[
\frac{S_n}{n} \overset{L^2(P)}{\to} \mu = \mathbb{E}X_1 \quad \text{as} \quad n \to \infty.
\]

To see this, simply apply Theorem 23.1 with \(a_n = n\). More generally if \(b_n \uparrow \infty\) such that \(\lim_{n \to \infty} (n/b_n^2) = 0\), then

\[
\text{Var}\left(\frac{S_n}{b_n}\right) = \frac{1}{b_n^2} \cdot n \text{Var}(X_1) \to 0 \quad \text{as} \quad n \to \infty
\]

and therefore

\[
(S_n - n\mu)/b_n \to 0 \quad \text{in} \quad L^2(P).
\]

**Note well:** since \(L^2(P)\) convergence implies \(L^p(P)\) - convergence for \(0 \leq p \leq 2\), where by \(L^p(P)\) - convergence we mean convergence in probability. The remainder of this chapter is mostly devoted to proving a.s. convergence for the quantities in Theorem 15.25 and Proposition 23.10 under various assumptions. These results will be described in the next section.

**Theorem 23.3 (Weak Law of Large Numbers).** Suppose that \(\{X_n\}_{n=1}^{\infty}\) is a sequence of independent random variables. Let and

\[
S_n := \sum_{j=1}^{n} X_j \quad \text{and} \quad a_n := \sum_{k=1}^{n} \mathbb{E}(X_k : |X_k| \leq n).
\]

If

\[
\lim_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E}(X_k^2 : |X_k| \leq n) = 0,
\]

then

\[
\frac{S_n - a_n}{n} \overset{p}{\to} 0.
\]

**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

\[
S'_n := \sum_{k=1}^{n} X_k 1_{|X_k| \leq n}.
\]

Since \(\{S_n \neq S'_n\} \subset \bigcup_{k=1}^{\infty} \{|X_k| > n\}\),

\[
P\left(\left|\frac{S_n - a_n}{n} - \frac{S'_n - a_n}{n}\right| > \varepsilon\right) = P\left(\left|\frac{S_n - S'_n}{n}\right| > \varepsilon\right) \leq P(S_n \neq S'_n) \leq \sum_{k=1}^{n} P(|X_k| > n) \to 0 \quad \text{as} \quad n \to \infty.
\]
Hence it suffices to show \( \frac{S'_n - a_n}{n} \xrightarrow{L^2(P)} 0 \) as \( n \to \infty \) and for this it suffices to show, \( \frac{S'_n - a_n}{n} \xrightarrow{L^2} 0 \) as \( n \to \infty \).

Observe that \( \mathbb{E} S'_n = a_n \) and therefore,
\[
\mathbb{E} \left( \frac{S'_n - a_n}{n} \right)^2 = \frac{1}{n^2} \text{Var} \left( S'_n \right) = \frac{1}{n^2} \sum_{k=1}^{n} \text{Var} \left( X_1 1_{|X_1| \leq n} \right)
\]
\[
\leq \frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E} \left( X_k^2 1_{|X_k| \leq n} \right) \xrightarrow{n \to \infty} 0,
\]
in which we have used \( \text{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \) 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 23.4.** If \( X \in L^1(P) \), Chebyshev’s inequality along with the dominated convergence theorem implies
\[
\tau(x) := xP(|X| \geq x) \leq \mathbb{E} [|X| : |X| \geq x] \xrightarrow{x \to \infty} 0.
\]

If \( X \) is a random variable such that \( \tau(x) = xP(|X| \geq x) \xrightarrow{x \to \infty} 0 \), we say that \( X \) is in “weak \( L^1 \).”

**Exercise 23.2.** 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 \( X(y) := (y \ln y)^{-1} \cdot 1_{y \leq 1/2} \) for \( y \in \Omega \). Show that \( X \notin L^1(P) \) yet \( \lim_{x \to \infty} xP(|X| \geq x) = 0 \).

**Lemma 23.5.** Let \( X \) be a random variable such that \( \tau(x) := xP(|X| \geq x) \xrightarrow{x \to \infty} 0 \), then
\[
\lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left( |X|^2 : |X| \leq n \right) = 0.
\]

**Proof.** To prove this we observe that
\[
\mathbb{E} \left( |X|^2 : |X| \leq n \right) = \mathbb{E} \left[ 2 \int_{0 \leq x \leq |X| \leq n} x \, dx \right] = 2 \int_0^n xP(0 \leq x \leq |X| \leq n) \, dx
\]
\[
\leq 2 \int_0^n xP(|X| \geq x) \, dx = 2 \int_0^n \tau(x) \, dx
\]
so that
\[
\frac{1}{n} \mathbb{E} \left( |X|^2 : |X| \leq n \right) = \frac{2}{n} \int_0^n \tau(x) \, dx.
\]
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 23.6 (Feller’s WLLN).** If \( \{X_n\}_{n=1}^\infty \) are i.i.d. and \( \tau(x) := xP(|X| > x) \xrightarrow{x \to \infty} 0 \) as \( x \to \infty \), then the hypothesis of Theorem 23.3 are satisfied so that
\[
\frac{S_n}{n} - \mathbb{E} (X_1 : |X_1| \leq n) \xrightarrow{P} 0.
\]

**Proof.** Since
\[
\sum_{k=1}^{n} P(|X_k| > n) = nP(|X| > n) = \tau(n) \xrightarrow{n \to \infty} 0,
\]
Eq. 23.1 is satisfied. Equation 23.2 follows from Lemma 23.5 and the identity,
\[
\frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E} (X_k^2) = \frac{1}{n} \mathbb{E} \left( |X_1|^2 : |X_1| \leq n \right).
\]

As a direct corollary of Feller’s WLLN and Remark 23.4 we get Khintchin’s weak law of large numbers.

**Corollary 23.7 (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 \( \{\frac{1}{n} S_n\}_{n=1}^\infty \) is uniformly integrable under these hypothesis.

This result is also clearly a consequence of Komogorov’s strong law of large numbers.

**23.1.1 A WLLN Example**

**Theorem 23.8 (Shannon’s Theorem).** Let \( \{X_k\}_{k=1}^r \) be a sequence of i.i.d. random variables with values in \( \{1, 2, \ldots, r\} \subset \mathbb{N} \), \( p(k) := P(X_k = k) > 0 \) for \( 1 \leq k \leq r \), and
\[
H(p) := -\mathbb{E} [\ln p(X_1)] = -\sum_{k=1}^{r} p(k) \ln p(k)
\]
be the entropy of \( p \). If we define \( \pi_n(\omega) := p(X_1(\omega)) \ldots p(X_n(\omega)) \) to be the “probability of the realization” \( \{X_1(\omega), \ldots, X_n(\omega)\} \), then for all \( \varepsilon > 0 \),
\[
P \left( e^{-nH(p)+\varepsilon} \leq \pi_n \leq e^{-nH(p)-\varepsilon} \right) \xrightarrow{n \to \infty} 1
\]

Thus the probability, \( \pi_n \), that the random sample \( \{X_1, \ldots, X_n\} \) 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
\[
-\frac{1}{n} \ln \pi_n = -\frac{1}{n} \sum_{i=1}^{n} \ln p(X_i) \stackrel{P}{\rightarrow} -\mathbb{E} [\ln p(X_1)] = -\sum_{k=1}^{r} p(k) \ln p(k) =: H(p),
\]
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
\[
\begin{align*}
\left\{ \left| H(p) + \frac{1}{n} \ln \pi_n \right| > \varepsilon \right\} &= \left\{ H(p) + \frac{1}{n} \ln \pi_n > \varepsilon \right\} \cup \left\{ H(p) + \frac{1}{n} \ln \pi_n < -\varepsilon \right\} \\
&= \left\{ \frac{1}{n} \ln \pi_n > -H(p) + \varepsilon \right\} \cup \left\{ \frac{1}{n} \ln \pi_n < -H(p) - \varepsilon \right\} \\
&= \left\{ \pi_n > e^{n(-H(p)+\varepsilon)} \right\} \cup \left\{ \pi_n < e^{n(-H(p)-\varepsilon)} \right\}
\end{align*}
\]
it follows that
\[
\left\{ \left| H(p) + \frac{1}{n} \ln \pi_n \right| > \varepsilon \right\} \subset \left\{ \pi_n \leq e^{n(-H(p)-\varepsilon)} \right\} \cap \left\{ \pi_n \geq e^{n(-H(p)+\varepsilon)} \right\}
\]
and therefore
\[
P \left( e^{-n(H(p)+\varepsilon)} \leq \pi_n \leq e^{-n(H(p)-\varepsilon)} \right) \to 1 \text{ as } n \to \infty.
\]

For our next example, let \( \{X_n\}_{n=1}^{\infty} \) be i.i.d. random variables with common distribution function, \( F(x) := P(X_n \leq x) \). For \( x \in \mathbb{R} \) let \( F_n(x) \) be the empirical distribution function defined by,
\[
F_n(x) := \frac{1}{n} \sum_{j=1}^{n} 1_{X_j \leq x} = \left( \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j} \right) ((-\infty, x]].
\]
Since \( \mathbb{E} 1_{X_j \leq x} = F(x) \) and \( \{1_{X_j \leq x}\}_{j=1}^{\infty} \) are Bernoulli random variables, the weak law of large numbers implies \( F_n(x) \stackrel{P}{\rightarrow} F(x) \) as \( n \to \infty \). As usual, for \( p \in (0,1) \) let
\[
F^+(p) := \inf \{ x : F(x) \geq p \}
\]
and recall that \( F^+(p) \leq x \) iff \( F(x) \geq p \). Let us notice that
\[
F_n^+(p) = \inf \{ x : F_n(x) \geq p \} = \inf \left\{ x : \frac{1}{n} \sum_{j=1}^{n} 1_{X_j \leq x} \geq np \right\} = \inf \{ x : \# \{ j \leq n : X_j \leq x \} \geq np \}.
\]
Recall from Definition [14.11] that the order statistic of \( (X_1, \ldots, X_n) \) is the finite sequence, \( (X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}) \), where \( (X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}) \) denotes \( (X_1, \ldots, X_n) \) arranged in increasing order with possible repetitions. It follows from the formula in Definition [14.11] that \( X_k^{(n)} \) are all random variables for \( k \leq n \), but it will be useful to give another proof. Indeed, \( X_k^{(n)} \leq x \) iff \( \# \{ j \leq n : X_j \leq x \} \geq k \) if \( \sum_{j=1}^{n} 1_{X_j \leq x} \geq k \), i.e.
\[
\left\{ X_k^{(n)} \leq x \right\} = \left\{ \sum_{j=1}^{n} 1_{X_j \leq x} \geq k \right\} \in \mathcal{B}.
\]
Moreover, if we let \([x] = \min \{ n \in \mathbb{Z} : n \geq x \}\), the reader may easily check that \( F_n^+(p) = X_{[np]}^{(n)} \).

Proposition 23.9. Keeping the notation above. Suppose that \( p \in (0,1) \) is a point where
\[
F(F^+(p) - \varepsilon) < p < F(F^+(p) + \varepsilon) \text{ for all } \varepsilon > 0 \]
then \( X_{[np]}^{(n)} = F_n^+(p) \) \( \stackrel{P}{\rightarrow} \) \( F^+(p) \text{ as } n \to \infty \). 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
\[
\{ F_n^+(p) - F^+(p) > \varepsilon \} \subset \{ F_n^+(p) - \varepsilon + F^+(p) \} = \{ F_n^+(p) \leq \varepsilon + F^+(p) \} = \{ F_n^+(p) \geq F^+(p) \} \geq p \}
\]
so that
\[
\{ F_n^+(p) - F^+(p) > \varepsilon \} = \{ F_n^+(p) \leq \varepsilon + F^+(p) \} = \{ F_n(\varepsilon + F^+(p)) \}
\]
and
\[
\{ F_n^+(p) - F^+(p) > \varepsilon \} = \{ F_n(\varepsilon + F^+(p)) \leq p - F(F^+(p) + \varepsilon) \}
\]
Letting \( \delta_\varepsilon := F(F^+(p) + \varepsilon) - p > 0 \), we have, as \( n \to \infty \), that
\[
P \{ \{ F_n^+(p) - F^+(p) > \varepsilon \} \} = P \{ F_n(\varepsilon + F^+(p)) \leq p - F(F^+(p) + \varepsilon) \}
\]
Similarly, let \( \delta_\varepsilon := p - F(F^+(p) - \varepsilon) > 0 \) and observe that
\[
\{ F_n^+(p) - F_n^+(p) > \varepsilon \} = \{ F_n^+(p) \geq F^+(p) - \varepsilon \} = \{ F_n(F^+(p) - \varepsilon) \}
\]
and
\[
\{ F_n^+(p) - F_n^+(p) > \varepsilon \} = \{ F_n^+(p) \leq F^+(p) - \varepsilon \} = \{ F_n(F^+(p) - \varepsilon) \geq p \}
23.2 Kolmogorov’s Convergence Criteria

Proposition 23.10 (\(L^2\) - Convergence of Random Sums). Suppose that \(\{Y_k\}_{k=1}^{\infty} \subset L^2(P)\) are uncorrelated. If \(\sum_{k=1}^{\infty} \text{Var}(Y_k) < \infty\) 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 15.25) to show \(\|S_n - S_m\|_2 \to 0\) as \(m, n \to \infty\). Supposing \(n > m\), we have
\[
\|S_n - S_m\|^2 = \mathbb{E}\left( \sum_{k=m+1}^{n} (Y_k - \mu_k) \right)^2 = \sum_{k=m+1}^{n} \text{Var}(Y_k) = \sum_{k=m+1}^{n} \sigma_k^2 \to 0 \text{ as } m, n \to \infty.
\]

Theorem 23.11 (Kolmogorov’s Convergence Criteria). Suppose that \(\{Y_n\}_{n=1}^{\infty}\) are independent square integrable random variables. If \(\sum_{j=1}^{\infty} \text{Var}(Y_j) < \infty\), then \(\sum_{j=1}^{\infty} (Y_j - \mathbb{E}Y_j)\) converges a.s. In particular if \(\sum_{j=1}^{\infty} \text{Var}(Y_j) < \infty\) 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 21.69. 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, \(\mathcal{F}_n = \sigma(S_0, \ldots, S_n)\). By assumption we have
\[
\mathbb{E}S_n^2 = \sum_{j=1}^{n} \text{Var}(Y_j) \leq \sum_{j=1}^{\infty} \text{Var}(Y_j) < \infty
\]

so that \(\{S_n\}_{n=0}^{\infty}\) is bounded in \(L^2(P)\). Therefore by Theorem 21.69
\[
\sum_{j=1}^{\infty} (Y_j - \mathbb{E}Y_j) = \lim_{n \to \infty} S_n \text{ exists a.s. and in } L^2(P).
\]

Another way to prove this is to appeal Proposition 23.10 above and Lévy’s Theorem 23.50 below. As second method is to make use of Kolmogorov’s inequality and we will give this proof below.

Exercise 23.3 (Resnik 7.1). Does \(\sum_{n=1}^{\infty} 1/n\) converge? Does \(\sum_{n=1}^{\infty} (-1)^n/n\) converge? Let \(\{X_n\}\) be iid with \(P[X_n = \pm 1] = 1/2\) Does \(\sum_{n=1}^{\infty} X_n/n\) converge? [See Example 23.41 below for a more thorough investigation of this sort.]

Example 23.12 (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
\[
\sum_{n=1}^{\infty} a_n N_n \sin \omega_n t \text{ converges a.s.}
\]

provided \(\sum_{n=1}^{\infty} a_n^2 < \infty\). This is a simple consequence of Kolmogorov’s convergence criteria, Theorem 23.11 and the facts that \(\mathbb{E}[a_n N_n \sin \omega_n t] = 0\) and
\[
\text{Var}(a_n N_n \sin \omega_n t) = a_n^2 \sin^2 \omega_n t \leq a_n^2.
\]

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
\[
B_t := \frac{2\sqrt{2}}{\pi} \sum_{k=1,3,5,\ldots} \frac{N_k}{k} \sin \left( k \frac{\pi}{2} t \right) \quad (23.5)
\]
is a.s. convergent for all \(t \in \mathbb{R}\). The factor \(\frac{2\sqrt{2}}{\pi k}\) has been determined by requiring,
\[
\int_0^1 \left[ \frac{d}{dt} \frac{2\sqrt{2}}{\pi k} \sin \left( k \frac{\pi}{2} t \right) \right]^2 dt = 1
\]
as seen by,
\[
\int_0^1 \left[ \frac{d}{dt} \frac{2\sqrt{2}}{\pi k} \sin \left( k \frac{\pi}{2} t \right) \right]^2 dt = \frac{k^2 \pi^2}{2^2} \int_0^1 \left[ \cos \left( k \frac{\pi}{2} t \right) \right]^2 dt = \frac{k^2 \pi^2}{2^2} \frac{2}{k} \frac{1}{4} \left[ \cos \left( k \frac{\pi}{2} t \right) \right]^2_0 = \frac{k^2 \pi^2}{2^3}.
\]

Fact: Wiener in 1923 showed the series in Eq. (23.5) is in fact almost surely uniformly convergent. Given this, the process, \(t \to B_t\) is almost surely continuous. The process \(\{B_t : 0 \leq t \leq 1\}\) is Brownian Motion.
Kolmogorov’s convergence criteria becomes a powerful tool when combined with the following real variable lemma.

**Lemma 23.13 (Kronecker’s Lemma).** Suppose that $\{x_k\}_{k=1}^{\infty} \subset \mathbb{R}$ 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

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} b_k y_k = \sum_{k=1}^{n} b_k (S_k - S_{k-1}) = \sum_{k=1}^{n} b_k S_k - \sum_{k=0}^{n-1} b_{k+1} S_k$$

$$= b_n S_n + \sum_{k=1}^{n-1} (b_k - b_{k+1}) S_k$$

and hence

$$\frac{1}{b_n} \sum_{k=1}^{n} x_k = S_n - \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) S_k$$

$$= S_n - \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) s + \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) (s - S_k)$$

$$= S_n - \frac{1}{b_n} (b_n - b_1) s + R_n,$$

where

$$R_n := \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) (s - S_k).$$

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 $\lim_{n \to \infty} R_n = 0$. 

To this end, for $N \in \mathbb{N}$ let $\varepsilon_N = \sup_{n \geq N} |S_n - s| \to 0$ so that $\varepsilon_N \downarrow 0$ as $N \uparrow \infty$. Then for $n, N \in \mathbb{N}$ with $n > N$ we have (using $b_{k+1} \geq b_k$ for all $k$) that

$$|R_n| \leq \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) |s - S_k|$$

$$\leq \frac{1}{b_n} \sum_{k=1}^{N} (b_{k+1} - b_k) \varepsilon_1 + \varepsilon_N \frac{1}{b_n} \sum_{k=N+1}^{n-1} (b_{k+1} - b_k)$$

$$= \frac{1}{b_n} (b_{N+1} - b_1) \varepsilon_1 + \varepsilon_N \frac{1}{b_n} (b_n - b_{N+1}).$$

Letting $n \to \infty$ in this last inequality then shows

$$\limsup_{n \to \infty} |R_n| \leq 0 + \varepsilon_N \to 0 \text{ as } N \to \infty.$$ 

**Remark 23.14.** Here is another proof of Lemma 23.13 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 $f_n(u) := \sum_{1 \leq k \leq n} y_k$. Then $|f_n(u)| \leq M := \sup_{N \in \mathbb{N}} |S_N| < \infty$ and $f_n(u) \to s$ as $n \to \infty$ for all $u > 0$. Making use of the fundamental theorem of calculus we learn,

$$S_n - \frac{1}{b_n} \sum_{k=1}^{n} b_k y_k = \sum_{k=1}^{n} \left( 1 - \frac{b_k}{b_n} \right) y_k$$

$$= \sum_{k=1}^{n} \left( 1 - \frac{k}{n} \right)^c y_k$$

$$= c \sum_{1 \leq k \leq n} \int_{k/n}^{1} u^{c-1} du y_k$$

$$= c \int_{0}^{1} du u^{c-1} \sum_{1 \leq k \leq n} \left( 1 - \frac{k}{n} \right)^c y_k$$

$$= c \int_{0}^{1} du u^{c-1} f_n(u) \to c \int_{0}^{1} du u^{c-1} s = s \text{ as } n \to \infty.$$ 

Since $S_n \to s$ as $n \to \infty$, the result is proved.

As an immediate corollary we have the following corollary.

**Corollary 23.15 (L^2 – SLLN).** 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 $\sum_{n=1}^{\infty} \frac{1}{b_n} < \infty$, then
\[ \frac{1}{b_n} (S_n - n\mu) \to 0 \text{ a.s. and in } L^2(P) \quad (23.6) \]

We may rewrite Eq. (23.6) as

\[ S_n = n\mu + o(1) b_n \text{ or } \frac{S_n}{n} = \mu + o(1) \frac{b_n}{n}. \]

Example 23.16. For example, we could take \( b_n = n \) or \( b_n = n^p \) for any \( p > 1/2 \), or \( b_n = n^{1/2} \ln(n)^{1/2+\varepsilon} \) for any \( \varepsilon > 0 \). The idea here is that

\[ \sum_{n=2}^{\infty} \frac{1}{(n^{1/2} \ln(n)^{1/2+\varepsilon})^2} = \sum_{n=2}^{\infty} \frac{1}{n (\ln(n))^{1+2\varepsilon}} \]

which may be analyzed by comparison with the integral

\[ \int_{x=2}^{\infty} \frac{1}{x \ln^{1+2\varepsilon} x} \, dx = \int_{\ln 2}^{\infty} \frac{1}{ey^{1+2\varepsilon}} e^y \, dy = \int_{\ln 2}^{\infty} \frac{1}{y^{1+2\varepsilon}} \, dy < \infty, \]

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 \( S_n/n \) about the mean, \( \mu \), have order smaller than \( n^{-1/2} \ln(n)^{1/2+\varepsilon} \).

Fact 23.17 (Missing Reference) Under the hypothesis in Corollary 23.15

\[ \lim_{n \to \infty} \frac{S_n - n\mu}{n^{1/2} \ln(n)^{1/2}} = \sqrt{2} \sigma \text{ a.s.} \]

We end this section with another example of using Kolmogorov’s convergence criteria in conjunction with Kronecker’s Lemma 23.13

Lemma 23.18. 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

\[ \sum_{n=1}^{\infty} \text{Var} \left( \frac{X_n}{\mathbb{E} S_n} \right) = \sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{(\mathbb{E} S_n)^2} < \infty \implies \frac{S_n}{\mathbb{E} S_n} \to 1 \text{ a.s.} \]

Proof. Kolmogorov’s convergence criteria, Theorem 23.11 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 23.13 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 23.19. Suppose that \( \{X_n\}_{n=1}^{\infty} \) are i.i.d. square integrable random variables with \( \mu := \mathbb{E} X_n > 0 \) and \( \sigma^2 := \text{Var}(X_n) < \infty \). Since \( \mathbb{E} S_n = n \mu \uparrow \infty \) and

\[ \sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{(\mathbb{E} S_n)^2} = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2 \mu^2 n^2} < \infty, \]

we may conclude that \( \lim_{n \to \infty} \frac{S_n}{\mathbb{E} S_n} = 1 \text{ a.s., i.e. } S_n/n \to \mu \text{ a.s. as we already know.} \)

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.

\[ A_j := \{X_j > \max\{X_1, X_2, \ldots, X_{k-1}\}\}. \]

Recall from Renyi Theorem 13.26 that \( \{A_j\}_{j=1}^{\infty} \) are independent and \( P(A_j) = \frac{1}{j} \) for all \( j \).

Proposition 23.20. Keeping the preceding notation and let \( S_n := \sum_{j=1}^{j} 1_{A_j} \), denote the number of records in the first \( n \) observations. Then \( \lim_{n \to \infty} \frac{S_n}{\ln n} = 1 \text{ a.s.} \)

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

\[ \text{Var}(1_{A_n}) = \mathbb{E}^2 1_{A_n} - (\mathbb{E} 1_{A_n})^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \]

so by that

\[ \sum_{n=1}^{\infty} \text{Var} \left( \frac{1_{A_n}}{\mathbb{E} S_n} \right) = \sum_{n=1}^{\infty} \left( \frac{1}{n} - \frac{1}{n^2} \right) \left( \sum_{j=1}^{1} \frac{1}{j^2} \right) \]

\[ \leq \sum_{n=1}^{\infty} \frac{1}{n} \left( \sum_{j=1}^{1} \frac{1}{j} \right)^2 n \]

\[ \leq 1 + \int_{2}^{\infty} \frac{1}{y^2} \, dy = 1 + \int_{2}^{\infty} \frac{1}{y^2} \, dy < \infty. \]
Therefore by Lemma 23.18 we may conclude that \( \lim_{n \to \infty} \frac{S_n}{E S_n} = 1 \) a.s.

So to finish the proof it only remains to show

\[
\lim_{n \to \infty} \frac{E S_n}{\ln n} = \lim_{n \to \infty} \frac{\sum_{j=1}^{n-1} \frac{1}{j} x^j}{\ln n} = 1. \tag{23.7}
\]

To see this write

\[
\ln(n+1) = \int_1^{n+1} \frac{1}{x} dx = \sum_{j=1}^{n} \int_{j}^{j+1} \frac{1}{x} dx = \sum_{j=1}^{n} \left( \int_{j}^{j+1} \left( \frac{1}{x} - 1 \right) dx \right) + \sum_{j=1}^{n} \frac{1}{j} = \rho_n + \sum_{j=1}^{n} \frac{1}{j} \tag{23.8}
\]

where

\[
|\rho_n| = \sum_{j=1}^{n} \left| \ln \frac{j+1}{j} - 1 \right| = \sum_{j=1}^{n} \left| \ln(1 + 1/j) - 1 \right| \sim \sum_{j=1}^{n} \frac{1}{j^2}
\]

and hence we conclude that \( \lim_{n \to \infty} \rho_n < \infty \). So dividing Eq. (23.8) by \( \ln n \) and letting \( n \to \infty \) gives the desired limit in Eq. (23.7).

\section{23.3 The Strong Law of Large Numbers Revisited}

\textbf{Remark 23.21.} 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. \( \lim_{n \to \infty} \frac{S_n}{\tilde{E} S_n} \) exists a.s. as an \( \mathbb{R} \)-valued random variable
2. there exists \( c \in \mathbb{R} \) such that \( \lim_{n \to \infty} \frac{X_n}{\tilde{E} X_n} = c \) a.s.
3. \( \lim_{n \to \infty} \frac{X_n}{\tilde{E} X_n} = 0 \) a.s.
4. \( E |X| < \infty \).

Indeed, 1. \( \implies \) 2. by Kolmogorov’s zero one law, see Corollary \ref{kolmogorov_zero_one}. 2. \( \implies \) 3. \( \implies \) 4. is contained in Proposition \ref{feller2} and Corollary \ref{feller3}. 4. \( \implies \) 1. is Theorem \ref{kl}

\textbf{Definition 23.22.} Two sequences, \( \{X_n\} \) and \( \{X'_n\} \), of random variables are tail equivalent if

\[
E \left[ \sum_{n=1}^{\infty} 1_{X_n \neq X'_n} \right] = \sum_{n=1}^{\infty} P(X_n \neq X'_n) < \infty.
\]

\textbf{Proposition 23.23.} 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

\[
\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=1}^{n} X_k = X \text{ a.s}
\]

then

\[
\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=1}^{n} X'_k = X \text{ a.s}
\]

\textbf{Proof.} If \( \{X_n\} \) and \( \{X'_n\} \) are tail equivalent, we know by the first Borel-Cantelli Lemma \ref{borel_cantelli} that \( P(X_n = X'_n) \) for a.a. \( n \) = 1. The proposition is an easy consequence of this observation.

\textbf{Remark 23.24.} 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 \ref{d}) 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.

\textbf{Lemma 23.25.} Suppose that \( X : \Omega \to \mathbb{R} \) is a random variable, then

\[
E |X|^p = \int_{0}^{\infty} ps^{p-1} P(|X| \geq s) \, ds = \int_{0}^{\infty} ps^{p-1} P(|X| > s) \, ds.
\]
Lemma 23.26. If $X$ is a random variable and $\varepsilon > 0$, then
\[
\sum_{n=1}^{\infty} P(|X| \geq n\varepsilon) \leq \frac{1}{\varepsilon} E|X| \leq \sum_{n=0}^{\infty} P(|X| \geq n\varepsilon).
\] (23.9)

Proof. First observe that for all $y \geq 0$ we have,
\[
\sum_{n=1}^{\infty} 1_{n\leq y} \leq y \leq \sum_{n=1}^{\infty} 1_{n\leq y} + 1 = \sum_{n=0}^{\infty} 1_{n\leq y}.
\] (23.10)
Taking $y = |X|/\varepsilon$ in Eq. (23.10) and then take expectations gives the estimate in Eq. (23.9).

Proposition 23.27. Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables, then the following are equivalent:
1. $E|X_1| < \infty$.
2. There exists $\varepsilon > 0$ such that $\sum_{n=1}^{\infty} P(|X_1| \geq \varepsilon n) < \infty$.
3. For all $\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|X_1| \geq \varepsilon n) < \infty$.
4. $\{X_n\}_{n=1}^{\infty}$ and $\{X_n 1_{|X_n| \leq n}\}_{n=1}^{\infty}$ are tail equivalent.
5. $\lim_{n \to \infty} |X_n|/n = 0$ a.s.

Proof. The equivalence of items 1., 2., and 3. easily follows from Lemma 23.26 and their equivalence with item 4. is explained in Remark 23.24. So to finish the proof it suffices to show 3. is equivalent to 5. To this end we start by noting that $\lim_{n \to \infty} |X_n|/n = 0$ a.s. iff
\[
0 = P\left(\frac{|X_n|}{n} \geq \varepsilon \text{ i.o.}\right) = P(|X_n| \geq n\varepsilon \text{ i.o.}) \text{ for all } \varepsilon > 0.
\] (23.11)
Because $\{|X_n| \geq n\varepsilon\}_{n=1}^{\infty}$ are independent sets, the Borel zero-one law (Lemma 13.47) shows the statement in Eq. (23.11) is equivalent to $\sum_{n=1}^{\infty} P(|X_n| \geq n\varepsilon) < \infty$ for all $\varepsilon > 0$.

Corollary 23.28. 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 $\mu := E|X_n| = c$.
\[ \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E} \left[ |X|^2 : 1_{|X| \leq n} \right] = \mathbb{E} \left[ |X|^2 \sum_{n=1}^{\infty} \frac{1}{n^2} 1_{|X| \leq n} \right] = \mathbb{E} \left[ |X|^2 \varphi(|X|) \right] \leq 2 \mathbb{E} \left[ |X|^2 \left( \frac{1}{|X|} \wedge 1 \right) \right] \leq 2 \mathbb{E} |X| . \]

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 19.10 and Example 21.82.

**Theorem 23.31 (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 \( X_n \in L^1(\mathbb{P}) \) and \( \mathbb{E} X_n = \mu \) has already been proved in Corollary 23.28. So let us now assume \( X_n \in L^1(\mathbb{P}) \) and let \( \mu := \mathbb{E} X_n \).

Let \( X'_n := X_n 1_{|X_n| \leq n} \). By Lemma 23.26,

\[ \sum_{n=1}^{\infty} P(X'_n \neq X_n) = \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \leq \mathbb{E} |X_1| < \infty , \]

and hence \( \{X_n\} \) and \( \{X'_n\} \) are tail equivalent. Therefore, by Proposition 23.23 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 23.30

\[ \sum_{n=1}^{\infty} \frac{\text{Var}(X'_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{\mathbb{E} |X'_n|^2}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbb{E} |X_n|^2 1_{|X_n| \leq n}}{n^2} \]

\[ = \sum_{n=1}^{\infty} \frac{\mathbb{E} |X_1|^2 1_{|X_1| \leq n}}{n^2} \leq 2 \mathbb{E} |X_1| < \infty . \quad (23.12) \]

Therefore by Kolmogorov’s convergence criteria, Theorem 23.11

\[ \sum_{n=1}^{\infty} \frac{X'_n - \mathbb{E} X'_n}{n} \]

is almost surely convergent.

Kronecker’s Lemma 23.13 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

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} X'_k = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} [X_k 1_{|X_k| \leq k}] = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} [X_1 1_{|X_1| \leq k}] = \mu . \]

Here we have used the dominated convergence theorem to see that \( a_k := \mathbb{E} [X_1 1_{|X_1| \leq k}] \to \mu \) 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 23.32.** 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 \( X''_M := X_n \wedge M \) and \( S''_M := \sum_{i=1}^{n} X''_M \). It follows from Theorem 23.31 that \( \frac{1}{n} S''_M \to \mu'' := \mathbb{E} X''_1 \) a.s. Since \( S_n \geq S''_M \), we may conclude that

\[ \lim_{n \to \infty} \frac{S_n}{n} \geq \liminf_{n \to \infty} \frac{1}{n} S''_M = \mu'' \text{ a.s.} \]

Since \( \mu'' \to \infty \) 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 23.4 (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 23.31 show for any bounded sequence \( \{c_n\}_{n=1}^{\infty} \) of real numbers that

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} c_k X_k = 0 \text{ a.s.} \]

### 23.3.1 Strong Law of Large Number Examples

**Example 23.33 (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 \( n \)th – bulb burns out. (We assume the bulbs are replaced immediately on burning out.) By convention, we set \( S_0 = 0 \).

Let

\[ N_t := \sup \{ n \geq 0 : S_n \leq t \} \]

denote the number of bulbs which have burned out up to time \( t \). Since \( \mathbb{E} X_i < \infty \), \( X_i < \infty \text{ a.s.} \) and therefore \( S_n < \infty \text{ a.s. for all } n \). From this observation it follows that \( N_t \uparrow \infty \) on the set, \( \Omega_1 := \bigcap_{i=1}^{\infty} \{X_i < \infty\} \) – a subset of \( \Omega \) with full measure.

It is reasonable to guess that \( N_t \sim t/\mu \) and indeed we will show:

\[ \lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mu} \text{ a.s.} \quad (23.13) \]
To prove Eq. (23.13), by the SSLN, if \( \Omega_0 := \{ \lim_{n \to \infty} \frac{1}{n} S_n = \mu \} \) then \( P(\Omega_0) = 1 \). From the definition of \( N_t, S_{N_t} \leq t < S_{N_t+1} \) and so

\[
\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t}.
\]

For \( \omega \in \Omega_0 \cap \Omega_1 \) we have

\[
\mu = \lim_{t \to \infty} \frac{S_{N_t}(\omega)}{N_t(\omega)} \leq \liminf_{t \to \infty} \frac{t}{N_t(\omega)} = \limsup_{t \to \infty} \frac{t}{N_t(\omega)} \leq \limsup_{t \to \infty} \left[ \frac{S_{N_t}(\omega) + 1}{N_t(\omega) + 1} + \frac{1}{N_t(\omega)} \right] = \mu.
\]

**Example 23.34 (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 \( i \)-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

\[
\lim_{t \to \infty} \frac{1}{t} R_t = \frac{\mathbb{E}X_1}{\mathbb{E}X_1 + \mathbb{E}Y_1} = \frac{\mu}{\mu + \nu}.
\]

To prove this, let \( S_n := \sum_{i=1}^{n} (X_i + Y_i) \) be the time that the \( n \)-th bulb is replaced and

\[
N_t := \text{sup} \{ n \geq 0 : S_n \leq t \}
\]

denote the number of bulbs which have burned out up to time \( n \). By Example 23.33 we know that

\[
\lim_{t \to \infty} \frac{1}{t} N_t = \frac{1}{\mu + \nu} \quad \text{a.s., i.e. } N_t = \frac{1}{\mu + \nu} t + o(t) \quad \text{a.s.}
\]

Let us now set \( \tilde{R}_t = \sum_{i=1}^{N_t} X_i \) and observe that

\[
\tilde{R}_t \leq R_t \leq \tilde{R}_t + X_{N_t+1}.
\]

By Proposition 23.27 we know that \( X_n/n \to 0 \) a.s. and therefore,

\[
\lim_{t \to \infty} \frac{X_{N_t+1}}{t} = \lim_{t \to \infty} \left[ \frac{X_{N_t+1}}{N_t} + \frac{N_t}{t} \right] = 0 \cdot \frac{1}{\mu + \nu} = 0 \quad \text{a.s.}
\]

Thus it follows that \( \lim_{t \to \infty} \frac{1}{t} \tilde{R}_t = \lim_{t \to \infty} \frac{1}{t} R_t \) a.s. and the latter limit may be computed using the strong law of large numbers;

\[
\frac{1}{t} \tilde{R}_t = \frac{1}{t} \sum_{i=1}^{N_t} X_i = \frac{N_t}{t} \cdot \frac{1}{N_t} \sum_{i=1}^{N_t} X_i \to \frac{1}{\mu + \nu} \quad \text{a.s.}
\]

**Theorem 23.35 (Glivenko-Cantelli Theorem).** Suppose that \( \{X_n\}_{n=1}^{\infty} \) are i.i.d. random variables and \( F(x) := P(X_1 \leq x) \). Further let \( \mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \) be the empirical distribution with empirical distribution function,

\[
F_n(x) := \mu_n \left( (-\infty, x] \right) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x}.
\]

Then

\[
\lim_{n \to \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \text{ a.s.}
\]

**Proof.** Since \( \{1_{X_i \leq x}\}_{i=1}^{\infty} \) are i.i.d random variables with \( \mathbb{E}1_{X_i \leq x} = P(X_1 \leq x) = F(x) \), it follows by the strong law of large numbers that

\[
\lim_{n \to \infty} F_n(x) = F(x) \text{ a.s. for all } x \in \mathbb{R}.
\]

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 \( \{1_{X_i \leq x}\} \) in order to conclude that, for all \( x \in \mathbb{R} \),

\[
\lim_{n \to \infty} F_n(x-) = F(x-) \text{ a.s. for all } x \in \mathbb{R}.
\]

Keep in mind that the exceptional set of probability zero depend on \( x \).

Given \( k \in \mathbb{N} \), let \( A_k := \left\{ \frac{i}{k} : i = 1, 2, \ldots, k-1 \right\} \) and let \( x_i := \inf \{ x : F(x) \geq i/k \} \) for \( i = 1, 2, \ldots, k-1 \), see Figure 23.3. 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. (23.14) and (23.15) hold for \( x \in \{ x_i : 1 \leq i \leq k-1 \} \). 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 \( n \geq N(\omega), 1 \leq i \leq k-1 \), 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 \( \Omega_0 \subseteq \Omega \) 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 \( r \leq x \leq s \) with \( r, s \in \mathbb{Q} \), we have

\[
F(r) = \lim_{n \to \infty} F_n(r) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq \lim_{n \to \infty} F_n(s) = F(s)
\]

We may now let \( s \downarrow x \) and \( r \uparrow x \) to conclude, on \( \Omega_0 \), on

\[
F(x) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x) \quad \text{for all } x \in \mathbb{R},
\]

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 \( x \in \mathbb{R} \).
From this it follows on\footnote{In fact if }\(F(x) = \delta_0([\infty, x]) = 1_{x \geq 0}\), then \(x_1 = \ldots = x_{k-1} = 0\) for all \(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 \(1/k\). See Figure 23.1. Now suppose \(i\) has been chosen so that \(x_{i-1} < x_i\) and let \(x \in (x_{i-1}, x_i)\). Then we have for \(\omega \in \Omega_k\) and \(n \geq N(\omega)\) that

\[
F_n(x) \leq F_n(x_i) - F(x_i) - 1/k \leq F(x) + 2/k
\]

and

\[
F_n(x) \geq F_n(x_{i-1}) \geq F(x_{i-1}) - 1/k \geq F(x_i) - 2/k \geq F(x) - 2/k.
\]

From this it follows on \(\Omega_k\) that \(|F(x) - F_n(x)| \leq 2/k\) for \(n \geq N\) and therefore,

\[
\sup_{x \in \mathbb{R}} |F(x) - F_n(x)| \leq 2/k.
\]

Hence it follows on \(\Omega_0 := \cap_{k=1}^{\infty} \Omega_k\) (a set with \(P(\Omega_0) = 1\)) that

\[
\lim_{n \to \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0.
\]

23.4 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 23.42 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 \(|X_k| \leq c\) a.s. for all \(k \in \mathbb{N}\). The main result here is Corollary 23.40 which states that \(\sum_{k=1}^{\infty} E X_k\) is a.s. convergent iff \(\sum_{k=1}^{\infty} E X_k + \sum_{k=1}^{\infty} \text{Var}(X_k)\) are convergent in \(\mathbb{R}\). Kolmogorov’s convergence criteria in Theorem 23.11 gives one direction. The converse direction will be based on the following (sub-martingale) inequality.

Theorem 23.36. Suppose that \(S_n := \sum_{k=1}^{n} X_k\) where \(\{X_k\}_{k=1}^{\infty}\) are independent random variables such that \(|X_k| \leq c < \infty\) and \(E X_k = 0\) for all \(k \in \mathbb{N}\). Then for all \(\lambda > 0\),

\[
P\left(\sup_n |S_n| \leq \lambda\right) \cdot \sum_{k=1}^{\infty} \text{Var}(X_k) \leq (\lambda + c)^2.
\]

In particular if \(P(\sup_n |S_n| < \infty) > 0\), then \(\sum_{k=1}^{\infty} \text{Var}(X_k) \leq \infty\) 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 \(S_0 = 0, B_n := \sigma(S_0, \ldots, S_n)\), 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 \(A_n = \mathbb{E} S_n^2 = \sum_{k=1}^{n} \text{Var}(X_k)\), see Examples 21.3 and 21.19. For completeness, here is a proof that \(\{M_n := S_n^2 - E S_n^2\}\) is a martingale;

\[
E_{B_n}[M_{n+1}] = E_{B_n}[S_n^2 - A_{n+1}] = E_{B_n}\left((S_n + X_{n+1})^2 - E S_n^2\right) = S_n^2 + 2 S_n E X_{n+1} + E S_n^2 - E S_n^2 = S_n^2 + 2 S_n E X_{n+1} + E X_{n+1} - E S_n^2 = S_n^2 - E S_n^2 = M_n.
\]

We now have

\[
E\left[S_{\tau \wedge N}^{2}\right] = E\left[S_{\tau}^{2} : \tau > N\right] + E\left[S_{\tau}^{2} : \tau \leq N\right] \leq \lambda^2 P(\tau > N) + E\left[(S_{\tau-1} + c)^2 : \tau \leq N\right] \leq \lambda^2 P(\tau > N) + (\lambda + c)^2 \lambda P(\tau \leq N) \leq (\lambda + c)^2.
\]

On the other hand, by the optional sampling theorem,

\[
E\left[S_{\tau \wedge N}^{2}\right] = E\left[M_{\tau \wedge N}\right] + E\left[A_{\tau \wedge N}\right] = E\left[M_0\right] + E\left[A_{\tau \wedge N}\right] = E\left[A_{\tau \wedge N}\right].
\]

Putting these together shows

\[
A_{\infty} P(\tau = \infty) \leq E\left[A_{\tau}\right] = \lim_{N \to \infty} E\left[A_{\tau \wedge N}\right] \leq (\lambda + c)^2
\]

which is equivalent to Eq. (23.16).
If we now further assume $P\left(\sup_n |S_n| < \infty\right) > 0$ then $P\left(\sup_n |S_n| \leq \lambda\right) > 0$ for sufficiently large $\lambda$ and it follows from Eq. (23.16) that $\sum_{k=1}^{\infty} \text{Var} (X_k) < \infty$. 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 23.37.** 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 $Y_{n+1} \leq f (Y_n)$ 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

$$E \left[ A_\infty : \sup_n Y_n \leq \lambda \right] \leq \lambda \vee f (\lambda)$$

(23.17)

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,

$$E X_{\tau \wedge N} = EM_{\tau \wedge N} + EA_{\tau \wedge N} = EM_0 + EA_{\tau \wedge N}.$$  

Using $X_N \leq \lambda$ on $\{\tau > N\}$ and $X_{\tau} \leq f (X_{\tau-1}) \leq f (\lambda)$ on $\{\tau \leq N\}$ allows us to find the estimate,

$$EA_{\tau \wedge N} = E X_{\tau \wedge N} = E [X_{\tau} : \tau \leq N] + E [X_N : \tau > N] \leq f (\lambda) P (\tau \leq N) + \lambda P (\tau > N) \leq \lambda \vee f (\lambda).$$

Letting $N \uparrow \infty$ in this estimate the shows

$$E [A_\infty : \tau = \infty] \leq EA_{\tau} \leq \lambda \vee f (\lambda)$$

which gives Eq. (23.17) since $\{\tau = \infty\} = \{\sup_n Y_n \leq \lambda\}$.

**Remark 23.38.** The above proof also shows

$$E \left[ A_N : \max_{n \leq N} Y_n \leq \lambda \right] \leq EA_{\tau \wedge N} \leq \lambda \vee f (\lambda) \ \forall \ N \in \mathbb{N}.$$

**Corollary 23.39.** Suppose that $\{X_n\}_{n=0}^{\infty}$ are independent random variables which are bounded by some $c < \infty$ and have mean zero, $E X_n = 0$ for all $n$. Then $\sum_{k=1}^{\infty} X_k$ exists a.s. iff $\sum_{k=1}^{\infty} \text{Var} (X_k) < \infty$.

**Proof.** We need only prove the forward direction since Theorem 23.11 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 $A_n = \mathbb{E} S_n^2 = \sum_{k=1}^{n} \text{Var} (X_k)$, see Examples 21.3 and 21.19.

Moreover

$$Y_{n+1} = S_{n+1}^2 \leq (|S_n| + c)^2 = (\sqrt{Y_n} + c)^2 = f (Y_n)$$

where $f (y) := (\sqrt{y} + c)^2$. Since we are assuming

$$\lim_{n \to \infty} Y_n = \lim_{n \to \infty} S_n^2 = \left( \sum_{k=1}^{\infty} X_k \right)^2$$

exists a.s.,

it follows that $P (\{\sup_n Y_n < \infty\}) = 1$ and so by Theorem 23.37, $A_\infty = \sum_{k=1}^{\infty} \text{Var} (X_k) < \infty$ a.s. This completes the proof since $A_\infty = \sum_{k=1}^{\infty} \text{Var} (X_k)$ is not random.

**Corollary 23.40.** 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 $\sum_{k=1}^{\infty} \text{Var} (X_k) < \infty$ 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 23.11 it suffices to prove the forward direction. Our goal is to make use of Corollary 23.39. 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 $|\hat{X}_n| \leq 2c$, $E \hat{X} = 0$, and $\text{Var} (\hat{X}_k) = 2 \text{Var} (X_n)$. 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 23.36 or Corollary 23.39

$$\sum_{k=1}^{\infty} \text{Var} (X_k) = \frac{1}{2} \sum_{k=1}^{\infty} \text{Var} (\hat{X}_k) < \infty.$$
We may now apply Theorem 23.11 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 23.41. Let \( \{Z_n\}_{n=1}^{\infty} \) be i.i.d. mean zero random variables, \( |Z_n| \leq c < \infty \) for some \( n \). Further assume \( \sigma^2 = \mathbb{E}Z_n^2 = \text{Var}(Z_n) > 0 \). 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 \( |X_n| \leq c \) and \( \mathbb{E}X_n = 0 \) for all \( n \). Moreover \( \text{Var}(X_n) = \frac{1}{n^{2p}} \sigma^2 \) and so
\[
\sum_{n=1}^{\infty} \text{Var}(X_n) = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^{2p}}
\]
which is convergent iff \( 2p > 1 \), i.e. \( p > 1/2 \). Thus according to Corollary 23.40 \( \sum_{n=1}^{\infty} \frac{1}{n^p} Z_n \) 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 \( c \leq \mathbb{E}Z_n^2 \leq c^{-1} \) for all \( n \).

We are now ready to state and prove Kolmogorov’s Three Series Theorem.

**Theorem 23.42 (Kolmogorov’s Three Series Theorem).** Suppose that \( \{X_n\}_{n=1}^{\infty} \) are independent random variables and for \( c > 0 \) any \( c > 0 \) let \( X_n^c := X_n 1_{|X_n| \leq c} \). 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. \( \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c) < \infty \),
2. \( \sum_{n=1}^{\infty} \text{Var}(X_n^c) < \infty \), 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 \( 0 = \mathbb{P}(X_n \neq X_n^c \text{ i.o.}) = \mathbb{P}(|X_n| > c \text{ i.o.}) \) which by the Borel Cantelli Lemma happens iff \( \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c) < \infty \).

(\( \Leftarrow \)) 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 23.40 happens iff the sums in items 2. and 3. of the theorem are convergent.

(\( \Rightarrow \)) 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 \( \sum_{n=1}^{\infty} \mathbb{P}(|X_n^c| > c) < \infty \) 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 23.40. [Another proof of this direction may be found in Chapter 26 see Theorem 26.17]}

**Remark 23.43.** We have seen another necessary and sufficient condition in Exercise 21.22 namely \( \sum_{n=1}^{\infty} X_n \), is almost surely convergent in \( \mathbb{R} \) iff \( \sum_{n=1}^{\infty} X_n^c \) is convergent in distribution. We will also see below that \( \sum_{n=1}^{\infty} X_n \), is almost surely convergent in \( \mathbb{R} \) if \( \sum_{n=1}^{\infty} X_n \), is convergent in probability, see Lévy’s Theorem 23.50 below.

**Exercise 23.5 (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 \( X_n \geq 0 \) are independent, then \( \sum_{n=1}^{\infty} X_n < \infty \) a.s. iff for any \( c > 0 \) we have
\[
\sum_{n} \mathbb{P}(X_n > c) < \infty \text{ and } \sum_{n} \mathbb{E}[X_n 1_{X_n \leq c}] < \infty,
\]
that is it is unnecessary to verify the convergence of the second series in Theorem 23.42 involving the variances.

**23.4.1 Examples**

**Lemma 23.44.** 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 “\( \Leftarrow \)” is true without the assumption that the \( Y_n \) are normal random variables as pointed out in Theorem 23.11. To prove the converse directions we will make use of the Kolmogorov’s three series Theorem 23.42. Namely, if \( \sum_{j=1}^{\infty} Y_j \) converges a.s. then the three series in Theorem 23.42 converge for all \( c > 0 \).

1. Since \( Y_n \overset{d}{=} \sigma_n^2 + \mu_n \), we have for any \( c > 0 \) that
\[
\sum_{n=1}^{\infty} \mathbb{P}(|\sigma_n^2 + \mu_n| > c) < \infty.
\]

If \( \lim_{n \to \infty} \mu_n \neq 0 \) then there is a \( c > 0 \) such that either \( \mu_n \geq 2c \) for infinitely many \( n \) or \( \mu_n \leq -2c \) for infinitely many \( n \). It then follows that either \( \{N > 0 \} \subset \)
which would imply then there exists $\alpha < c$ such that

$$
N \geq \alpha \subset \{ |\sigma_n N + \mu_n | > c \} \quad n \text{ i.o.}
$$

which would imply $P (|\sigma_n N + \mu_n | > 1) \geq 1/2 \text{ n - i.o.}$ which would violate Eq. (23.20) and so we may conclude 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

$$
N \geq \alpha \subset \{ |\sigma_n N + \mu_n | > c \} \quad n \text{ i.o.}
$$

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

$$
\lim_{n \to \infty} \sum_{n=1}^{\infty} \chi_n = 0
$$

If we can show

$$
\lim_{n \to \infty} \sum_{n=1}^{\infty} \chi_n = 0
$$

it would then follow from Eq. (23.21) that $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. We may now use Kolmogorov's convergence criteria (Theorem 23.11) to infer that $\sum_{n=1}^{\infty} \text{Var}(Y_n - \mu_n) \leq \sum_{n=1}^{\infty} \text{Var}(Y_n - \mu_n)$ and hence

$$
\sigma_n^2 = \text{Var}(\mu_n) + \text{Var}(\sum_{n=1}^{\infty} \chi_n)
$$

Therefore, using Var $(\sigma X + \mu Y) = \sigma^2 \text{Var}(X) + \mu^2 \text{Var}(Y) + 2\sigma \mu \text{Cov}(X, Y)$, we find

$$
\text{Var}(\sum_{n=1}^{\infty} \chi_n) = \sum_{n=1}^{\infty} \chi_n^2 \text{Var}(\mu_n) + \sum_{n=1}^{\infty} \text{Var}(\chi_n)
$$

Making use of the estimate, $2ab \leq a^2 + b^2$ valid for all $a, b \geq 0$, it follows that

$$
\text{Var}(\sum_{n=1}^{\infty} \chi_n) \geq \sum_{n=1}^{\infty} \text{Var}(\chi_n) = \sum_{n=1}^{\infty} \chi_n^2 \text{Var}(\mu_n)
$$

An alternative proof that $\sum_{n=1}^{\infty} \mu_n$ is convergent using the the third series in Theorem 23.42 For all $c > 0$ the third series implies

$$
\sum_{n=1}^{\infty} \sum_{\chi_n \leq c} \text{Var}(\mu_n) \leq \sum_{n=1}^{\infty} \sum_{\chi_n \leq c} \text{Var}(Y_n) < \infty
$$

where $\delta_n := \mathbb{E} (\mu_n \mid \{ \sigma_n N + \mu_n \leq c \})$ and $\beta_n := \mathbb{E} (\mu_n \mid \{ \sigma_n N + \mu_n \leq c \})$. 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,

$$
\sum_{n=1}^{\infty} |\mu_n (\beta_n - 1)| \leq C \sum_{n=1}^{\infty} |\mu_n| \sigma_n^2 < \infty
$$

Example 23.45. As another simple application of Theorem 23.42, let us use it to give a proof of Theorem 23.11. We will apply Theorem 23.42 with $X_n := Y_n - EY_n$. We need to then check the three series in the statement of Theorem 23.42 converge. For the first series we have by the Markov inequality,

$$
\sum_{n=1}^{\infty} P (|X_n| > c) \leq \sum_{n=1}^{\infty} \frac{1}{c^2} \mathbb{E}|X_n|^2 = \frac{1}{c^2} \sum_{n=1}^{\infty} \mathbb{E}|Y_n|^2 < \infty
$$

For the second series, observe that

$$
\sum_{n=1}^{\infty} \text{Var}(X_n \mid |X_n| \leq c) \leq \sum_{n=1}^{\infty} \mathbb{E} \left[ \left( X_n \mid |X_n| \leq c \right)^2 \right] \leq \sum_{n=1}^{\infty} \mathbb{E}[X_n^2] = \sum_{n=1}^{\infty} \text{Var}(Y_n) < \infty
$$

and we estimate the third series as;

$$
\sum_{n=1}^{\infty} \mathbb{E} (X_n \mid |X_n| \leq c) \leq \sum_{n=1}^{\infty} \mathbb{E} \left( \frac{1}{c} |X_n|^2 \mid |X_n| \leq c \right) \leq \frac{1}{c} \sum_{n=1}^{\infty} \mathbb{E}|Y_n|^2 < \infty.
$$
23.5 Maximal Inequalities

Theorem 23.46 (Kolmogorov’s Inequality). Let \( \{X_n\} \) be a sequence of independent random variables with mean zero, \( S_n := X_1 + \cdots + X_n \), and \( S_n^* := \max_{j \leq n} |S_j| \). Then for any \( \alpha > 0 \) we have

\[
P ( S_n^* \geq \alpha ) \leq \frac{1}{\alpha^2} \mathbb{E} \left[ S_n^2 : S_n^* \geq \alpha \right].
\]

(23.23)

**Proof. First proof.** As \( \{S_n\}_{n=1}^{\infty} \) is a martingale relative to the filtration, \( B_n = \sigma(S_1, \ldots, S_n) \), the inequality in Eq. (23.23) is a special case of Proposition 21.4 with \( X_n = S_n^2 \), also see Example 21.49.

*Second direct proof.* Let \( \tau = \inf \{ j : |S_j| \geq \alpha \} \) with the infimum of the empty set being taken to be equal to \( \infty \). Observe that

\[
\{ \tau = j \} = \{ |S_1| < \alpha, \ldots, |S_{j-1}| < \alpha, |S_j| \geq \alpha \} \in \sigma(X_1, \ldots, X_j).
\]

Now

\[
\mathbb{E} \left[ S_n^2 : |S_n^*| > \alpha \right] = \mathbb{E} \left[ S_n^2 : \tau \leq N \right] = \sum_{j=1}^{N} \mathbb{E} \left[ S_n^2 : \tau = j \right]
\]

\[
= \sum_{j=1}^{N} \mathbb{E} \left[ (S_j + S_N - S_j)^2 : \tau = j \right]
\]

\[
= \sum_{j=1}^{N} \mathbb{E} \left[ S_j^2 + (S_N - S_j)^2 + 2S_j(S_N - S_j) : \tau = j \right]
\]

\[
= \sum_{j=1}^{N} \mathbb{E} \left[ S_j^2 + (S_N - S_j)^2 : \tau = j \right]
\]

\[
\geq \sum_{j=1}^{N} \mathbb{E} \left[ S_j^2 : \tau = j \right] = \alpha^2 \sum_{j=1}^{N} P [\tau = j] = \alpha^2 P (|S_n^*| > \alpha).
\]

The equality, (\( \ast \)), is a consequence of the observations: 1) \( 1_{\tau = j}S_j \) is \( \sigma(X_1, \ldots, X_j) \) measurable, 2) \( S_n - S_j \) is \( \sigma(X_{j+1}, \ldots, X_n) \) measurable and hence \( 1_{\tau = j}S_j \) and \( (S_n - S_j) \) are independent, and so 3)

\[
\mathbb{E} \left[ S_j(S_N - S_j) : \tau = j \right] = \mathbb{E} \left[ S_j 1_{\tau = j}(S_N - S_j) \right] = \mathbb{E} \left[ S_j1_{\tau = j} \right] \cdot \mathbb{E} \left[ S_N - S_j \right] = \mathbb{E} \left[ S_j1_{\tau = j} \right] \cdot 0 = 0.
\]

Remark 23.47 (Another proof of Theorem 23.11). Suppose that \( \{Y_j\}_{j=1}^{\infty} \) are independent random variables such that \( \sum_{j=1}^{\infty} \text{Var}(Y_j) < \infty \) 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 23.46 for all \( M < N \),

\[
P \left( \max_{M \leq j \leq N} |S_j - S_M| \geq \alpha \right) \leq \frac{1}{\alpha^2} \mathbb{E} \left[ (S_N - S_M)^2 \right] = \frac{1}{\alpha^2} \sum_{j=M+1}^{N} \mathbb{E} \left[ X_j^2 \right]
\]

\[
= \frac{1}{\alpha^2} \sum_{j=M+1}^{N} \text{Var}(X_j).
\]

Letting \( N \to \infty \) in this inequality shows, with \( Q_M := \sup_{j \geq M} |S_j - S_M| \),

\[
P (Q_M \geq \alpha) \leq \frac{1}{\alpha^2} \sum_{j=M+1}^{\infty} \text{Var}(X_j).
\]

Since

\[
\delta_M := \sup_{j,k \geq M} |S_j - S_k| \leq \sup_{j,k \geq M} \left( |S_j - S_M| + |S_M - S_k| \right) \leq 2Q_M
\]

we may further conclude,

\[
P (\delta_M \geq 2\alpha) \leq \frac{1}{\alpha^2} \sum_{j=M+1}^{\infty} \text{Var}(X_j) \to 0 \text{ as } M \to \infty,
\]

i.e. \( \delta_M \xrightarrow{p} 0 \) as \( M \to \infty \). Since \( \delta_M \) is decreasing in \( M \), it follows that \( \lim_{M \to \infty} \delta_M =: \delta \) exists and because \( \delta_M \xrightarrow{p} 0 \) 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. This gives a second proof of Kolmogorov’s convergence criteria in Theorem 23.11.

Corollary 23.48 (\( L^2 - \text{SSLN} \)). 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

\[
\frac{1}{n^p} S_n \to 0 \text{ a.s.}
\]

If \( \{Y_n\} \) is a sequence of independent random variables \( \mathbb{E}Y_n = \mu \) and \( \sigma^2 = \text{Var}(X_n) < \infty \), then for any \( \beta \in (0, 1/2) \),

\[
\frac{1}{n^\beta} \sum_{k=1}^{n} Y_k - \mu = O \left( \frac{1}{n^\beta} \right).
\]
Proof. (The proof of this Corollary may be skipped as it has already been proved, see Corollary 23.15) From Theorem 23.46 we have for every $\varepsilon > 0$ that

$$P \left( \frac{S_n^*}{N_n^p} \geq \varepsilon \right) = P \left( S_n^* \geq \varepsilon N_n^p \right) = \frac{1}{\varepsilon^2 N_n^p} \mathbb{E} \left[ S_n^2 \right] = \frac{1}{\varepsilon^2 N_n^p} C N = \frac{C}{\varepsilon^2 N_n^{p-1}}.$$ 

Hence if we suppose that $N_n = n^\alpha$ with $\alpha (2p - 1) > 1$, then we have

$$\sum_{n=1}^{\infty} P \left( \frac{S_n^*}{N_n^p} \geq \varepsilon \right) \leq \sum_{n=1}^{\infty} \frac{C}{\varepsilon^2 n^\alpha (2p - 1)} < \infty,$$

and so by the first Borel–Cantelli lemma we have

$$P \left( \left\{ \frac{S_n^*}{N_n^p} \geq \varepsilon \text{ for } n \text{ i.o.} \right\} \right) = 0.$$ 

From this it follows that $\lim_{n \to \infty} \frac{S_n^*}{N_n^p} = 0$ a.s.

To finish the proof, for $m \in \mathbb{N}$, we may choose $n = n (m)$ such that

$$n^\alpha \geq N_n \geq m < N_{n+1} = (n + 1)^\alpha.$$ 

Since

$$\frac{S_{n(m+1)}^*}{N_{n(m+1)}^p} \leq \frac{S_{n(m)}^*}{m^p} \leq \frac{S_{n(m+1)}^*}{N_{n(m)}^p},$$

and

$$N_{n+1}/N_n \to 1 \text{ as } n \to \infty,$$

it follows that

$$0 = \lim_{m \to \infty} \frac{S_{n(m)}^*}{N_{n(m)}^p} = \lim_{m \to \infty} \frac{S_{n(m)}^*}{N_{n(m)}^p} \leq \lim_{m \to \infty} \frac{S_{n(m)}^*}{m^p} \leq \lim_{m \to \infty} \frac{S_{n(m)}^*}{N_{n(m)}^p} = 0 \text{ a.s.}$$

That is $\lim_{m \to \infty} \frac{S_{n(m)}^*}{m^p} = 0$ a.s.

We are going to give three more maximal inequalities before ending this section. In all cases 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 $S_n := \sum_{k=1}^{n} X_k$ and $S_n^* := \max_{1 \leq k \leq n} \|S_k\|$. If $\tau$ is any $B^X$ – stopping time and $f$ is a non-negative function on $Y$, then

$$\mathbb{E} \left[ f \left( S_n - S_\tau \right) : \tau \leq n \right] = \sum_{k=1}^{n} \mathbb{E} \left[ f \left( S_n - S_k \right) : \tau = k \right] = \sum_{k=1}^{n} \mathbb{E} \left[ f \left( S_n - S_k \right) \right] P(\tau = k).$$

Theorem 23.49 (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

$$P \left( \|S_N\| \geq \alpha \right) \geq (1 - c_N (\alpha)) P \left( S_N^* \geq 2\alpha \right) \quad \text{and}$$

$$P \left( \|S_N\| > \alpha \right) \geq (1 - c_N (\alpha)) P \left( S_N^* > 2\alpha \right)$$

where

$$c_N (\alpha) := \max_{1 \leq k \leq N} P \left( \|S_N - S_k\| > \alpha \right).$$

Proof. We only prove Eq. (23.25) since the proof of Eq. (23.26) is similar and in fact can be deduced from Eq. (23.25) 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,

$$\|S_N\| = \|S_\tau + S_N - S_\tau\| \geq \|S_\tau\| - \|S_N - S_\tau\| \geq 2\alpha - \|S_N - S_\tau\|.$$ 

From this it follows that

$$\{\tau \leq N \& \|S_N - S_\tau\| \leq \alpha\} \subset \{\|S_N\| \geq \alpha\}$$

and therefore,

$$P \left( \|S_N\| \geq \alpha \right) \geq P \left( \tau \leq N \& \|S_N - S_\tau\| \leq \alpha \right)$$

$$= \sum_{k=1}^{N} P(\tau = k) \cdot P \left( \|S_N - S_k\| \leq \alpha \right)$$

$$\geq \min_{1 \leq k \leq N} P \left( \|S_N - S_k\| \leq \alpha \right) \cdot \sum_{k=1}^{N} P(\tau = k)$$

$$= \min_{1 \leq k \leq N} \left[ 1 - P \left( \|S_N - S_k\| > \alpha \right) \right] \cdot \sum_{k=1}^{N} P(\tau = k)$$

$$= \left[ 1 - c_N (\alpha) \right] \cdot P \left( S_N^* \geq 2\alpha \right).$$

As an application of Theorem 23.49 we have the following convergence result.

Theorem 23.50 (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 if $\sum_{n=1}^{\infty} X_n$ 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 $M \in \mathbb{N}$, 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 $\max_{M \leq n \leq N} P(\|S_n - S_j\| > \varepsilon) < \varepsilon$. 

\end{document}
for all $N \geq M$. An application of Skorohod’s inequality (Theorem 23.49), 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, $P(\delta_M > 4\varepsilon) \leq \frac{\varepsilon}{4}$ 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. 

\[\text{Remark 23.51 (Yet another proof of Theorem 23.11). Suppose that } \{Y_j\}_{j=1}^\infty \text{ are independent random variables such that } \sum_{j=1}^\infty \text{Var}(Y_j) < \infty. \text{ By Proposition 23.10, the sum, } \sum_{j=1}^\infty (Y_j - \text{E}Y_j), \text{ is } L^2(P) \text{ convergent and hence convergent in probability. An application of Lévy’s Theorem 23.50 then shows } \sum_{j=1}^\infty (Y_j - \text{E}Y_j) \text{ is almost surely convergent which gives another proof of Kolmogorov’s convergence criteria in Theorem 23.11.} \]

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 23.49) along with Chebyshev’s inequality.

\[\text{Corollary 23.52 (Ottaviani’s maximal inequality). Suppose that } \{X_n\}_{n=1}^\infty \text{ are independent real or Banach valued square integrable random variables. Then for all } \alpha > 0 \text{ we have} \]

$$P(\|S_N\| \geq \alpha) \geq \left(1 - \frac{1}{\alpha^2} \max_{1 \leq k \leq N} \text{E}\|S_N - S_k\|^2\right) P(S_N^* \geq 2\alpha)$$

and in particular if $\alpha^2 > \max_{1 \leq k \leq N} \text{E}\|S_N - S_k\|^2$, then

$$P(S_N^* \geq 2\alpha) \leq \left(1 - \frac{1}{\alpha^2} \max_{1 \leq k \leq N} \text{E}\|S_N - S_k\|^2\right)^{-1} P(\|S_N\| \geq \alpha).$$

If we further assume that $\{X_n\}$ are real (or Hilbert valued) mean zero random variables, then

$$P(S_N^* \geq 2\alpha) \leq \left(1 - \frac{1}{\alpha^2} \text{E}\|S_N - X_1\|^2\right)^{-1} P(\|S_N\| \geq \alpha).$$

\[\text{Proof. The first and second inequalities follow by Chebyshev’s inequality and Skorohod’s Theorem 23.49. When the } \{X_n\} \text{ are real or Hilbert valued mean zero square integrable random variables, we have} \]

$$\max_{1 \leq k \leq N} \text{E}\|S_N - S_k\|^2 = \max_{1 \leq k \leq N} \sum_{j=k+1}^N \text{E}\|X_j\|^2 = \sum_{j=2}^N \text{E}\|X_j\|^2 = \|S_N - X_1\|^2.$$ 

\[\text{Corollary 23.53. Suppose } \lambda > 1 \text{ and } \{X_n\}_{n=1}^\infty \text{ are independent real square integrable random variables with } \text{E}X_n = 0 \text{ and Var}(X_n) = 1 \text{ for all } n. \text{ Then} \]

$$P(\|S_n\| \geq 2\lambda \sqrt{n}) \leq \left(1 - \frac{1}{\lambda^2}\right)^{-1} \cdot P(\|S_n\| \geq \lambda \sqrt{n})$$

and if we further assume that $\{X_n\}_{n=1}^\infty$ are i.i.d, then

$$\lim_{n \to \infty} P(S_n^* \geq 2\lambda \sqrt{n}) \leq \sqrt{\frac{n}{\pi}} \left(1 - \frac{1}{\lambda^2}\right)^{-1} \frac{1}{\lambda} e^{-\lambda^2/2}.$$ 

\[\text{Proof. The first inequality follows from Eq. (23.28) of Corollary 23.52 with } \alpha = \lambda \sqrt{n}. \text{ For the second inequality we use the central limit theorem to conclude that} \]

$$P(\|S_n\| \geq \lambda \sqrt{n}) = P\left(\frac{|S_n|}{\sqrt{n}} \geq \lambda\right) \rightarrow P(|Z| \geq \lambda)$$

where $Z$ is a standard normal random variable. We then estimate $P(|Z| \geq \lambda)$ using the Gaussian tail estimates in Lemma 10.71. \[\text{We can significantly improve on Corollary 23.52 if we further assume that } X_n \text{ is symmetric for } n \text{ in which case the following reflection principle holds.} \]

\[\text{Theorem 23.54. Suppose that } \{X_n\}_{n=1}^\infty \text{ are independent real or Banach valued random variables such that } X_n = -X_n \text{ for all } n \text{ and } \tau \text{ is any stopping time. If we set} \]

$$B_n^\tau = \sigma(X_1, \ldots, X_n) \text{ the stopping time. Then if we set}$$
Proposition 23.55 (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

\[
P(S_N^r \geq r) \leq P(\|S_N\| \geq r) + P(\|S_N\| > r) \leq 2P(\|S_N\| \geq r). \quad (23.29)
\]

Proof. Let \( \tau := \inf \{ n : \|S_n\| \geq r \} \) (a \( \{B_k^X\}_{n=1}^{\infty} \) stopping time), then

\[
P(S_N^r \geq r) = P(S_N^r \geq r, \|S_N\| \geq r) + P(S_N^r \geq r, \|S_N\| < r)
\]

\[
= P(\|S_N\| \geq r) + P(\tau \leq N, \|S_N\| < r).
\]

Moreover by \( \tau \leq N,\|S_N\| < r \), we have

\[
P(\tau \leq N,\|S_N\| \geq r) = P(\tau \leq N,\|S_N - S_\tau\| < r).
\]

If \( \|S_\tau\| \geq r \) 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

\[
P(\tau \leq N,\|S_\tau - (S_N - S_\tau)\| < r) \geq P(\tau \leq N,\|S_N\| > r) = P(\|S_N\| > r).
\]

Combining this inequality with the first displayed inequality in the proof easily gives the result.

Exercise 23.6 (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 \( S_n := \sum_{k \leq n} X_k \) be the standard simple random walk on \( \mathbb{Z} \). Show for every \( r \in \mathbb{N} \) that

\[
P(\max_{k \leq n} S_k \geq r) = P(S_n \geq r) + P(S_n > r).
\]

23.6 Bone Yards

23.6.1 Kronecker’s Lemma

Remark 23.56. Here is a continuous version of Lemma 23.13. 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

\[
\limsup_{s \to \infty} \frac{|X(s)|}{a(s)} = \limsup_{s \to \infty} \left[ \frac{a(s_0) r(s_0) - a(s_0) r(s)}{a(s)} + \frac{1}{a(s)} \int_{s_0}^s r(u) a'(u) du \right]
\]

\[
\leq \limsup_{s \to \infty} \left[ -r(s) + \frac{1}{a(s)} \int_{s_0}^s |r(u)| a'(u) du \right]
\]

\[
\leq \limsup_{s \to \infty} \left[ \frac{a(s) - a(s_0)}{a(s)} \sup_{u \geq s_0} |r(u)| \right] = \sup_{u \geq s_0} |r(u)| \to 0 \text{ as } s_0 \to \infty.
\]
Corollary 23.57. 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
\[
\sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{b_k^2} < \infty
\]
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 23.11,
\[
\sum_{k=1}^{\infty} \frac{X_k - \mathbb{E}X_k}{b_k}
\]
is convergent a.s. and in \( L^2(P) \).

Therefore an application of Kronecker’s Lemma 23.13 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 23.13 we know that
\[
0 = \lim_{n \to \infty} \frac{1}{b_n^2} \sum_{k=1}^{n} \text{Var}(X_k) = \lim_{n \to \infty} \mathbb{E} \left( \frac{S_n - \mathbb{E}S_n}{b_n} \right)^2
\]
which gives the \( L^2(P) \) – convergence statement as well. \( \blacksquare \)

23.6.2 Older variants on the proof of Kolmogorov’s three series Theorem

[The reader should skip this section.

Lemma 23.58. Suppose that \( \{Y_n\}_{n=1}^{\infty} \) are independent random variables such that there exists \( c < \infty \) such that \( |Y_n| \leq c < \infty \) 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,
\[
\sum_{j=1}^{\infty} \mathbb{E} Y_j^2 \leq \frac{(\lambda + c)^2}{P(\sup_n |S_n| \leq \lambda)} \text{ for all } \lambda > 0,
\]
where as usual, \( S_n := \sum_{j=1}^{n} Y_j \).

Remark 23.59. It follows from Eq. (23.30) 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 23.11), \( \sum_{j=1}^{\infty} Y_j = \lim_{n \to \infty} S_n \) 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
\[
\mathbb{E} [S_N^2] \leq \frac{(\lambda + c)^2}{P(\sup_{n \leq N} |S_n| \leq \lambda)} \leq \frac{(\lambda + c)^2}{P(\sup_n |S_n| \leq \lambda)}.
\]
To prove Eq. (23.31), let \( S_0 := 0 \), \( B_n := \sigma(S_0, \ldots, S_n) \), and \( \tau \) be the stopping time,
\[
\tau = \tau_\lambda := \inf \{ n \geq 1 : |S_n| > \lambda \},
\]
where \( \inf \emptyset = \infty \). Recall that \( \{S_n\}_{n=0}^{\infty} \) and \( \{M_n := S_n^2 - \mathbb{E} S_n^2\}_{n=0}^{\infty} \) are martingales \( ^3 \) see Examples 21.13 and 21.19 Simple estimates along with Theorem 21.40 then shows
\[
\mathbb{E} [S_N^2] = \mathbb{E} [S_{N,\tau}\tau] = \mathbb{E} [S_N^2 : \tau > N]
\]
\[
\leq \mathbb{E} [\mathbb{E}_B S_N^2 : \tau \leq N] + \lambda^2 P(\tau > N)
\]
\[
= \mathbb{E} [M_\tau + \mathbb{E} S_N^2 : \tau \leq N] + \lambda^2 P(\tau > N)
\]
\[
= \mathbb{E} [M_\tau : \tau \leq N] + \mathbb{E} [S_N^2] \cdot P(\tau \leq N) + \lambda^2 P(\tau > N),
\]
or equivalently that
\[
\mathbb{E} [S_N^2] \cdot P(\tau > N) \leq \mathbb{E} [M_\tau : \tau \leq N] + \lambda^2 P(\tau > N).
\]
As \( M_n \leq S_n^2 \) and so \( M_\tau \leq S_\tau^2 \leq (\lambda + c)^2 \) on \( \{\tau < \infty\} \cap \{\tau \leq N\} \), we learn that
\[
\mathbb{E} [S_N^2] \leq \frac{(\lambda + c)^2}{P(\tau \leq N)} + \lambda^2 P(\tau > N)
\]
\[
\leq \frac{(\lambda + c)^2}{P(\tau > N)} = \frac{(\lambda + c)^2}{P(\sup_{n \leq N} |S_n| \leq \lambda)}
\]
\[
\leq \frac{(\lambda + c)^2}{P(\sup_n |S_n| \leq \lambda)}
\]
\[\]
\( ^3 \) Here is the proof for \( \{M_n\} \) again;
\[
\mathbb{E}_B [S_{n+1}^2 - S_n^2] = \mathbb{E}_B [(S_{n+1} + S_n)(S_{n+1} - S_n)]
\]
\[
= \mathbb{E}_B [(S_{n+1} - S_n)(S_{n+1} - S_n)]
\]
\[
= \mathbb{E}_B [S_{n+1}^2 - S_n^2].
\]
which proves Eq. (23.31).

Since $S_n$ is convergent a.s., it follows that $P(\sup_n |S_n| < \infty) = 1$ and therefore,

$$\lim_{\lambda \to \infty} P\left(\sup_n |S_n| \leq \lambda\right) = 1.$$ 

Hence for $\lambda$ sufficiently large, $P(\sup_n |S_n| \leq \lambda) > 0$ and we learn from Eq. (23.31) that

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 = \lim_{N \to \infty} \mathbb{E}\left[\left(S_N^2 j = \tau = j\right) = \sum_{j=1}^{N} \mathbb{E}\left[(S_N - S_j)^2 : \tau = j\right]
\right.$$ 

$$\leq \sum_{j=1}^{N} \mathbb{E}\left[(S_{j-1} + Y_j)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] \sum_{j=1}^{N} P[\tau = j]
\right.$$ 

$$\leq \sum_{j=1}^{N} \mathbb{E}\left[(\lambda^2 + \mathbb{E}[S_N^2]) P[\tau \leq N]\right]
$$

$$= \left[(\lambda^2 + \mathbb{E}[S_N^2]) P[\tau \leq N]\right].$$

**Remark 23.60.** For those skipping the martingale section here is another way to estimate $\mathbb{E}\left[S_N^2 : \tau \leq N\right]$:

$$\mathbb{E}\left[S_N^2 : \tau \leq N\right] = \sum_{j=1}^{N} \mathbb{E}\left[S_j + S_N - S_j^2 : \tau = j\right]
$$

Since $\sum_{n=1}^{\infty} 2\mathbb{E} Y_j^2 < \infty$, it follows that $\lim_{n \to \infty} \mathbb{E} Y_j^2 = 0$. 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 23.42.

**Proof of Theorem 23.42.** 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. $\sum_{n=1}^{\infty} P(\{|X_n| > c\} < \infty$, and
2. $\sum_{n=1}^{\infty} \text{Var}(X_n 1_{|X_n| \leq c}) < \infty$, and
3. $\sum_{n=1}^{\infty} \text{E}(X_n 1_{|X_n| \leq c})$ 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 $c > 0$, $P(\{|X_n| \geq c \text{ i.o.}\}) = 0$. According to the Borel zero one law (Lemma 13.47) 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 $\{X_n^c := X_n 1_{|X_n| \leq c}\}$ 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 23.61 (with $Y_n = X_n^c$),

$$\sum_{n=1}^{\infty} \mathbb{E} X_n^c = \sum_{n=1}^{\infty} \mathbb{E}(X_n 1_{|X_n| \leq c})$$

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 23.58 allows us to conclude

$$\sum_{n=1}^{\infty} \text{Var}(X_n 1_{|X_n| \leq c}) = \sum_{n=1}^{\infty} \mathbb{E} Y_n^2 < \infty.$$
Weak Convergence Results

In this chapter we will discuss a couple of different ways to decide whether 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 \( \mu_n := \text{Law}(Y_n) \) close to \( \mu = \text{Law}(Y) \) for large \( n \).

We will often be the case that \( Y_n = X_1 + \cdots + X_n \) 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 \( \mathbb{R}^n \).

24.1 Convolutions

**Definition 24.1.** 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

\[
(\mu * \nu)(A) = P(X + Y \in A) = \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y) 1_A(x + y) = \int_{\mathbb{R}^n} \nu(A - x) d\mu(x) = \int_{\mathbb{R}^n} \mu(A - x) d\nu(x).
\]

This may also be expressed as,

\[
(\mu * \nu)(A) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x + y) d\mu(x) d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} 1_A(x + y) d(\mu \otimes \nu)(x, y).
\]

**Exercise 24.1.** 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 \( x \in \mathbb{R}^n \) 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 \( \mu_i = \text{Law}(Y_i) \), then

\[
\text{Law}(Y_1 + \cdots + Y_n) = \mu_1 \ast \mu_2 \ast \cdots \ast \mu_n.
\]

**Remark 24.2.** 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

\[
\mu * \nu(f) = \int_{\mathbb{R}^n} f(x + y) u(x) \, dx \, d\nu(y) = \int_{\mathbb{R}^n} f(x) u(x - y) \, d\nu(y)
\]

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 24.3.** 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

\[
d(\mu * \nu)(x) = (1_{[0,1]} * 1_{[-1,0]}) (x) \, dx
\]
where
\[
(1_{[0,1]} * 1_{[-1,0]})(x) = \int_{\mathbb{R}} 1_{[-1,0]}(y) 1_{[0,1]}(y) \, dy
= \int_{\mathbb{R}} 1_{[0,1]}(y-x) 1_{[0,1]}(y) \, dy
= \int_{\mathbb{R}} 1_{[0,1]+x}(y) 1_{[0,1]}(y) \, dy
= m([0,1] \cap (x + [0,1])) = (1 - |x|)_+.
\]

### 24.2 Total Variation Distance

**Definition 24.4.** Let \(\mu\) and \(\nu\) be two probability measures on a measurable space, \((\Omega, \mathcal{B})\). The total variation distance, \(d_{TV}(\mu, \nu)\), is defined as
\[
d_{TV}(\mu, \nu) := \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|,
\]i.e. \(d_{TV}(\mu, \nu)\) is simply the supremum norm of \(\mu - \nu\) as a function on \(\mathcal{B}\).

**Notation 24.5** 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 24.6.** 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 24.2.** 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 24.3.** Suppose that \(\mu, \nu,\) and \(\gamma\) are probability measures on \((\mathbb{R}^n, \mathcal{B}_\mathbb{R}^n)\). Show \(d_{TV}(\mu * \nu, \mu * \gamma) \leq d_{TV}(\nu, \gamma)\). Use this fact along with Exercise 24.2 to show,
\[
d_{TV}(\mu_1 * \mu_2 * \cdots * \mu_n, \nu_1 * \nu_2 * \cdots * \nu_n) \leq \sum_{i=1}^n d_{TV}(\mu_i, \nu_i)
\]for all choices probability measures, \(\mu_i\) and \(\nu_i\) on \((\mathbb{R}^n, \mathcal{B}_\mathbb{R}^n)\).

**Remark 24.7.** 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
\[
\|\lambda\|_{TV} := \sup \left\{ \sum_{i=1}^n |\lambda(A_i)| : n \in \mathbb{N} \text{ and partitions, } A_1, \ldots, A_n \in \mathcal{B} \text{ of } \Omega \right\}.
\]
You are asked to show in Exercise 24.4 below, that when \(\lambda = \mu - \nu, d_{TV}(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_{TV}\).

**Lemma 24.8 (Sheffe’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 \(dm = f dm\) and \(\nu = g dm\). Then
\[
d_{TV}(\mu, \nu) = \frac{1}{2} \int \! |f - g| \, dm.
\]
Let us now further suppose that \(\mu_n \to \mu, \nu_n \to \nu\) \(\mathcal{B}\)-a.e. with \(\nu = g dm\) 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 \(d\lambda = h dm\) and \(d\nu = g dm\) are probability measures of the form, \(d\mu_n = f_n dm\) and \(d\nu = g dm\) with \(g, f_n : \Omega \to [0, \infty)\). If \(f_n \to f, m\)-a.e. with \(\nu_n \to \nu\) \(\mathcal{B}\)-a.e. with \(\nu = g dm\) still being a probability measure, then \(d_{TV}(\mu_n, \nu) \to 0\) as \(n \to \infty\).

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,
\[
|\lambda(A)| = \frac{1}{2} \left( |\lambda(A)| + |\lambda(A^c)| \right) = \frac{1}{2} \left( \int_A h dm + \int_{A^c} h dm \right)
\] \(\leq \frac{1}{2} \left( \int_A |h| dm + \int_{A^c} |h| dm \right) = \frac{1}{2} \int \! |h| \, dm.
\]
This shows
\[
d_{TV}(\mu, \nu) = \sup_{A \in \mathcal{B}} |\lambda(A)| \leq \frac{1}{2} \int \! |h| \, dm.
\]To prove the converse inequality, simply take \(A = \{h > 0\}\) (note \(A^c = \{h \leq 0\}\)) in Eq. (24.7) to find

---

1 Fact: it is always possible to do this by taking \(m = \mu + \nu\) for example.
For the second assertion, observe that \(|f_n - g| \to 0\) \(m\) a.e., \(|f_n - g| \leq G_n := f_n + g \in L^1(\Omega)\), \(G_n \to G := 2g\) 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 10.27, \[\lim_{n \to \infty} d_{TV}(\mu_n, \nu) = \frac{1}{2} \lim_{n \to \infty} \int |f_n - g| \, dm = 0.\]

For a concrete application of Scheffé’s Lemma 24.8, see Proposition 24.53 below.

**Corollary 24.9.** 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

\[d_{TV}(\mu, \nu) = \frac{1}{2} \sup \left\{ \left| \int \Omega h \, d\mu - \int \Omega h \, d\nu \right| : \|h\|_\infty \leq 1 \right\}.\]  

Consequently, \[\left| \int \Omega h \, d\mu - \int \Omega h \, d\nu \right| \leq 2d_{TV}(\mu, \nu) \cdot \|h\|_\infty\] and in particular, for all bounded and measurable functions, \(h : \Omega \to \mathbb{R}\),

\[\lim_{n \to \infty} d_{TV}(\mu_n, \nu) = 0 \implies \lim_{n \to \infty} \int \Omega h \, d\mu_n = \int \Omega h \, d\nu.\]  

**Proof.** We begin by observing that

\[\left| \int \Omega h \, d\mu - \int \Omega h \, d\nu \right| = \left| \int \Omega h(f - g) \, d\mu \right| \leq \int \Omega |h| |f - g| \, d\mu \leq \|h\|_\infty \int \Omega |f - g| \, d\mu = 2d_{TV}(\mu, \nu) \cdot \|h\|_\infty.\]

Moreover, from the proof of Scheffé’s Lemma 24.8 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. (24.8) and (24.9) and the latter implies Eq. (24.10). 

**Exercise 24.4.** Under the hypothesis of Scheffé’s Lemma 24.8 show

\[\|\mu - \nu\|_{TV} = \int \Omega |f - g| \, dm = 2d_{TV}(\mu, \nu).\]  

**Exercise 24.5.** 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 24.6.** Let \(\mu_p(\{1\}) = p \quad \text{and} \quad \mu_p(\{0\}) = 1 - p\). Show \(\nu_\lambda(n) = e^{-\lambda} \lambda^n / n!\) for all \(n \in \mathbb{N}\).

1. Find \(d_{TV}(\mu_p, \mu_q)\) for all \(0 \leq p, q \leq 1\).
2. Show \(d_{TV}(\mu_p, \nu_p) = \mu(1 - e^{-p})\) for all \(0 \leq p \leq 1\). From this estimate and the estimate,

\[1 - e^{-p} = \int_0^p e^{-x} \, dx \leq \int_0^p 1 \, dx = p,\]  

it follows that \(d_{TV}(\mu_p, \nu_p) \leq p^2\) for all \(0 \leq p \leq 1\).
3. Show

\[d_{TV}(\nu_\lambda, \nu_\gamma) \leq |\lambda - \gamma| \quad \text{for all} \quad \lambda, \gamma \in \mathbb{R}_+.\]  

**Hints:** (Andy Parrish’s method – a former 280 student.)

a) Observe that for any \(n \in \mathbb{N}\) we have \(\nu_\lambda(n) = e^{-\lambda} \lambda^n / n!\) and use this to conclude

\[d_{TV}(\nu_\lambda, \nu_\gamma) \leq nd_{TV}(\nu_\lambda/n, \nu_\gamma/n).\]  

b) Using item 2. of this exercise, show

\[|d_{TV}(\nu_\lambda/n, \nu_\gamma/n) - d_{TV}(\mu_\lambda/n, \mu_\gamma/n)| \leq Cn^{-2}.\]

c) Finally make use of your results in item 1. part b. in order to let \(n \to \infty\) in Eq. (24.13).

The next theorem should be compared with Exercise 10.6 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 \(P(S = k) \approx P(\text{Poisson}(pn) = k)\) which is valid for \(k \ll n\).
Theorem 24.10 (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 \( P(Z_i = 0) = 1 - p_i \), \( S := Z_1 + \cdots + Z_n \), \( a := p_1 + \cdots + p_n \), and \( X \overset{d}{=} \text{Poi} (a) \). Then for any \( A \in \mathcal{B}_\mathbb{R} \) we have
\[
|P(S \in A) - P(X \in A)| \leq \sum_{i=1}^n p_i^2, \tag{24.14}
\]
or in short,
\[
d_{TV}\left( \sum_{i=1}^n Z_i, X \right) \leq \sum_{i=1}^n p_i^2.
\]
(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 \( X_i \overset{d}{=} \text{Poi} (p_i) \) for each \( i \). It then follows from Exercises 24.3 and 24.6 that,
\[
d_{TV}\left( \sum_{i=1}^n Z_i, \sum_{i=1}^n X_i \right) \leq \sum_{i=1}^n d_{TV}(Z_i, X_i) = \sum_{i=1}^n p_i \left( 1 - e^{-p_i} \right) \leq \sum_{i=1}^n p_i^2.
\]

The reader should compare the proof of this theorem with the proof of the central limit theorem in Theorem 13.42. For another less quantitative Poisson limit theorem, see Theorem 26.23.

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 \( \{A_n\}_{n=1}^{N(\varepsilon)} \subset \mathcal{M} \) of \( A \) such that \( \mu(A_n) \leq \varepsilon \) for all \( n \). (This assumption actually follows from assumption the no-atom assumption 1. above, see Lemma 15.55.)

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 \( \delta(x) \to 0 \) as \( x \to 0 \) and \( |F(A)| \leq \mu(A) \delta(\mu(A)) \) for all \( A \in \mathcal{M} \).

**Proposition 24.11 (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 \( \{A_i\}_{i=1}^n \subset \mathcal{M} \) 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.

\(^2\) Actually, since \( S \) and \( X \) are \( \mathbb{N}_0 \) - valued, we may as well assume that \( A \subset \mathbb{N}_0 \).

2. \( \mu(N(A) \geq 2) = o(\mu(A)) \).
3. \( |\mu(N(A) \geq 1) - \mu(A)| = o(\mu(A)) \).

Then \( N(A) \overset{d}{=} \text{Poi}(\mu(A)) \) 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 \( \{A_i\}_{i=1}^N \subset \mathcal{M} \) of \( A \) such that \( \mu(A_i) \leq \varepsilon \) for all \( i \). Let \( Z_i := 1_{N(A_i) \geq 1} \) and \( S := \sum_{i=1}^N Z_i \). Using
\[
N(A) = \sum_{i=1}^N N(A_i)
\]
and Lemma 24.13, we have
\[
|F(N(A) = k) - P(S = k)| \leq P(N(A) \neq S) \leq \sum_{i=1}^N P(Z_i \neq N(A_i)).
\]
Since \( \{Z_i \neq N(A_i)\} \cap \{N(A_i) \geq 2\} \cap \{N(A_i) \geq 2\} = o(\mu(A_i)) \), it follows that
\[
|F(N(A) = k) - P(S = k)| \leq \sum_{i=1}^N \mu(A_i) \delta(\mu(A_i)) \leq \sum_{i=1}^N \mu(A_i) \delta(\varepsilon) = \delta(\varepsilon) \mu(A). \tag{24.15}
\]

On the other hand, \( \{Z_i\}_{i=1}^N \) are independent Bernoulli random variables with \( P(Z_i = 1) = P(N(A_i) \geq 1) \)
and \( a_\varepsilon = \sum_{i=1}^N P(N(A_i) \geq 1) \). Then by the Law of rare events Theorem 24.10
\[
\left| P(S = k) - \frac{a_\varepsilon}{k!} e^{-a_\varepsilon} \right| \leq \sum_{i=1}^N \left[ P(N(A_i) \geq 1) \right]^2 \leq \sum_{i=1}^N \left[ \mu(A_i) + o(\mu(A_i)) \right]^2
\]
\[
\leq \sum_{i=1}^N \mu(A_i)^2 (1 + \delta'(\varepsilon))^2 = (1 + \delta'(\varepsilon))^2 \varepsilon \mu(A). \tag{24.16}
\]
Combining Eqs. (24.15) and (24.16) shows
\[
\left| P(N(A) = k) - \frac{a_\varepsilon}{k!} e^{-a_\varepsilon} \right| \leq \left[ \delta(\varepsilon) + (1 + \delta'(\varepsilon))^2 \varepsilon \right] \mu(A). \tag{24.17}
\]
where \( a_\varepsilon \) satisfies
\[
|a_\varepsilon - \mu(A)| = \sum_{i=1}^{N} |P(N(A_i^\varepsilon) \geq 1) - \mu(A_i^\varepsilon)|
\leq \sum_{i=1}^{N} |P(N(A_i^\varepsilon) \geq 1) - \mu(A_i^\varepsilon)| \leq \sum_{i=1}^{N} o(\mu(A_i^\varepsilon))
\leq \sum_{i=1}^{N} \mu(A_i^\varepsilon) \delta'(\mu(A_i^\varepsilon)) \leq \mu(A) \delta'(\varepsilon).
\]

Hence we may let \( \varepsilon \downarrow 0 \) in Eq. (24.17) 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 14.12 and Exercises 14.6, 14.8 for concrete constructions of Poisson processes.

**Remark 24.12.** 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,
\[
N(A) \overset{d}{=} \text{Bern}(\mu(A)) + o(\mu(A)). \tag{24.18}
\]

This last statement is short hand for the assumptions that
\[
P(N(A) = 0) = 1 - \mu(A) + o(\mu(A)) \quad \text{and} \quad P(N(A) = 1) = \mu(A) + o(\mu(A)),
\]
and as a consequence,
\[
P(N(A) \geq 2) = 1 - P(N(A) = 0) + P(N(A) = 1) = o(\mu(A)).
\]

Assumption (24.18) implies
\[
E[z^{N(A)}] = (1 - \mu(A) + o(\mu(A))) + (\mu(A) + o(\mu(A))) z + o(\mu(A)) = 1 + \mu(A)(z - 1) + o(\mu(A)) = \exp(1 + \mu(A)(z - 1) + o(\mu(A))). \tag{24.19}
\]

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. (24.19) and the independence of the \( \{N(A_i)\}_{i=1}^n \) that
\[
E[z^{N(A)}] = \prod_{i=1}^n E[z^{N(A_i)}] = \prod_{i=1}^n \exp(1 + \mu/n(z - 1) + o(\mu/n)) = \exp(\mu(z - 1) + o(1/n)).
\]

Letting \( n \to \infty \) in this equation shows \( E[z^{N(A)}] = \exp(\mu(z - 1)) \) from which it follows that that \( N(A) \overset{d}{=} \text{Poi}(\mu(A)) \).

### 24.3 A Coupling Estimate

**Lemma 24.13 (Coupling Estimates).** Suppose \( X \) and \( Y \) are any random variables on a probability space, \((\Omega, \mathcal{B}, P)\) and \( A \in \mathcal{B}_R \). Then
\[
|P(X \in A) - P(Y \in A)| \leq P(\{X \in A\} \cup \{Y \in A\}) \leq P(X \neq Y). \tag{24.20}
\]

**Proof.** The proof is simply;
\[
|P(X \in A) - P(Y \in A)| = |E[1_A(X) - 1_A(Y)]| \leq E[1_A(X) - 1_A(Y)] = E[1_{X \in A \cup Y \in A}] \leq E[1_{X \neq Y}] = P(X \neq Y).
\]

Pushing the above proof a little more we have, if \( \{A_i\} \) is a partition of \( \Omega \), then
\[
\sum_i |P(X \in A_i) - P(Y \in A_i)| = \sum_i |E[1_{A_i}(X) - 1_{A_i}(Y)]| \leq \sum_i E[1_{A_i}(X) - 1_{A_i}(Y)] \leq E[1_{X \neq Y} : X \in A_i \text{ or } Y \in A_i] \leq E[1_{X \neq Y} : X \in A_i] + E[1_{X \neq Y} : Y \in A_i] = 2P(X \neq Y).
\]

This shows
\[
\|X_*P - Y_*P\|_{TV} \leq 2P(X \neq Y).
\]

This is really not more general than Eq. (24.20) since the Hahn decomposition theorem we know that in fact the signed measure, \( \mu := X_*P - Y_*P \), has total variation given by
\[
\|\mu\|_{TV} = \mu(\Omega^+) - \mu(\Omega^-)
\]
and observe that

$$\sum_{x \in \Omega} |J_x| + \sum_{x \in \Omega^+} [\mu_x (x) - \nu_x (x)] = \sum_{x \in \Omega} [\mu_x (x) \wedge \nu_x (x)] + \sum_{x \in \Omega^+} [\mu_x (x) - \nu_x (x)]$$

and similarly,

$$\sum_{x \in \Omega} |J_x| + \sum_{x \in \Omega^+} [\nu_x (x) - \mu_x (x)] = \sum_{x \in \Omega} \nu_x (x) = 1.$$  

Assuming we have lined up the \{J_x\} in \(0, \sum_{x \in \Omega} |J_x|\) we may choose two partitions \(\{K_x\}_{x \in \Omega^+}\) and \(\{K_x\}_{x \in \Omega^-}\) of \(\{\sum_{x \in \Omega} |J_x|, 1\}\) such that \(|K_x| = |\mu_x (x) - \nu_x (x)|\) for \(x \in \Omega^+ \cup \Omega^-\). If we then let

\[ Y := \sum_{x \in \Omega^+} x1_{J_x \in K_x} + \sum_{x \in \Omega^-} x1_{J_x \in K_x} \quad \text{and} \quad Z := \sum_{x \in \Omega^+} x1_{J_x \in K_x} + \sum_{x \in \Omega^-} x1_{J_x \in K_x} \]

then \(P(Y = s) = \mu(s)\) and \(P(Z = s) = \nu(s)\) for all \(s \in \Omega\). Moreover (as you should verify),

\[ \{Y \neq Z\} = \bigcup_{x \in \Omega^+} K_x = \bigcup_{x \in \Omega^-} K_x \]

so that

\[ P(Y \neq Z) = \sum_{x \in \Omega^+} |K_x| = \frac{1}{2} \sum_{x \in \Omega} |K_x| = \frac{1}{2} \sum_{x \in \Omega} |\mu_x (x) - \nu_x (x)|. \]

Taking \(\rho := \text{Law}(Y, Z)\) as a measure on \(\Omega \times \Omega\) then completes the proof. 

\[ \Box \]

**Proof.** Suppose that \(S\) is a finite (or countable) set and \(\mu\) and \(\nu\) are two probabilities on \(S\). Then there exists a coupling \(\rho\) on \(\Omega \times \Omega\) such that

\[ \|\mu - \nu\|_{TV} = \frac{2}{\sqrt{2}} \left( S^2 \setminus \Delta \right) . \]

**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

\[ \sum_{x \in S} |J_x| + \sum_{x \in S_+} [\mu(x) - \nu(x)] = \sum_{x \in S} \mu(x) \wedge \nu(x) + \sum_{x \in S_+} [\mu(x) - \nu(x)] \]

\[ = \sum_{x \in S} \mu(x) = 1 \]

and similarly,

\[ \sum_{x \in S} |J_x| + \sum_{x \in S_-} [\nu(x) - \mu(x)] = \sum_{x \in S} \nu(x) = 1. \]
\[ \alpha_i (k) := P (\text{Poi} (p_i) \leq k) = e^{-p_i} \sum_{j=0}^{k} \frac{p_i^j}{j!} \]

with the convention that \( \alpha_i (-1) = 0 \). Notice that
\[ e^{-p_i} \leq \alpha_i (k) \leq \alpha_i (k + 1) \leq 1 \text{ for all } k \in \mathbb{N}_0 \]
and for \( p_i \) small, we have
\[ \alpha_i (0) = e^{-p_i} \approx 1 - p_i \]
\[ \alpha_i (1) = e^{-p_i} (1 + p_i) \approx 1 - p_i^2. \]
If we define (see Figure 24.1) by
\[ X_i := \sum_{k=0}^{\infty} k 1_{\alpha_i (k-1) < U_i \leq \alpha_i (k)}, \]
then \( X_i = \text{Poi} (p_i) \) since
\[ P (X_i = k) = P (\alpha_i (k-1) < U_i \leq \alpha_i (k)) = \alpha_i (k) - \alpha_i (k-1) = e^{-p_i} \frac{p_i^k}{k!}. \]

An application of Lemma 24.13 now shows
\[ |P (S \in A) - P (X \in A)| \leq P (S \neq X) \]
and since \( \{S \neq X\} \subset \cup_{i=1}^{n} \{X_i \neq Z_i\} \), we may conclude
\[ |P (S \in A) - P (X \in A)| \leq \sum_{i=1}^{n} P (X_i \neq Z_i). \]
As is easily seen from Figure 24.1
\[ P (X_i \neq Z_i) = [\alpha_i (0) - (1 - p_i)] + 1 - \alpha_i (1) = [e^{-p_i} - (1 - p_i)] + 1 - e^{-p_i} (1 + p_i) = p_i (1 - e^{-p_i}) \leq p_i^2 \]
where we have used the estimate in Eq. (24.11) for the last inequality. \( \blacksquare \)

### 24.4 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 \( -\infty < a \leq b < \infty \). 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 \( -\infty < a \leq b < \infty \).

Example 24.15. Suppose that \( P (X_n = \frac{i}{n}) = \frac{1}{n} \) for \( i \in \{1, 2, \ldots, n\} \) 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 \( A_n = \{ \frac{i}{n} : i = 1, 2, \ldots, n\} \), 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 \( n \).

Nevertheless we would like \( X_n \) to be close to \( U \) in distribution. Let us observe that if we let \( F_n (y) := P (X_n \leq y) \) and \( F (y) := P (U \leq y) \), then
\[ F_n (y) = P (X_n \leq y) = \frac{1}{n} \# \left\{ i \in \{1, 2, \ldots, n\} : \frac{i}{n} \leq y \right\} \]
and
\[ F (y) := P (U \leq y) = (y \wedge 1) \lor 0. \]
From these formula, it easily follows that \( F (y) = \lim_{n \to \infty} F_n (y) \) for all \( y \in \mathbb{R} \), see Figure 24.2. This suggest that we should say that \( X_n \) converges in distribution to \( X \) iff \( P (X_n \leq y) \to P (X \leq y) \) for all \( y \in \mathbb{R} \). However, the next simple example shows this definition is also too restrictive.

More generally, if \( \mu \) and \( \nu \) are two probability measures 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 \( d_{TV} (\mu, \nu) = 1 \).
2. \( F_n \) converges to \( F \) \textbf{weakly} and write, \( F_n \overset{w}{\to} F \), iff \( F_n(x) \to F(x) \) for all \( x \in C(F) \).

3. We say \( F \) is \textbf{proper}, if \( F \) is a distribution function of a probability measure, i.e. if \( F(\infty) = 1 \) and \( F(-\infty) = 0 \).

Example 24.20. If \( X_n \) and \( U \) are as in Example\(^{[24.15]} \) and \( F_n(y) := P(X_n \leq y) \) and \( F(y) := P(Y \leq y) \), then \( F_n \overset{w}{\to} F \) and \( F_n \overset{w}{\to} F \).

Example 24.21. Suppose that \( Z \) is a random variable and \( F(x) := P(Z \leq x) \). Let \( X_n = n + Z \) and

\[
F_n(x) := P(X_n \leq x) = P(n + Z \leq x) = P(Z \leq x - n) = F(x - n) \to 0 \text{ as } n \to \infty.
\]

Thus it follows that \( F_n \overset{w}{\to} 0 = F_\infty \). Notice that limit, \( F_\infty \), is no longer a distribution function, i.e. \( F_\infty \) is not proper.

Lemma 24.22. Let \( \{F, F_n : n = 1, 2, \ldots \} \) be a collection of \textbf{proper} distribution functions. Then \( F_n \overset{w}{\to} F \) iff \( F_n \overset{w}{\to} F \).

**Proof.** If \( F_n \overset{w}{\to} F \), then \( F_n([a, b]) = F_n(b) - F_n(a) \to F(b) - F(a) = F([a, b]) \) for all \( a, b \in C(F) \) and therefore \( F_n \overset{w}{\to} F \). So now suppose \( F_n \overset{w}{\to} F \) and let \( a < x \) with \( a, x \in C(F) \). 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,

\[
F(x) \geq \limsup_{n \to \infty} F_n(x).
\]

\[\blacksquare\]

**Definition 24.23.** In the case where \( F_n \) and \( F \) are proper and \( F_n \overset{w}{\to} F \), we will write \( F_n \underset{w}{\Rightarrow} F \). Moreover if \( \{X_n\} \) is a sequence of random variables, we say \( X_n \) \textbf{converges weakly} or \textbf{converges in distribution} to \( X_0 \) x (iff \( F_n(y) := P(X_n \leq y) \Rightarrow F(y) := P(X \leq y) \) and we abbreviate this convergence by writing \( X_n \Rightarrow X \).
Example 24.24 (Central Limit Theorem). The central limit theorem (see Theorems 10.74, Corollary 13.44 and Theorem 25.29) states; if \( \{X_n\}_{n=1}^{\infty} \) are i.i.d. \( L^2 (P) \) random variables with \( \mu := \mathbb{E} X_1 \) and \( \sigma^2 = \text{Var} (X_1) \), then
\[
\frac{S_n - n \mu}{\sqrt{n}} \xrightarrow{d} N (0, \sigma).
\]
Written out explicitly we find
\[
\lim_{n \to \infty} P \left( a < \frac{S_n - n \mu}{\sigma \sqrt{n}} \leq b \right) = P (a < N (0, 1) \leq b) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{- \frac{1}{2} x^2} \, dx
\]
or equivalently put
\[
\lim_{n \to \infty} P \left( (n \mu + \sqrt{n}a < S_n \leq n \mu + \sqrt{nb}) \right) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{- \frac{1}{2} x^2} \, dx.
\]
More intuitively, we have
\[
S_n \xrightarrow{d} n \mu + \sqrt{n} \sigma N (0, 1) \xrightarrow{d} N (n \mu, n \sigma^2).
\]

Example 24.25. Suppose that \( P (X_n = n) = 1 \) for all \( n \), then \( F_n (y) = 1_{y \geq n} \to 0 \) 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 \( F_n = \frac{1}{2} 1_{[\pm n, \infty)} + \frac{1}{2} = F (y) \) 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 24.26. Suppose \( X \) is a non-zero random variables such that \( X \xrightarrow{d} -X \), then \( X_n := (-1)^n X \xrightarrow{d} X \) for all \( n \) and therefore, \( X_n \xrightarrow{d} X \) as \( n \to \infty \). On the other hand, \( X_n \) does not converge to \( X \) almost surely or in probability.

Lemma 24.28. Suppose \( \{X_n\}_{n=1}^{\infty} \) is a sequence of random variables on a common probability space and \( c \in \mathbb{R} \). Then \( X_n \xrightarrow{d} c \) iff \( X_n \xrightarrow{P} c \).

Proof. Recall that \( X_n \xrightarrow{P} c \) for all \( \epsilon > 0 \), then \( P (|X_n - c| > \epsilon) \to 0 \). Since
\[
\{X_n - c| \geq \epsilon\} \to \{X_n > c + \epsilon\} \cup \{X_n < c - \epsilon\}
\]
it follows \( X_n \xrightarrow{P} c \) as \( P (X_n > x) \to 0 \) for all \( x > c \) and \( P (X_n < x) \to 0 \) for all \( x < c \). These conditions are also equivalent to \( P (X_n \leq x) \to 1 \) for all \( x > c \) and \( P (X_n \leq x) \leq P (X_n < x') \to 0 \) for all \( x < c \) (where \( x < x' < c \)). So \( X_n \xrightarrow{P} c \) iff \( X_n \xrightarrow{d} c \).

Alternative proof using Theorem 24.32 below. For the implication that \( X_n \xrightarrow{P} c \) implies \( X_n \xrightarrow{d} c \), see Corollary 24.35 Conversely if \( X_n \xrightarrow{d} c \) and \( \epsilon > 0 \) let \( f_\epsilon \in C ([\mathbb{R}, 0, 1]) \) such that \( f_\epsilon (c) = 0 \) and \( f_\epsilon (x) = 1 \) for \( |x - c| \geq \epsilon \). (A simple piecewise linear function will do here.) Then
\[
P (|X_n - c| \geq \epsilon) \leq \mathbb{E} |f_\epsilon (X_n)| \to \mathbb{E} |f (c)| = 0 \text{ as } n \to \infty.
\]

Notation 24.29. Given a proper distribution function, \( F : \mathbb{R} \to [0, 1] \), let \( Y = F^* : (0, 1) \to \mathbb{R} \) be the function defined by
\[
Y (x) = F^* (x) := \sup \{y \in \mathbb{R} : F (y) < x\}.
\]
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 24.3
1. \( Y (x) \leq Y^* (x) \) and \( Y (x) < Y^* (x) \) if \( x \) is the height of a “flat spot” of \( F \).
2. The set
\[
E := \{x \in (0, 1) : Y (x) < Y^* (x)\}
\]
of flat spot heights is at most countable. This is because, \( \{(Y (x), Y^* (x))\}_{x \in E} \) is a collection of pairwise disjoint intervals which is necessarily countable. (Each such interval contains a rational number.)
Theorem 24.30 (Baby Skorohod Theorem). Suppose \( \{F_n\}_{n=0}^\infty \) is a collection of distribution functions such that \( F_n \Rightarrow F_0 \). Then there exists a probability space, \( (\Omega, B, P) \) and random variables, \( \{Y_n\}_{n=1}^\infty \) such that \( P(Y_n \leq y) = F_n(y) \) for all \( y \in \mathbb{R} \) and \( \lim_{n \to \infty} Y_n = Y \) a.s.

Proof. We will take \( \Omega := (0, 1), B = B_{(0, 1)}, \) and \( P = m - \text{Lebesgue measure} \) on \( \Omega \) and let \( Y_n := F_n^- \) and \( Y := F_0^- \) as in Notation 24.29. 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. 24.21.

We now suppose \( x \notin E \). If \( y \in \mathbb{C}(F_0) \) 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 \( y \in \mathbb{C}(F_0) \) then implies

\[
\liminf_{n \to \infty} Y_n(x) \geq Y(x).
\]

Similarly, for \( x \notin E \) and \( y \in \mathbb{C}(F_0) \) 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 \( y \in \mathbb{C}(F_0) \) 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

\[
\lim_{n \to \infty} F_n^+(x) = \lim_{n \to \infty} Y_n(x) = Y(x) = F^+(x) \quad \text{for all } x \notin E. \tag{24.23}
\]

In preparation for the full version of Skorohod’s Theorem 24.86, it will be useful to record a special case of Theorem 24.30 which has both a stronger hypothesis and a stronger conclusion.

Theorem 24.31 (Prenatal Skorohod Theorem). Suppose \( S = \{1, 2, \ldots, m\} \subset \mathbb{R} \) and \( \{\mu_n\}_{n=1}^\infty \) is a sequence of probabilities on \( S \) such that \( \mu_n \Rightarrow \mu \) for some probability \( \mu \) on \( S \). Let \( P := \mu \otimes m \) on \( \Omega := S \times (0, 1], \) \( Y(x, \theta) = i \) for all \( (i, \theta) \in \Omega. \) Then there exists \( Y_n : \Omega \to S \) such that \( \text{Law}_P(Y_n) = \mu_n \) 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 \( \text{Law}_P(Y_\nu) = \nu. \) 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 \( a_i = \left( \sum_{j \leq i} \alpha_j \right) a \) for \( 1 \leq i \leq m. \)

Let \( \nu \) be any probability on \( S \) and let

\[
A_i := \{i\} \times \left( 0, \frac{\nu(i)}{\mu(i)} \right] \land 1
\]

and

\[
\lim_{n \to \infty} Y_n(x) \geq Y(x).
\]
\[ C = \Omega \setminus \left( \sum_{i=1}^{m} A_i \right) = \sum_{i=1}^{m} \{ i \} \times \left( \frac{\nu(i)}{\mu(i)} \wedge 1, 1 \right) \]

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
\[ Y_{\nu}(i, \theta) := \sum_{i=1}^{m} iI_{A_i \cup C_i} \]

so that \( Y_{\nu} = i \) on \( A_i \) and in particular \( Y_{\nu}(i, \theta) = i \) when \( \theta \leq \nu(i) / \mu(i) \).

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,
\[ \nu(i) = P(Y_{\nu} = i) = P(A_i \cup C_i) = P(A_i) + \alpha_i P(C) = \nu(i) \wedge \mu(i) + \alpha_i P(C) \]

and therefore we must define
\[ \alpha_i = (\nu(i) - \nu(i) \wedge \mu(i)) / P(C) \geq 0. \]

To see this is an admissible choice (i.e. \( \sum_{i=1}^{m} \alpha_i = 1 \)) notice that
\[ P(C) = \sum_{i} [\mu(i) - \nu(i) \wedge \mu(i)] = \sum_{\nu(i) < \mu(i)} (\mu(i) - \nu(i)) = \sum_{\mu(i) \leq \nu(i)} (\nu(i) - \mu(i)), \]

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,
\[ \sum_{i \in S} \alpha_i = \frac{1}{P(C)} \sum_{\mu(i) \leq \nu(i)} (\nu(i) - \mu(i)) = 1. \]

The next theorem summarizes a number of useful equivalent characterizations of weak convergence. (The reader should compare Theorem 24.32 with Corollary 24.9.) 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 \( C(\mathbb{R}) \) for those \( f \in C(\mathbb{R}) \) which have compact support, i.e. \( f(x) \equiv 0 \) if \( |x| \) is sufficiently large.

**Theorem 24.32.** Suppose that \( \{ \mu_n \}_{n=0}^{\infty} \) is a sequence of probability measures on \( (\mathbb{R}, \mathcal{B}) \) 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 \( f \in BC(\mathbb{R}) \),
   \[ \int_{\mathbb{R}} f d\mu_n \to \int_{\mathbb{R}} f d\mu_0 \text{ as } n \to \infty. \]  
   (24.26)

2. Eq. (24.20) holds for all \( f \in BC(\mathbb{R}) \) which are uniformly continuous.

3. Eq. (24.26) holds for all \( f \in C(\mathbb{R}) \).

4. \( F_n \Rightarrow F \).

5. There exists a probability space \( (\Omega, \mathcal{B}, P) \) and random variables, \( Y_n \), on this space such that \( \mathbb{P} \circ Y_n^{-1} = \mu_n \) 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
\[ \int_{\mathbb{R}} f d\mu_n = \mathbb{E} [f(Y_n)]^{D.C.T.} \mathbb{E} [f(Y)] = \int_{\mathbb{R}} f d\mu_0 \]

for all \( f \in BC(\mathbb{R}) \). The implication that 4. \( \implies \) 5. is Skorohod’s Theorem 24.30 above. Therefore it suffices to prove 3. \( \implies \) 4.

(3. \( \implies \) 4.) Let \( -\infty < a < b < \infty \) with \( a, b \in C(F_0) \) and for \( \varepsilon > 0 \), let \( f_\varepsilon(x) \geq 1_{(a,b]} \) and \( g_\varepsilon(x) \leq 1_{(a,b]} \) be the functions in \( C_\varepsilon(\mathbb{R}) \) pictured in Figure 24.4. Then
\[ \limsup_{n \to \infty} \mu_n ((a, b]) \leq \limsup_{n \to \infty} \int_{\mathbb{R}} f_\varepsilon d\mu_n = \int_{\mathbb{R}} f_\varepsilon d\mu_0 \]

(24.27) and
\[ \liminf_{n \to \infty} \mu_n ((a, b]) \geq \liminf_{n \to \infty} \int_{\mathbb{R}} g_\varepsilon d\mu_n = \int_{\mathbb{R}} g_\varepsilon d\mu_0. \]  

(24.28) 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. (24.27) and (24.28) to conclude,
points of continuity of \( \mu \) and is equal to \( a \), \( b \) \( \in \mathbb{R} \) \( \lim \sup \mu_n ((a, b]) = \mu_0 ((a, b]) \)

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 \( \mu_0 ((a, b]) \).

*Proposition 24.33.* Suppose that \( \{ \mu_n \}_{n=1}^\infty \) are measures on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) such that \( \mu_n (\mathbb{Z}) = 1 \) for all \( n \). Then \( \mu_n \implies \mu \) if and only if \( \lim_{n \to \infty} d_{TV} (\mu_n, \mu) = 0 \).

**Proof.** In light of Example 24.33 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 \( |x - m| \geq 1 \). 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 24.5.

*Corollary 24.35.* 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 24.26 shows the converse is in general false.)

**Proof.** Let \( g \in B C (\mathbb{R}) \), then by Corollary 15.12, \( g (X_n) \xrightarrow{P} g (X_0) \) and since \( g \) is bounded, we may apply the dominated convergence theorem (see Corollary 15.9) to conclude that \( E[g (X_n)] \to E[g (X_0)] \).

We end this section with a few more equivalent characterizations of weak convergence. The combination of Theorem 24.32 and 24.36 is often called the Portmanteau Theorem. A review of the notions of closure, interior, and boundary of a set \( A \) which are used in the next theorem may be found in Subsection 24.10.1 below.

**Theorem 24.36 (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. \( \lim_{n \to \infty} F_n (U) \geq F_0 (U) \) for open subsets, \( U \subset \mathbb{R} \).
3. \( \lim_{n \to \infty} F_n (C) \leq F_0 (C) \) 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 \( \mathbb{F}_0 (bd (A)) = 0 \).

\(^5\) Portmanteau: 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, we may choose random variables, \(Y_n\), such that \(P(Y_n \leq y) = F_n(y)\) 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
\[
1_U(Y) \leq \liminf_{n \to \infty} 1_U(Y_n) \text{ a.s.}
\]
and so by Fatou’s lemma,
\[
F(U) = P(Y \in U) = \mathbb{E}[1_U(Y)] \\
\leq \liminf_{n \to \infty} \mathbb{E}[1_U(Y_n)] = \liminf_{n \to \infty} P(Y_n \in U) = \liminf_{n \to \infty} F_n(U).
\]

(2. \(\iff\) 3.) This follows from the observations: 1) \(C \subseteq \mathbb{R}\) is closed iff \(U := C^c\) is open, 2) \(F(U) = 1 - F(C)\), and 3) \(\lim \inf_{n \to \infty} (-F_n(C)) = -\lim \sup_{n \to \infty} F_n(C)\).

(2. and 3. \(\iff\) 4.) If \(F_0(\partial(A)) = 0\), then \(A^c \subseteq A \subseteq \bar{A}\) with \(F_0(\bar{A} \setminus A^c) = F_0(\partial(A)) = 0\). Therefore
\[
F_0(A) = F_0(A^c) \leq \liminf_{n \to \infty} F_n(A^c) \leq \limsup_{n \to \infty} F_n(A) \leq F_0(\bar{A}) = F_0(A).
\]

(4. \(\implies\) 1.) Let \(a, b \in C(F_0)\) and take \(A := (a, b)\). Then \(F_0(\partial(A)) = F_0((a, b)) = 0\) and therefore, \(\lim_{n \to \infty} F_n((a, b)) = F_0((a, b))\), i.e. \(F_n \Rightarrow F_0\)\.

**Exercise 24.7.** 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
\[
\sup_{a < b} |\mu_F((a, b]) - \mu_{F_n}((a, b])|
\leq \sup_{a < b} |F(b) - F(a) - (F_n(b) - F_n(a))|
\leq \sup_{a < b} |F(b) - F_n(b)| + \sup_{a} |F_n(a) - F_n(a)| \to 0 \text{ as } n \to \infty.
\]

**Hints for part 2.** Given \(\varepsilon > 0\), show that there exist \(a_0 < \alpha_1 < \cdots < \alpha_n = \infty\), such that \(|F(\alpha_{i+1}) - F(\alpha_i)|\leq \varepsilon\) for all \(i\). Now, show, for \(x \in [\alpha_i, \alpha_{i+1})\), that
\[
|F(x) - F_n(x)|
\leq (F(\alpha_{i+1}) - F(\alpha_i)) + |F(\alpha_i) - F_n(\alpha_i)| + (F_n(\alpha_{i+1}) - F_n(\alpha_i)).
\]

Most of the results above generalize to the case where \(\mathbb{R}\) is replaced by a complete separable metric space as described in Section 24.10 below. The definition of weak convergence in this generality is as follows.

**Definition 24.37 (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 \(f \in BC(S)\)\.

We will write this convergence as \(\mu_n \Rightarrow \mu\) or \(\mu_n \rightharpoonup \mu\) 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 24.38.** 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 \(0 \leq \rho \leq 1_{C_1}\) and \(\int_S \rho(z) dz = 1\). For \(f \in C_c(S)\) and \(\varepsilon > 0\), let
\[
f_\varepsilon(x) := \int_S f(x + \varepsilon z) \rho(z) dz.
\]
It then follows that
\[
M_\varepsilon := \max_x |f(x) - f_\varepsilon(x)| = \max_x \left| \int_S (f(x) - f(x + \varepsilon z)) \rho(z) dz \right|
\leq \max_x \int_S |f(x) - f(x + \varepsilon z)| \rho(z) dz
\leq \max \max_x |f(x) - f(x + z)|
\]
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. 24.29 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,
\[
\lim_{n \to \infty} \sup_{x \in S} |\mu(f) - \mu_n(f)|
\leq \lim \sup_{n \to \infty} \left| \mu(f) - \mu_n(f) \right|
\leq 2M_\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0.
\]

\(\Box\)

\(\text{This is actually “weak-* convergence” when viewing } \mu_n \in BC(S)\).
Theorem 24.39. Suppose that \( \{ \mu_n \}_{n=1}^{\infty} \cup \{ \mu \} \) are probability measures on \( (S := \mathbb{R}^d, \mathcal{B} = \mathcal{B}_{\mathbb{R}^d}) \) (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 \( \mu(K_\varepsilon) \geq 1 - \varepsilon \) and \( \mu_n(K_\varepsilon) \geq 1 - \varepsilon \) 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 \( C_R := \{ x \in S : |x| \leq R \} \) and then choose \( \varphi_R \in C_c(S) \) such that \( \varphi_R = 1 \) on \( C_{R/2} \) and \( 0 \leq \varphi_R \leq 1_{C_R} \).

1. With this notation it follows that

\[
\mu_n (C_R) \geq \mu_n (\varphi_R) \rightarrow \mu (\varphi_R) \geq \mu (C_{R/2}).
\]

Choose \( R \) so large that \( \mu_n (C_R) \geq \mu (C_{R/2}) \geq 1 - \varepsilon / 2 \). Then for \( n \geq N_\varepsilon \) we will have \( \mu_n (C_R) \geq 1 - \varepsilon \) for all \( n \geq N_\varepsilon \). By increasing \( R \) more if necessary we may also assume that \( \mu_n (C_R) \geq 1 - \varepsilon \) 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

\[
\limsup_{n \to \infty} |\mu (f) - \mu (f_R)| \leq \limsup_{n \to \infty} \left[ |\mu (f) - \mu (f_R)| + |\mu (f) - \mu_n (f_R) - \mu_n (f)| \right] \\
= |\mu (f) - \mu (f_R)| + \sup_{n \to \infty} |\mu_n (f_R) - \mu_n (f)|.
\] (24.30)

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 have

\[
\sup_n |\mu_n (f_R) - \mu_n (f)| \leq \sup_n |f_R - f| \leq M \cdot \sup_n |\mu_n (\varphi_R - 1)| \leq M \cdot \sup_n |\mu_n (S \setminus C_{R/2})|.
\]

However, by item 1, it follows that \( \lim_{R \to \infty} \sup_n \mu_n (S \setminus C_{R/2}) = 0 \). Therefore letting \( R \to \infty \) in Eq. (24.30) show that \( \limsup_{n \to \infty} |\mu (f) - \mu_n (f)| = 0 \).

\[
24.5 \text{ “Derived” Weak Convergence}
\]

Lemma 24.40. 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. \( \{ f^\delta \leq a \} \) is closed (or equivalently \( \{ f^\delta > a \} \) 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 in \( B_x (\delta) \) whenever \( d(x,x') < \delta - d(x,y) \) (because then \( d(x',y) \leq d(x,y) + d(x,x') < \delta \)) it follows that \( f^\delta (x') > a \) for all \( x' \in B_x (\delta - d(x,y)) \). This shows \( \{ f^\delta > a \} \) 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

\[
\{ f_\delta \geq a \} = \{ -(f^\delta) \leq -a \}
\]

is closed for all \( a \in \mathbb{R} \), i.e. \( f_\delta \) is upper semi-continuous. Moreover, \( f_\delta \leq f \leq f^\delta \) for all \( \delta > 0 \) and \( f^\delta \downarrow f^0 \) and \( f_\delta \uparrow f_0 \) as \( \delta \downarrow 0 \), where \( f_0 = f \) and \( f^0 = \sup f \). The proofs are now complete since it is easy to see that

\[
\mathcal{D}(f) = \{ f^0 > f_0 \} = \{ f^0 - f_0 \neq 0 \} \in \mathcal{B}_X.
\]

Remark 24.41. Suppose that \( x_n \to x \) with \( X := \mathcal{D}(f)^c \). Then \( f(x_n) \to f(x) \) as \( n \to \infty \).

Theorem 24.42 (Continuous Mapping Theorem). Let \( f : \mathbb{R} \to \mathbb{R} \) be a Borel measurable function. If \( X_n \xrightarrow{\text{prob}} X_0 \) and \( P(X_0 \in \mathcal{D}(f)) = 0 \), then \( f(X_n) \xrightarrow{\text{prob}} f(X_0) \). If in addition, \( f \) is bounded, \( \lim_{n \to \infty} E f(X_n) = E f(X_0) \). (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 [24.30]. For \( g \in BC(\mathbb{R}) \) we observe that \( \mathcal{D}(g \circ f) \subset \mathcal{D}(f) \) and therefore,

\[
P(Y_0 \in \mathcal{D}(g \circ f)) \leq P(Y_0 \in \mathcal{D}(f)) = P(X_0 \in \mathcal{D}(f)) = 0.
\]

Hence it follows that \( g \circ f \circ Y_n \xrightarrow{\text{prob}} g \circ f \circ Y_0 \) a.s. So an application of the dominated convergence theorem (see Corollary [15.9]) implies
\[ \mathbb{E} [g (f (X_n))] = \mathbb{E} [g (f (Y_n))] \to \mathbb{E} [g (f (Y_0))] = \mathbb{E} [g (f (X_0))]. \] (24.31)

This proves the first assertion. For the second assertion we take \( g (x) = (x \land M) \lor (-M) \) in Eq. (24.31) where \( M \) is a bound on \( |f| \).

\textbf{Theorem 24.43 (Slutzky’s Theorem).} Suppose that \( X_n \Rightarrow X \in \mathbb{R}^m \) and \( Y_n \overset{d}{\to} c \in \mathbb{R}^n \) where \( c \in \mathbb{R}^n \) is constant. Assuming all random vectors are on the same probability space we will have \( (X_n, Y_n) \Rightarrow (X, c) \) – see Definition 24.37.

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 \( X_n + Y_n \Rightarrow X + c \) and \( X_n \cdot Y_n \Rightarrow X \cdot c \) respectively. (The first part of this theorem generalizes to metric spaces as well.)

\textbf{Proof.} According to Theorem 24.39 it suffices to show for
\[ \lim_{n \to \infty} \mathbb{E} [f (X_n, Y_n)] = \mathbb{E} [f (X, c)] \] (24.32)

for all \( f \in BC (\mathbb{R}^{m \times n}) \) which are uniformly continuous or even only \( f \in C_c (\mathbb{R}^{m \times n}) \). For a uniformly continuous function we have for every \( \varepsilon > 0 \) a \( \delta = \delta (\varepsilon) > 0 \) such that
\[ |f (x, y) - f (x', y')| \leq \varepsilon \text{ if } \|(x, y) - (x', y')\| \leq \delta. \]

Then
\[ |\mathbb{E} [f (X_n, Y_n) - f (X_n, c)]| \leq \mathbb{E} [\|f (X_n, Y_n) - f (X_n, c)\| : \|Y_n - c\| \leq \delta] + \mathbb{E} [\|f (X_n, Y_n) - f (X_n, c)\| : \|Y_n - c\| > \delta] \leq \varepsilon + 2MP (\|Y_n - c\| > \delta) \to \varepsilon \text{ as } n \to \infty, \]

where \( M = \sup |f| \). Since, \( X_n \Rightarrow X \), we know \( \mathbb{E} [f (X_n, c)] \to \mathbb{E} [f (X, c)] \) and hence we have shown,
\[ \limsup_{n \to \infty} |\mathbb{E} [f (X_n, Y_n) - f (X_n, c)]| \leq \limsup_{n \to \infty} |\mathbb{E} [f (X_n, Y_n) - f (X_n, c)]| + \limsup_{n \to \infty} |\mathbb{E} [f (X_n, c) - f (X_n, c)]| \leq \varepsilon. \]

As \( \varepsilon > 0 \) was arbitrary this proves Eq. (24.32).

\textbf{Theorem 24.44 (\( \delta \)-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
\[ Y_n := \frac{X_n - b}{a_n} \Rightarrow Z. \]

If \( g: \mathbb{R} \to \mathbb{R} \) be a measurable function which is differentiable at \( b \), then
\[ \frac{g(X_n) - g(b)}{a_n} \Rightarrow g'(b) Z. \] (24.33)

Put more informally, if
\[ X_n \overset{d}{\to} b + a_n Z \text{ then } g(X_n) \overset{d}{\to} g(b) + g'(b) a_n Z. \] (24.34)

\textbf{Proof.} Informally we have \( X_n = a_n Y_n + b \overset{d}{\Rightarrow} a_n Z + b \) and therefore
\[ \frac{g(X_n) - g(b)}{a_n} = \frac{g(a_n Y_n + b) - g(b)}{a_n Z} \to g'(b) Z \text{ as } n \to \infty. \]

We now make the proof rigorous.

By Skorohod’s Theorem 24.30 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 \( X_n := a_n Y_n + b \).

By the definition of the derivative of \( g \) at \( b \), we have
\[ g(b + \Delta) - g(b) = g'(b) \Delta + \varepsilon (\Delta) \Delta \]

where \( \varepsilon (\Delta) \to 0 \) as \( \Delta \to 0 \). Taking \( \Delta = a_n Y_n \) in this equation shows
\[ \frac{g(X_n) - g(b)}{a_n} = \frac{g(a_n Y_n + b) - g(b)}{a_n} = \frac{g'(b) a_n Y_n + \varepsilon (a_n Y_n) a_n Y_n}{a_n} \to g'(b) Z \text{ a.s.} \]

which implies Eq. (24.33) because of Corollary 24.35.

\textbf{Example 24.45.} Suppose that \( \{U_n\}_{n=1}^\infty \) are i.i.d. random variables which are uniformly distributed on \( [0, 1] \) and let \( Y_n := \prod_{j=1}^{n} \frac{U_j}{n} \). 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
\[ \mathbb{E} [\ln U_1] = \int_{0}^{1} \ln x \, dx = -1, \]
\[ \mathbb{E} [\ln U_1]^2 = \int_{0}^{1} \ln^2 x \, dx = 2, \]
\[ \text{Var}(\ln U_1) = 1 \] and so by the central limit theorem
\[ \sqrt{n} [X_n - (-1)] = \frac{\sum_{j=1}^{n} [\ln U_j + 1]}{\sqrt{n}} \implies Z \overset{d}{=} N(0,1). \]

In other words, \( \ln Y_n = X_n \overset{d}{=} -1 + \frac{1}{\sqrt{n}} Z \) and so we expect,
\[ Y_n \overset{d}{=} e^{-1 + \frac{1}{\sqrt{n}} Z} \simeq e^{-1} \left( 1 + \frac{1}{\sqrt{n}} Z + \ldots \right) \]
and thus we conjecture
\[ \sqrt{n} [Y_n - e^{-1}] \implies e^{-1} Z \overset{d}{=} N(0,e^{-2}). \]
To verify this is correct recall that by the \( \delta \)-method if \( g'(1) \) exits, then
\[ g(X_n) \overset{d}{=} g(1) + g'(1) \frac{1}{\sqrt{n}} Z. \]
Taking \( g(x) = e^x \) then implies \( Y_n \overset{d}{=} e^{-1} + e^{-1} \frac{1}{\sqrt{n}} Z \) or more precisely,
\[ \sqrt{n} \left[ \prod_{j=1}^{n} U_j^\frac{1}{n} - e^{-1} \right] = \sqrt{n} (Y_n - e^{-1}) \implies e^{-1} Z = N(0,e^{-2}). \]

**Exercise 24.8.** Given a function, \( f : X \to \mathbb{R} \) and a point \( x \in X \), let
\[ \liminf_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \inf_{y \in B_x'(\varepsilon)} f(y) \] and
\[ \limsup_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \sup_{y \in B_x'(\varepsilon)} f(y), \]
where
\[ B_x'(\varepsilon) := \{ y \in X : 0 < d(x,y) < \varepsilon \}. \]
Show \( f \) is lower (upper) semi-continuous iff \( \liminf_{y \to x} f(y) \geq f(x) \) \( \limsup_{y \to x} f(y) \leq f(x) \) for all \( x \in X \).

### 24.6 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
\[ \frac{X_n - b_n}{a_n} \implies Y. \] (24.37)

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
\[ \frac{X_n - \beta_n}{\alpha_n} \implies Z. \] (24.38)

Working informally we expect that
\[ X_n \overset{d}{=} \alpha_n Z + \beta_n \] and putting this expression back into Eq. (24.38) leads us to expect;
\[ \frac{\alpha_n Z + \beta_n - b_n}{a_n} = \frac{\alpha_n Z + \beta_n - b_n}{a_n} \implies Y. \]

It is reasonable to expect that this can only happen if the limits
\[ A = \lim_{n \to \infty} \frac{\alpha_n}{a_n} \in (0,\infty) \] and \( B := \lim_{n \to \infty} \frac{\beta_n - b_n}{a_n} \) (24.39)
exist and
\[ Y \overset{d}{=} AZ + B. \] (24.40)

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 24.49 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. (24.37) 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,
\[ \lim_{n \to \infty} F_n(b_n) = P(Y \leq 0) = \gamma_1 \in (0,1) \] 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
\[ P(Y \leq 0) = P(AZ + B \leq 0) = P(Z \leq -B/A) \] and
\[ P(Y \leq 1) = P(AZ + B \leq 1) = P(Z \leq (1 - B)/A). \]
Definition 24.46. Two random variables, Y and Z, are said to be of the same type if there exists constants, A > 0 and B ∈ ℜ such that Eq. (24.40) holds. Alternatively put, if U (y) := P (Y ≤ y) and V (z) := P (Z ≤ z), then U and V should satisfy

\[ V (z) = P (Z ≤ z) = P (Y ≤ Az + B) = U (Az + B) \]

for all z ∈ ℜ.

Remark 24.47. Suppose that Y ∼ AZ + B and Y and Z are square integrable random variables. Then

\[ EY = A · EZ + B \]

and \( Var (Y) = A^2 Var (Z) \)

from which it follows that \( A^2 = Var (Y) / Var (Z) \) and \( B = EY - A · EZ \). In particular, given \( Y \in L^2 (P) \) there is a unique Z of the same type such that \( EZ = 0 \) and \( Var (Z) = 1 \). On these grounds it is often reasonable to try to choose \( \{b_n\} \) and \( \{a_n > 0\} \) so that \( X_n := a_n^{-1} (X_n - b_n) \) has mean zero and variance one.

We will need the following elementary observation for the proof of Theorem 24.49.

Lemma 24.48. If Y is non-constant (a.s.) random variable and U (y) := P (Y ≤ 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 24.27 for the meaning of \( U^\leftarrow \).

Proof. Observe that Y is constant iff U (y) = 1_{y≥c} for some c ∈ ℜ, i.e., iff U only takes on the values, \{0, 1\}. So since Y is not constant, there exists y ∈ ℜ such that 0 < U (y) < 1. Hence if \( \gamma_2 > U (y) \) then \( U^\leftarrow (\gamma_2) \geq y \) and if \( \gamma_1 < U (y) \) then \( U^\leftarrow (\gamma_1) \leq y \). 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 24.49 (Convergence of Types). Suppose \( \{X_n\}_{n=1}^\infty \) is a sequence of random variables and \( a_n, \alpha_n ∈ (0, ∞) \), \( b_n, β_n ∈ ℜ \) are constants and Y and Z are non-constant random variables. Then

1. if both Eq. (24.37) and Eq. (24.38) hold then the limits, in Eq. (24.39) exists and Y ∼ AZ + B and in particular Y and Z are of the same type.
2. If the limits in Eq. (24.39) hold then either of the convergences in Eqs. (24.37) or (24.38) implies the others with Z and Y related by Eq. (24.41).
3. If there are some constants, \( a_n > 0 \) and \( b_n ∈ ℜ \) and a non-constant random variable Y, such that Eq. (24.37) holds, then Eq. (24.38) holds using \( α_n \) and \( β_n \) of the form,

\[ \alpha_n := F_n^\leftarrow (\gamma_2) - F_n^\rightarrow (\gamma_1) \]

and \( \beta_n := F_n^\rightarrow (\gamma_1) \)

for some \( 0 < \gamma_1 < \gamma_2 < 1 \). If the \( F_n \) are invertible functions, Eq. (24.41) may be written as

\[ F_n (\beta_n) = \gamma_1 \]

and \( F_n (\alpha_n + \beta_n) = \gamma_2 \).

Proof. (2) Assume the limits in Eq. (24.39) hold. If Eq. (24.37) is satisfied, then by Slutsky’s Theorem 16.23,

\[ \frac{X_n - b_n}{a_n} = \frac{X_n - b_n + b_n - \beta_n a_n}{a_n} \]

\[ = \frac{X_n - b_n}{a_n} - \frac{\beta_n - b_n a_n}{a_n} \]

\[ \Rightarrow A^{-1} (Y - B) =: Z \]

Similarly, if Eq. (24.38) is satisfied, then

\[ \frac{X_n - b_n}{a_n} = \frac{X_n - \beta_n a_n + \beta_n - b_n}{a_n} \]

\[ \Rightarrow AZ + B =: Y. \]

(1) If \( F_n (y) := P (X_n ≤ y) \), then

\[ P \left( \frac{X_n - b_n}{a_n} ≤ y \right) = F_n (a_n y + b_n) \]

and \( P \left( \frac{X_n - \beta_n}{a_n} ≤ y \right) = F_n (\alpha_n y + \beta_n) \).

By assumption we have

\[ F_n (a_n y + b_n) \Rightarrow U (y) \]

and \( F_n (\alpha_n y + \beta_n) \Rightarrow V (y) \).

If \( w := \text{sup} \{y : F_n (a_n y + b_n) < x\} \), then \( a_n w + b_n = F_n^\rightarrow (x) \) and hence

\[ \text{sup} \{y : F_n (a_n y + b_n) < x\} = \frac{F_n^\rightarrow (x) - b_n}{a_n}. \]

Similarly,

\[ \text{sup} \{y : F_n (\alpha_n y + \beta_n) < x\} = \frac{F_n^\rightarrow (x) - \beta_n}{\alpha_n}. \]

With these identities, it now follows from the proof of Skorohod’s Theorem 24.30 (see Eq. (24.23)) that there exists an at most countable subset, A, of \( (0, 1) \) such that,

\[ \frac{F_n^\rightarrow (x) - b_n}{a_n} = \text{sup} \{y : F_n (a_n y + b_n) < x\} \Rightarrow U^\rightarrow (x) \]

and

\[ \frac{F_n^\rightarrow (x) - \beta_n}{\alpha_n} = \text{sup} \{y : F_n (\alpha_n y + \beta_n) < x\} \Rightarrow V^\rightarrow (x) \]
for all \( x \notin A \). Since \( Y \) and \( Z \) are not constants a.s., we can choose, by Lemma 24.48, \( \gamma_1 < \gamma_2 \) not in \( A \) such that \( U^\leftarrow(\gamma_1) < U^\leftarrow(\gamma_2) \) and \( V^\leftarrow(\gamma_1) < V^\leftarrow(\gamma_2) \). In particular it follows that
\[
\frac{F_n^\leftarrow(\gamma_2) - F_n^\leftarrow(\gamma_1)}{a_n} = \frac{F_n^\leftarrow(\gamma_2) - b_n}{a_n} - \frac{F_n^\leftarrow(\gamma_1)}{a_n} \to U^\leftarrow(\gamma_2) - U^\leftarrow(\gamma_1) > 0 \quad (24.43)
\]
and similarly
\[
\frac{F_n^\leftarrow(\gamma_2) - F_n^\leftarrow(\gamma_1)}{a_n} \to V^\leftarrow(\gamma_2) - V^\leftarrow(\gamma_1) > 0.
\]

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,
\[
\frac{F_n^\leftarrow(\gamma_1) - b_n}{a_n} \to U^\leftarrow(\gamma_1) \quad (24.44)
\]
\[
\frac{F_n^\leftarrow(\gamma_1) - \beta_n}{a_n} = \frac{F_n^\leftarrow(\gamma_1) - \beta_n \alpha_n}{a_n} \to AV^\leftarrow(\gamma_1)
\]
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 \( \alpha_n := \frac{F_n^\leftarrow(\gamma_2) - F_n^\leftarrow(\gamma_1)}{a_n} \) and \( \beta_n := F_n^\leftarrow(\gamma_1) \), then according to Eqs. (24.43) and (24.44) we have
\[
\frac{\alpha_n}{a_n} \to U^\leftarrow(\gamma_2) - U^\leftarrow(\gamma_1) \in (0, 1) \quad \text{and} \quad \frac{\beta_n - b_n}{a_n} \to U^\leftarrow(\gamma_1) \text{ as } n \to \infty.
\]
Thus we may always center and scale the \( \{X_n\} \) using \( \alpha_n \) and \( \beta_n \) of the form described in Eq. (24.41). ■

### 24.7 Weak Convergence Examples

**Example 24.50.** 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
\[
F(x) := P(X_1 \leq x) = 1 - e^{-\lambda(x/\lambda)} = (1 - e^{-\lambda x})_+
\]
Consider \( M_n := \max(X_1, \ldots, X_n) \). We have, for \( x \geq 0 \) and \( c_n \in (0, \infty) \) that
\[
F_n(x) := P(M_n \leq x) = P(\cap_{j=1}^n \{X_j \leq x\}) = \prod_{j=1}^n P(X_j \leq x) = [F(x)]^n = (1 - e^{-\lambda x})^n.
\]
We now wish to find \( a_n > 0 \) and \( b_n \in \mathbb{R} \) such that \( \frac{M_n - b_n}{a_n} \Rightarrow Y \).

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. (24.42) 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
\[
\ln \gamma_1 \sim n \ln F(b_n) = n \ln (1 - e^{-\lambda b_n}) \sim -n e^{-\lambda b_n}.
\]
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
\[
\lim_{n \to \infty} P(M_n - \lambda^{-1} \ln n \leq x) = \lim_{n \to \infty} \left(1 - e^{-\lambda(x + \lambda^{-1} \ln n)}\right)^n = \lim_{n \to \infty} \left(1 - \frac{e^{-\lambda x}}{n}\right)^n = e^{-e^{-\lambda x}}.
\]

The function \( F(x) = \exp(-e^{-\lambda x}) \) is the CDF for a “Gumbel distribution,” see Figure 24.5. Thus letting \( Y \) be a random variable with this distribution (i.e. \( P(Y \leq x) = \exp(-e^{-\lambda x}) \)) we have shown \( M_n - \frac{1}{\lambda} \ln n \Rightarrow Y \), i.e.
\[
\max(X_1, \ldots, X_n) - \frac{1}{\lambda} \ln n \Rightarrow Y.
\]
Example 24.51. 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 $x > 0$,

$$P(pX_p > x) = P(X_p > \frac{x}{p}) = (1-p)^{\frac{x}{p}} = e^{\frac{x}{p}\ln(1-p)} \sim e^{-\frac{px}{p}} \to e^{-x} \text{ as } p \to 0.$$ 

Therefore $pX_p \implies T$ where $T \overset{d}{=} \exp(1)$, i.e. $P(T > x) = e^{-x}$ for $x \geq 0$ or alternatively, $P(T \leq y) = 1 - e^{-y}$. 

**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 $b_p = 1/p$. 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

$$\ln(1 - F_0(x)) = (a_p x + b_p) \ln q \sim (a_p x + b_p)(-p) = -pa_p x - 1.$$ 

From this (setting $x = 1$) we see that $pa_p \sim c > 0$. Hence we might take $a_p = 1/p$ as well. We then have

$$F_p(a_p x + b_p) = F_p(p^{-1} x + p^{-1}) = 1 - (1-p)^{[p^{-1}(x+1)]}$$

which is equal to 0 if $x \leq -1$, and for $x > -1$ we find

$$(1-p)^{[p^{-1}(x+1)}] = \exp (\left[ p^{-1}(x + 1) \right] \ln (1-p)) \to \exp (- (x + 1)).$$

Hence we have shown,

$$\lim_{p \downarrow 0} F_p(a_p x + b_p) = [1 - \exp(- (x + 1))] 1_{x \geq -1}$$

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

$$f(z) := E[z^{X_p}] = \sum_{n=1}^{\infty} z^n q^{n-1}p = \frac{pz}{1 - qz}.$$ 

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

$$f'(z) = E[X_p z^{X_p-1}], \quad f''(z) = E[X_p (X_p - 1) z^{X_p-2}], \ldots$$

$$f^{(k)}(z) = E[X_p (X_p - 1) \ldots (X_p - k + 1) z^{X_p-k}]$$

and in particular,

$$E[X_p (X_p - 1) \ldots (X_p - k + 1)] = f^{(k)}(1) = \left( \frac{d}{dz} \right)_{z=1} \frac{pz}{1 - qz}. $$
Since
\[
\frac{d}{dz} \frac{pz}{1 - qz} = p \frac{(1 - qz) + qpz}{(1 - qz)^2} = \frac{p}{(1 - qz)^2}
\]
and
\[
\frac{d^2}{dz^2} \frac{pz}{1 - qz} = 2 \frac{pq}{(1 - qz)^3}
\]
it follows that
\[
\mu_p := \mathbb{E}X_p = \frac{p}{(1 - qz)^2} = \frac{1}{p}
\]
and therefore,
\[
\mathbb{E}[X_p(X_p - 1)] = 2 \frac{pq}{(1 - qz)^3} = \frac{2q}{p^2}.
\]
Therefore,
\[
\sigma^2_p = \text{Var}(X_p) = \mathbb{E}X_p^2 - (\mathbb{E}X_p)^2 = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2
\]
\[
= \frac{2q + p - 1}{p^2} = \frac{q}{p^2} = \frac{1 - p}{p^2}.
\]
Thus, if we had used \(\mu_p\) and \(\sigma_p\) to center and scale \(X_p\) we would have considered,
\[
\frac{X_p - \frac{1}{p}}{\sqrt{\frac{1}{p}}} = \frac{pX_p - 1}{\sqrt{1 - p}} \Rightarrow T - 1
\]
instead.

**Theorem 24.52** (This is already done in Theorem 10.74). 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 \(\frac{S_n}{\sqrt{n}} \Rightarrow N(0, 1)\) as \(m \to \infty\).

**Proof.** (Sketch of the proof.) We start by observing that \(S_{2n} = 2k\) iff
\[
\#\{i \leq 2n : X_i = 1\} = n + k \quad \text{while}\quad \#\{i \leq 2n : X_i = -1\} = 2n - (n + k) = n - k
\]
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,
\[
n! \sim n^n e^{-n} \sqrt{2\pi n} \quad \text{as} \quad n \to \infty
\]
and therefore,
\[
P(S_{2n} = 2k)
\]
\[
\sim (n+k)^{n+k} e^{-(n+k)} 2^n (n+k)! \cdot (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi (n-k)} \left(\frac{1}{2}\right)^{2n}
\]
\[
= \sqrt{\frac{n}{\pi}} \frac{1}{(n+k)(n-k)} \left(1 + \frac{k}{n}\right)^{-(n+k)} \left(1 - \frac{k}{n}\right)^{-(n-k)}
\]
\[
= \frac{1}{\sqrt{\pi n}} \left(1 - \frac{k^2}{n^2}\right)^{-n} \left(1 + \frac{k}{n}\right)^{-k} \left(1 - \frac{k}{n}\right)^{k-1/2}
\]
So if we let \(x := 2k/\sqrt{2n}\), i.e. \(k = x \sqrt{n/2}\) and \(k/n = x/\sqrt{2n}\), we have
\[
P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right)
\]
\[
\sim \frac{1}{\sqrt{\pi n}} \left(1 - \frac{x^2}{2n}\right)^{-n} \left(1 + \frac{x}{\sqrt{2n}}\right)^{-x \sqrt{n/2 - 1/2}} \left(1 - \frac{x}{\sqrt{2n}}\right)^{x \sqrt{n/2 - 1/2}}
\]
\[
= \frac{1}{\sqrt{\pi n}} e^{x^2/2} e^{\frac{x}{\sqrt{2n}} \left(-x \sqrt{n/2 - 1/2}\right)} e^{-\frac{x}{\sqrt{2n}} \left(x \sqrt{n/2 - 1/2}\right)}
\]
\[
\sim \frac{1}{\sqrt{\pi n}} e^{-x^2/2},
\]
wherein we have repeatedly used
\[
(1 + a_n)^{b_n} = e^{b_n \ln (1 + a_n)} \sim e^{b_n a_n} \quad \text{when} \quad a_n \to 0.
\]
We now compute
\[
P\left(a \leq \frac{S_{2n}}{\sqrt{2n}} \leq b\right) = \sum_{a \leq x \leq b} P\left(\frac{S_{2n}}{\sqrt{2n}} = x\right)
\]
\[
= \frac{1}{\sqrt{2\pi}} \sum_{a \leq x \leq b} e^{-x^2/2} \frac{2}{\sqrt{2n}}
\]
\[
\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.
\]

This proves \( \frac{S_{2n}}{\sqrt{2n}} \rightarrow N(0,1) \). 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 \[\text{24.36}\]) that \( \frac{S_{2n+1}}{\sqrt{2n+1}} \rightarrow N(0,1) \) as well.

**Proposition 24.53.** 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, \( \{U_1, U_2, \ldots, U_n\} \). 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

\[
X_n := \frac{U_{(k(n),n)} - k(n)/n}{\sqrt{k(n)/n}}.
\]

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

\[
P \left( U_{(k,n)} \leq x \right) = P \left( \sum_{i=1}^{n} X_i \geq k \right) = \sum_{l=k}^{n} \binom{n}{l} x^l (1-x)^{n-l}.
\]

From this it follows that \( \rho_n(x) := 1_{(0,1)}(x) \frac{d}{dx} P \left( U_{(k,n)} \leq x \right) \) is the probability density for \( U_{(k,n)} \). It now turns out that \( \rho_n (x) \) is a Beta distribution,

\[
\rho_n(x) = \binom{n}{k} x^k (1-x)^{n-k}.
\]

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, \( U_{(k,n)} \leq x + \Delta \) would imply \( U_{(k,n)} \leq x \) as well and hence \( U_{(k,n)} \notin (x, x+\Delta) \). Let

\[
\Omega_i := \{ U_i \in (x, x+\Delta) \text{ and } U_j \notin (x, x+\Delta) \text{ for } j \neq i \}.
\]

Since

\[
P \left( U_i, U_j \in (x, x+\Delta) \text{ for some } i \neq j \text{ with } i, j \leq n \right) \leq \sum_{i<j \leq n} P \left( U_i, U_j \in (x, x+\Delta) \right) \leq \frac{n^2-n}{2} \Delta^2,
\]

we see that

\[
P \left( U_{(k,n)} \in (x, x+\Delta) \right) = \sum_{i=1}^{n} P \left( U_{(k,n)} \in (x, x+\Delta), \Omega_i \right) + O(\Delta^2) = nP \left( U_{(k,n)} \in (x, x+\Delta), \Omega_i \right) + O(\Delta^2).
\]

Now on the set, \( \Omega_i; U_{(k,n)} \in (x, x+\Delta) \) iff there are exactly \( k-1 \) of \( U_1, \ldots, U_n \) in \([0,x]\) and \( n-k \) of these in \([x+\Delta,1]\). This leads to the conclusion that

\[
P \left( U_{(k,n)} \in (x, x+\Delta) \right) = n \binom{n-1}{k-1} (1-x) \Delta + O(\Delta^2)
\]

and therefore,

\[
\rho_n(x) = \frac{N}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k}.
\]

By Stirling’s formula,

\[
\frac{n!}{(k-1)! \cdot (n-k)!} \sim \frac{n^ke^{-n} \sqrt{2\pi n}}{(k-1)! e^{-(k-1)} \sqrt{2\pi (k-1)} (n-k)! e^{-(n-k)} \sqrt{2\pi (n-k)}} = \frac{n^k e^{-n} \sqrt{2\pi}}{(k-1)! \sqrt{2\pi (k-1)} (n-k)! \sqrt{2\pi (n-k)}} = \frac{\sqrt{\pi} e^{-1}}{2} \left( \frac{k-1}{n} \right)^{(k-1)/2} \frac{1}{(1 - \frac{k}{n})(n-k+1/2)}.
\]

Since

\[
\left( \frac{k-1}{n} \right)^{(k-1)/2} = \left( \frac{k}{n} \right)^{(k-1)/2} \cdot \left( 1 - \frac{1}{k} \right)^{(k-1)/2} \sim e^{-1} \left( \frac{k}{n} \right)^{(k-1)/2},
\]

\[
\rho_n(x) = \left( \frac{k}{n} \right)^{(k-1)/2} \cdot \left( 1 - \frac{1}{k} \right)^{(k-1)/2} \sim e^{-1} \left( \frac{k}{n} \right)^{(k-1)/2}.
\]
we arrive at
\[
\frac{n!}{(k-1)! \cdot (n-k)!} \sim \frac{\sqrt{n}}{2\pi} \left( \frac{k}{n} \right)^{(k-1)/2} \left( 1 - \frac{k}{n} \right)^{(n-k+1)/2}.
\]

By the change of variables formula, with
\[
x = \frac{u - k(n)/n}{\sqrt{k(n)}},
\]
on noting the \(du = \frac{\sqrt{k(n)}}{n} dx\), \(x = -\sqrt{k(n)}\) at \(u = 0\), and
\[
x = \frac{1 - k(n)/n}{\sqrt{k(n)}} = \frac{n - k(n)}{\sqrt{k(n)}},
\]
we arrive at
\[
\frac{n}{\sqrt{k(n)}} \left( 1 - \frac{k(n)}{n} \right) = \sqrt{n} \frac{n}{k(n)} \left( 1 - \frac{k(n)}{n} \right) =: b_n,
\]
\[
\mathbb{E}[F(X_n)] = \int_{0}^{1} \rho_n(u) \frac{F \left( \frac{u - k(n)/n}{\sqrt{k(n)}} \right)}{\sqrt{k(n)}} du
\]
\[
= \int_{-b_n}^{b_n} \frac{\sqrt{k(n)}}{n} \rho_n \left( \frac{\sqrt{k(n)}}{n} x + k(n)/n \right) F(x) du.
\]
Using this information, it is then shown in Resnick that
\[
\frac{\sqrt{k(n)}}{n} \rho_n \left( \frac{\sqrt{k(n)}}{n} x + k(n)/n \right) \rightarrow e^{-x^2/2}/2\pi
\]
which upon an application of Scheffé’s Lemma 24.8 completes the proof. \(\blacksquare\)

**Remark 24.54.** 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
\[
\mathbb{E}[U_{(k,n)}] = \int_{0}^{1} \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k} x dx.
\]
\[
= \frac{k}{n+1} \sim \frac{k}{n}.
\]
\[
\mathbb{E}[U_{(k,n)}^2] = \int_{0}^{1} \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k} x^2 dx.
\]
\[
= \frac{(k+1)k}{(n+2)(n+1)} + \frac{k}{n+1} \sim \frac{k}{n^2}.
\]

### 24.8 Compactness and tightness of measures on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\)

**Notation 24.55.** 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 24.56.** A subset \(\Gamma \subset \mathcal{P}_1(\mathbb{R}, \mathcal{B}_\mathbb{R})\) is weakly sequentially pre-compact iff for every sequence, \(\{\mu_n\}_{n=1}^\infty \subset \Gamma\), there exists a subsequence \(\{\mu_{n_k}\}_{k=1}^\infty\) such that \(\mu_{n_k} \Rightarrow \mu_0\) for some probability measure \(\mu_0 \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_\mathbb{R})\).

**Definition 24.57.** 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
\[
\inf_{\mu \in \Gamma} \mu([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon.
\]

We further say that a collection of random variables, \(\{X_\lambda : \lambda \in \Lambda\}\) is tight iff the collection probability measures, \(\{P \circ X_\lambda^{-1} : \lambda \in \Lambda\}\) is tight. Equivalently put, \(\{X_\lambda : \lambda \in \Lambda\}\) is tight iff
\[
\lim_{M \rightarrow \infty} \sup_{\lambda \in \Lambda} P(|X_\lambda| \geq M) = 0.
\]

Observe that the definition of uniform integrability (see Definition 15.38) is considerably stronger than the notion of tightness. It is also worth observing that if \(\alpha > 0\) and \(C := \sup_{\lambda \in \Lambda} \mathbb{E}[|X_\lambda|^\alpha] < \infty\), then by Chebyshev’s inequality,
\[
\sup_{\lambda} P(|X_\lambda| \geq M) \leq \sup_{\lambda} \left[ \frac{1}{M^\alpha} \mathbb{E}[|X_\lambda|^\alpha] \right] \leq \frac{C}{M^\alpha} \rightarrow 0 \text{ as } M \rightarrow \infty
\]
and therefore \(\{X_\lambda : \lambda \in \Lambda\}\) is tight.
Proposition 24.58. If $\Gamma \subset \mathcal{P}_1(\mathbb{R}, \mathcal{B})$ is weakly sequentially pre-compact, then $\Gamma$ is tight.

**Proof.** For sake of contradiction, suppose there exists $\Gamma \subset \mathcal{P}_1(\mathbb{R}, \mathcal{B})$ which is weakly sequentially pre-compact but is not tight. Since $\Gamma$ is not tight, there exists an $\varepsilon > 0$ such that $\inf_{\nu \in \Gamma} \mu([-M, M]) < 1 - \varepsilon$ 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, $\{\nu_k := \mu_{n_k}\}_{k=1}^{\infty}$ such that $\nu_k \implies \mu_0$. Hence for every $M \in \mathbb{R}_+$ such that $M \in C(\mu_0)$ we have

$$
\mu_0((-M, M]) = \lim_{k \to \infty} \nu_k((-M, M]) \leq \liminf_{k \to \infty} \nu_k([-n_k, n_k]) \leq 1 - \varepsilon.
$$

We may now let $M \in C(\mu_0)$ tend to infinity in order to conclude $\mu_0(\mathbb{R}) \leq 1 - \varepsilon < 1$ which contradicts the assumption that $\mu_0$ is a probability measure. \hfill $\blacksquare$

The goal of this section is to prove the converse of Proposition 24.58, see Theorem 24.64. 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 $A \subset \mathbb{R}$ is a dense set and $F$ and $\tilde{F}$ are two right continuous functions. If $F = \tilde{F}$ on $A$, then $F = \tilde{F}$ on $\mathbb{R}$. Indeed, for $x \in \mathbb{R}$ we have

$$
F(x) = \lim_{A \ni \lambda \to x} F(\lambda) = \lim_{A \ni \lambda \to x} \tilde{F}(\lambda) = \tilde{F}(x).
$$

**Lemma 24.59.** If $G : A \to \mathbb{R}$ is a non-decreasing function, then

$$
F(x) := G_+(x) := \inf\{G(\lambda) : x < \lambda \in A\} \tag{24.48}
$$

is a non-decreasing right continuous function.

**Proof.** To show $F$ is right continuous, let $x \in \mathbb{R}$ and $\lambda \in A$ such that $\lambda > x$. Then for any $y \in (x, \lambda)$,

$$
F(y) = G_+(y) \leq G(\lambda)
$$

and therefore,

$$
F(x) \leq F(x+) := \lim_{y \ni x} F(y) \leq G(\lambda).
$$

Since $\lambda > x$ with $\lambda \in A$ is arbitrary, we may conclude, $F(x) \leq F(x+) \leq G_+(x) = F(x)$, i.e. $F(x+) = F(x)$.

**Proposition 24.60.** Suppose that $\{F_n\}_{n=1}^{\infty}$ is a sequence of distribution functions and $A \subset \mathbb{R}$ is a dense set such that $G(\lambda) := \lim_{n \to \infty} F_n(\lambda)$ is $[0, 1]$ exists for all $\lambda \in A$. If, for all $x \in \mathbb{R}$, we define $F = G_+$ as in Eq. (24.48), then $F_n(x) \to F(x)$ for all $x \in C(F)$ [see Notation 24.17]. (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 $(\mathbb{R}, \mathcal{B})$.)

**Proof.** Suppose that $x, y \in \mathbb{R}$ with $x < y$ and and $s, t, \lambda \in A$ are chosen so that $x < s < y < t$. Then passing to the limit in the inequality,

$$
F_n(s) \leq F_n(y) \leq F_n(t)
$$

implies

$$
F(x) = G_+(x) \leq G(s) \leq \liminf_{n \to \infty} F_n(y) \leq \limsup_{n \to \infty} F_n(y) \leq G(t).
$$

Taking the infimum over $t \in A \cap (y, \infty)$ and then letting $x \in \mathbb{R}$ tend up to $y$, we may conclude

$$
F(y-) \leq \liminf_{n \to \infty} F_n(y) \leq \limsup_{n \to \infty} F_n(y) \leq F(y) \text{ for all } y \in \mathbb{R}.
$$

This completes the proof, since $F(y-) = F(y)$ for $y \in C(F)$.

The next theorem deals with weak convergence of measures on $(\mathbb{R}, \mathcal{B})$. So as not have to introduce any new machinery, the reader should identify $\mathbb{R}$ with $[-1, 1] \subset \mathbb{R}$ via the map,

$$
[-1, 1] \ni x \to \frac{\pi}{2} x \in \mathbb{R}.
$$

Hence a probability measure on $(\mathbb{R}, \mathcal{B})$ may be identified with a probability measure on $(\mathbb{R}, \mathcal{B})$ 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, $F : \mathbb{R} \to [0, 1]$ if and only if $F(-\infty) = 0$. On the other hand, $\infty$ is always a point of continuity.

**Theorem 24.61 (Helly’s Selection Theorem).** Every sequence of probability measures, $\{\mu_n\}_{n=1}^{\infty}$, on $(\mathbb{R}, \mathcal{B})$ has a sub-sequence which is weakly convergent to a probability measure, $\mu_0$ on $(\mathbb{R}, \mathcal{B})$.

**Proof.** Using the identification described above, rather than viewing $\mu_n$ as probability measures on $(\mathbb{R}, \mathcal{B})$, we may view them as probability measures on $(\mathbb{R}, \mathcal{B})$ 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 $\{F_n(x)\}_{n=1}^{\infty}$ such that $G(x) := \lim_{n \to \infty} G_k(x)$ exists for all $x \in A := \mathbb{Q}$.

Letting $F(x) := G(x)$ as in Eq. (24.48), it follows from Lemma 24.59 and Proposition 24.60 that $G_k = F_{n_k} \implies F_k$. 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 $x \geq 1$ 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 $(\mathbb{R}, \mathcal{B})$.
Example 24.62. Here are there simple examples showing that probabilities may indeed transfer to the points at ±∞: 1) \( \delta_{-n} \Rightarrow \delta_{-\infty} \), 2) \( \delta_{n} \Rightarrow \delta_{\infty} \) and 3) \( \frac{1}{2} (\delta_{n} + \delta_{-n}) \Rightarrow \frac{1}{2} (\delta_{\infty} + \delta_{-\infty}) \).

Theorem 24.63 (Helly’s Selection Theorem for \( \mathbb{R}^n \)). Every sequence of probability measures, \( \{\mu_n\}_{n=1}^{\infty} \), on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) has a sub-sequence which is weakly convergent to a probability measure, \( \mu_0 \) on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \).

Proof. The proof is very similar to Theorem 24.61 provided we replace Lemma 24.59 by Lemma 6.63 and use a multi-variate version of Proposition 24.60. The next question we would like to address is when is the limiting measure, \( \mu_0 \) on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) concentrated on \( \mathbb{R} \).

Theorem 24.64. A subset, \( \Gamma \subset \mathcal{P}_1 (\mathbb{R}, \mathcal{B}_\mathbb{R}) \subset \mathcal{P}_1 (\overline{\mathbb{R}}, \mathcal{B}_\overline{\mathbb{R}}) \), is tight, iff every subsequently limit measure, \( \mu_0 \in \mathcal{P}_1 (\overline{\mathbb{R}}, \mathcal{B}_\overline{\mathbb{R}}) \), 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 24.89 below.) [This theorem generalizes to \( \mathbb{R}^n \) in a fairly obvious way.]

Proof. Suppose that \( \Gamma \ni \mu_{nk} \Rightarrow \mu_0 \) with \( \mu_0 \) being a probability measure on \( (\overline{\mathbb{R}}, \mathcal{B}_\overline{\mathbb{R}}) \) – 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 \( M < \infty \) such that \( M \varepsilon < \infty \) such that \( M \varepsilon, -M \varepsilon \in C (F_0) \) and \( \mu ([M \varepsilon, M \varepsilon]) \geq 1 - \varepsilon \) for all \( \mu \in \Gamma \). Hence it follows that

\[
\mu_0 ((-M \varepsilon, M \varepsilon)) = \lim_{k \to \infty} \mu_{nk} ((-M \varepsilon, M \varepsilon)) \geq 1 - \varepsilon
\]

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 \( \{\mu_{nk}\}_{k=1}^{\infty} \) such that \( \mu_{nk} \Rightarrow \mu_0 \) with \( \mu_0 \) a probability measure on \( (\overline{\mathbb{R}}, \mathcal{B}_\overline{\mathbb{R}}) \) 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

\[
\mu_0([-M,M]) \leq \mu_0(\overline{\mathbb{R}}) - \mu_0(-\infty,\infty) = 1 - \varepsilon_0.
\]

By choosing \( M \) so that \(-M \) and \( M \) are points of continuity of \( F_0 \), it then follows that

\[
\lim_{k \to \infty} \mu_{nk}([-M,M]) = \mu_0([-M,M]) \leq 1 - \varepsilon_0.
\]

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.

24.9 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 6.9. The main goal should be Theorem 24.84.

Definition 24.65. 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

\[
(a,b) = \mathbb{R} \setminus \big[ S_{(a_1,b_2,...,b_n)} \cup S_{(b_1,a_2,...,b_n)} \cup \cdots \cup S_{(b_1,b_2,...,b_{n-1},a_n)} \big];
\]

it follows that \( \{S_x\}_{x \in \mathbb{R}^n} \) is a \( \sigma \) – 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 24.77 at the end of this section. Let us begin with the following basic analytic result.

Theorem 24.66. Suppose \( \alpha, \beta \in \mathbb{R}^n \) with \( \alpha < \beta \) and \( \mu : C := A_{(\alpha,\beta)} \to [0,\infty) \) is a finitely additive measure. Then \( \mu \) extends to a measure on \( B_{(\alpha,\beta)} = \sigma (C) \) 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 \( C \) and the \( \sigma \) – algebra \( B_{(\alpha,\beta)} \). 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

\[
\mu ((a,b]) \leq \sum_{n=1}^{\infty} \mu ((a_n, b_n]).
\]

(24.49)

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

\[
\mu ((a_n, \tilde{b}_n \wedge b]) \leq \mu ((a_n, b_n]) + \varepsilon 2^{-n} \text{ for } n \in \mathbb{N}.
\]
We have 

\[ \hat{a}, \hat{b} \subset (a, b) \subset \bigcup_n (a_n, \hat{b}_n) . \]

So by compactness there exists \( N < \infty \) such that 

\[ (\hat{a}, \hat{b}) \subset [\hat{a}, \hat{b}] \subset \bigcup_{n=1}^N (a_n, \hat{b}_n) \cap (a, b) \subset \bigcup_{n=1}^N (a_n, \hat{b}_n \land b) \]

By finite sub-additivity it now follows that 

\[
\mu((\hat{a}, \hat{b})) \leq \sum_{n=1}^N \mu((a_n, \hat{b}_n \land b)) \\
\leq \sum_{n=1}^N \mu((a_n, \hat{b}_n \land b)) \\
\leq \sum_{n=1}^N [\mu((a_n, b_n)] + \varepsilon 2^{-n}] \\
\leq \sum_{n=1}^\infty \mu((a_n, b_n)] + \varepsilon .
\]

As \( \varepsilon > 0 \) was arbitrary we may conclude,

\[
\mu((\hat{a}, \hat{b})) \leq \sum_{n=1}^\infty \mu((a_n, b_n)].
\]

Then using the right continuity in the first variable we may now let \( \hat{a} \downarrow a \) to conclude that Eq. (24.49) holds.

\[ \nabla \]

**Lemma 24.67.** 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

\[
F(x) := \inf_{y \leq x} G(y).
\]

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

\[
\inf_{x \in (a, b) \cap D^n} F(x) = F(a) = \inf_{y \in (a, b) \cap D^n} G(y) \tag{24.51}
\]

\[
\sup_{x \in (a, b) \cap D^n} F(x) = \sup_{y \in (a, b) \cap D^n} G(y) \text{ and } \tag{24.52}
\]

**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. (24.51) 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

\[
G(y) \leq F(x) \leq \sup_{x \in (a, b)} F(x) \implies \sup_{y \in (a, b) \cap D^n} G(y) \leq \sup_{x \in (a, b)} F(x).
\]

\[ \nabla \]

**Lemma 24.68.** 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

\[
\lim_{k \to \infty} F_k(x) = F(x) \text{ for all } x \in C(F)
\]

where \( F \) is given as in Eq. (24.50) and \( C(F) \) are the points of continuity of \( F \).

**Proof.** Let \( x \in C(F) \) 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 24.67 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.

\[ \nabla \]

**Corollary 24.69.** 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 \( A \subset \mathbb{N} \) such that

\[
\lim_{A \ni k \to \infty} F_k(x) = F(x) \text{ exists for all } x \in C(F).
\]
Proof. Let $D \subset \mathbb{R}$ be a countable dense set and then use Cantor’s diagonalization argument to find a $A \subset \mathbb{N}$ such that

$$G(y) := \lim_{A \ni k \to \infty} F_k(y)$$

exists for $y \in D^n$. Now apply Lemma 24.68 to finish the proof.

24.9.1 Finitely additive measures for $\mathbb{R}^n$

Suppose $V$ is a vector space and $\mu : A \to V$ is a finitely additive measure. Let $F(x) := \mu(S_x)$ for all $x \in \mathbb{R}^n$. If $a < b \in \mathbb{R}^n$, then

$$1_{(a,b]} = \prod_{i=1}^n 1_{(a_i,b_i]} = \prod_{i=1}^n \left[ 1_{(-\infty,b_i]} - 1_{(-\infty,a_i]} \right]$$

$$= \sum_{y \in \prod_{i=1}^n (a_i,b_i]} \text{sgn}(y)1_{(-\infty,y]} = \sum_{k \in \{0,1\}^n} (-1)^{|k|} 1_{(-\infty,y^k(a,b))},$$

where

$$\text{sgn}(y) := (-1)^{\# \{i : y_i = a_i \}},$$

$$|k| := k_1 + \cdots + k_n$$

and $\{y^k(a,b)\}_{k \in \{0,1\}^n}$ are the $2^n$ vertices of $[a,b]$ indexed as:

$$[y^k(a,b)]_i := \begin{cases} a_i & \text{if } k_i = 1 \\ b_i & \text{if } k_i = 0. \end{cases}$$

Proposition 24.70. Suppose $F$ and $\mu$ are as above, then for every $a, b \in \mathbb{R}^n$ with $a < b$,

$$\mu((a,b]) = \sum_{y \in \prod_{i=1}^n (a_i,b_i]} \text{sgn}(y)F(y)$$

(24.53)

$$= \sum_{k \in \{0,1\}^n} (-1)^{|k|} F(y^k(a,b)).$$

(24.54)

This is really a statement of the inclusion-exclusion formula. For example, when $n = 2$ we have

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

and by the inclusion-exclusion formula (using $S_{(a_1,b_2)} \cap S_{(b_1,a_2)} = S_a$) we have

$$\mu((a,b]) = \mu(S_b) - \mu(S_{(a_1,b_2)}) - \mu(S_{(b_1,a_2)}) + \mu(S_a)$$

$$= F(b) - F(a_1,b_1) - F(b_1,a_2) + F(a).$$

It will be convenient to rewrite Eq. (24.53) in another more useful way.

Notation 24.71. For $i \in \{1, 2, \ldots, n\}$ and $\alpha \in \mathbb{R}$ let

$$(e_i^\alpha)(x) := G(x) \mid_{x_i = \alpha} = G(x_1, \ldots, x_{i-1}, \alpha, x_{i+1}, \ldots, x_n)$$

and for $\alpha < \beta$ in $\mathbb{R}$ let

$$\Delta_i^{(\alpha,\beta)} G(x) = G(x) \mid_{x_i = \beta} - G(x_1, \ldots, x_{i-1}, \beta, x_{i+1}, \ldots, x_n).$$

Since $e_i^\alpha e_j^\beta = e_j^\beta e_i^\alpha$ for all $i \neq j$, it follows that the $\Delta_i$ and $\Delta_j$ also commute. With this notation we have,

$$\mu((a,b]) = \mu((a_1,b_1] \times (\tilde{a}, \tilde{b})) = \Delta_1^{(a_1,b_1)} \mu((\infty, x_1) \times (\tilde{a}, \tilde{b})).$$

Then working inductively we find

$$\mu((a,b]) = \Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} \mu(S_x)$$

$$= \Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} F$$

(24.55)

This formula agrees with Eq. (24.53) since,

$$\Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} F = \prod_{i=1}^n (e_i^{b_i} - e_i^{a_i}) F$$

$$= \sum_{y \in \prod_{i=1}^n (a_i,b_i]} \text{sgn}(y) \prod_{i=1}^n e_i^{y_i} F$$

$$= \sum_{y \in \prod_{i=1}^n (a_i,b_i]} \text{sgn}(y) F(y).$$

If $F$ happens to be sufficiently differentiable the fundamental theorem of calculus along with Eq. (24.55) implies

$$\mu((a,b]) = \left( \prod_{i=1}^n \int_{a_i}^{b_i} dx_i \partial_i \right) F$$

$$= \int_{[a,b]} (\partial_1 \ldots \partial_n F)(x) \, dx_1 \ldots dx_n.$$

From this we see that if $G$ is another function such that $\partial_1 \ldots \partial_n F = \partial_1 \ldots \partial_n G$ then

$$\Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} F = \Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} G.$$

As Nul $(\partial_1 \ldots \partial_n)$ 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 24.72. If $F : \mathbb{R}^n \to V$ is a function such that $\Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} F = 0$ 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 $\Delta_1 \Delta_2 F = 0$ 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,

$$F(x, y) = F(x, a_2) + C(a_2, y)$$

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

$$F(x, y) = f_\alpha(x) + g_\alpha(y)$$

for $y > \alpha$.

Now suppose that $\alpha < \beta$ so that

$$f_\alpha(x) + g_\alpha(y) = F(x, y) = f_\beta(x) + g_\beta(y)$$

for $y > \beta$.

i.e.

$$f_\alpha(x) - f_\beta(x) = g_\beta(y) - g_\alpha(y)$$

for $y > \beta$.

From this we see that $f_\alpha(x) - f_\beta(x) = g_\beta(y) - g_\alpha(y) = C$ for $y > \beta$, i.e.

$$f_\alpha(x) = f_\beta(x) + C \text{ and } g_\alpha(y) = g_\beta(y) - C$$

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 24.72 is that when $n \geq 2$, there is considerable ambiguity in the choice of $F$ that we may use so that Eq. (24.55) 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. (24.55) we have to be careful to have chosen a good representative of $F$ where we write $F \sim G$ iff $\Delta_1^{(a_1,b_1)} \ldots \Delta_n^{(a_n,b_n)} (F - G) = 0$ for all $a < b$.

Our next goal is to prove Theorem 24.74 which shows that Eq. (24.55) or equivalently Eq. (24.53) 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 24.73. Suppose that $A \subset 2^X$ is an algebra and for each $t \in \mathbb{R}$ let $\mathcal{B} \subset 2^X$ 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 $A \times \mathcal{B}$ such that

$$\mu(\{a, b\}) = \mu_b(A) - \mu_a(A) \forall (a, b) \in \mathcal{E} \text{ and } A \in \mathcal{A}.$$
\[
\mu((a, b]) = \sum_{\varepsilon \in \{0, 1\}^n} (-1)^{|\varepsilon|} F(y^\varepsilon(a, b))
\]

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 24.76.** Suppose that \(F : \mathbb{R}^n \to \mathbb{R}\) is a right continuous and satisfies
\[
\Delta_{(a_1, b_1)} \ldots \Delta_{(a_n, b_n)} F \geq 0
\]
for all \(a < b\) in \(\mathbb{R}^n\). Then there exists a unique measure \(\mu = \mu_F\) on \(\mathcal{B}\) such that Eq. (24.57) holds for all \(a < b\) in \(\mathbb{R}^n\). Moreover this measure is finite on compact sets.

**Proof.** Let \(\beta_k := (k, k, \ldots, k)\) for \(k \in \mathbb{N}\). Then by Theorem 24.75 there exists measure \(\mu_k\) on \(\mathcal{B}_{[\beta_k, \beta_k]}\) such that Eq. (24.57) holds for all \(a < b\) with \(a, b \notin (-\beta_k, \beta_k]\). 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 24.77 (Characterizing CDF’s).** If \(F : \mathbb{R}^n \to [0, 1]\) is the CDF of a probability measure \((\mu)\) on \((\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})\) \(\text{i.e.} F(x) = \mu(S_x)\) then,

1. \(F\) is right continuous,
2. \(F\) is increasing in the sense that \(F(a) \leq F(b)\) whenever \(a \leq b\),
3. \(\lim_{x \to -\infty} F(a_1, \ldots, a_{i-1}, x, a_i + \varepsilon, \ldots, a_n) = 0\) 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. \(\Delta_{(a_1, b_1)} \ldots \Delta_{(a_n, b_n)} F \geq 0\) 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 24.76. To see that \(\mu\) is the desired measure we compute,

\[
\mu(S_b) = \lim_{a \to -\infty} \mu((a, b]) = \lim_{a \to -\infty} \sum_{y \in \prod_{i=1}^n \{a_i, b_i\}} \text{sgn}(y) F(y) = F(b)
\]

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 24.78.** One might think that item 5 of Theorem 24.77 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 \(\Delta_{(1,1)} \Delta_{(2,2)} F = -(b_1 - a_1)(b_2 - a_2)\). This remark is not so surprising in light of Lemma 24.72.

**Lemma 24.79.** 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. \(x \in \mathcal{C}(F)\) \(-\text{the continuity points of} F\),
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. \(\lim_{n \to \infty} F(x - 1/n) = F(x)\) where \(1 := (1, \ldots, 1)\), 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 \(F(x) - \varepsilon \leq F(y_n) < F(a) < F(x) + \varepsilon\) which shows that \(F\) is continuous at \(x\).

**Exercise 24.9.** If \(F(x) = \mu(S_x)\) is a CDF then \(\mathcal{C}(F) = \{x \in \mathbb{R}^n : \mu(\text{bd}(S_x)) = 0\}\). 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 \(\text{bd}(S_x) \subset \cup_i \{\pi_i = x_i\}\) 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 24.80.** 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 24.81 (See Lemma 24.67).** 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,

\[
F(a) = \lim_{n \to \infty} F_n(a) \leq \lim_{n \to \infty} F_n(x) \leq \lim_{n \to \infty} F_n(b) = F(b) \leq F(x) + \varepsilon
\]

Letting \(a \uparrow x\) then implies,
Theorem 24.83. Suppose that 

\[ \int F(x) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x) + \varepsilon \]

and as \( \varepsilon > 0 \) was arbitrary, it follows that From this it follows that 

\[ F(x) \leq \limsup_{n \to \infty} F_n(x) \leq \liminf_{n \to \infty} F_n(x) = F(x). \]

Lemma 24.82 (Right Continuous Versions). 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 \( a < b \) 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

\[ 0 \leq G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon \quad \text{for all } b < y < \beta. \]

From this it follows that

\[ 0 \leq F(b) - F(a) \leq \inf_{y < \beta} G(y) - F(a) \leq G(\beta) - F(a) < \varepsilon \]

which proves the right continuity of \( F \).

Theorem 24.84 (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 24.61 provided we replace Lemma 24.59 by Lemma 24.82 and use a multi-variate version of Proposition 24.60. In a bit more detail. Choose \( \{m_k\} \) such that \( G(y) := \lim_{k \to \infty} F_{m_k}(y) \) exists for all \( y \in \mathbb{R}^n \). 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 \( y \in \mathbb{R}^n \). Notice that

\[ \Delta_{(a,b)} G := \lim_{k \to \infty} \Delta_{(a,b)} F_k = \lim_{k \to \infty} \mu_k((a,b)) \geq 0 \quad \text{for all } a < b \in \mathbb{R}^n. \]

We then define \( F(x) := \inf \{G(y) : y > x \} \) where \( G(y) \) is any sequence in \( \mathbb{R}^n \) with \( y_i \to x \) and \( y_i \to x \) as \( i \to \infty \). Therefore for \( a < b \in \mathbb{R}^n \) choose \( a_i < b_i \) with \( a_i, b_i \in \mathbb{Q}^n \) and \( a_i \uparrow a \) and \( b_i \uparrow b \), then

\[ \Delta_{(a,b)} F = \sum_{\varepsilon} (-1)^{|\varepsilon|} G(y^\varepsilon(a, b)) = \sum_{\varepsilon} (-1)^{|\varepsilon|} \lim_{l \to \infty} G(y^\varepsilon(a_l, b_l)) \]

\[ = \lim_{l \to \infty} \sum_{\varepsilon} (-1)^{|\varepsilon|} G(y^\varepsilon(a_l, b_l)) = \lim_{l \to \infty} \Delta_{(a_l, b_l)} G \geq 0. \quad (24.58) \]

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. (24.58) that \( \mu((\alpha, \beta]) = \Delta_{(\alpha, \beta]} F \geq 1 - \varepsilon \).

[Clean this up!] Finally I claim that \( F_n \to F \). Indeed, suppose that \( x \in C(F) \) and \( \varepsilon > 0 \) be given. Choose \( a, b \in \mathbb{R}^n \) such that \( \alpha < x < b \) and \( F(a) \leq G(b) \leq F(a) + \varepsilon \). 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

\[ F(x) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) = F(x). \]
24.10 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 \(B_S\) 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, \(A \subset S\)), then every open set may be written as a union of balls with \(x \in A\) and \(r \in \mathbb{Q}\) and so in the separable case

\[
B_S = \sigma \{ B(x, r) : x \in A, \ r \in \mathbb{Q} \} = \sigma \{ B(x, r) : x \in S, \ r \geq 0 \}.
\]

Let us now state the theorems of this section.

**Definition 24.85.** Let \((S, \tau)\) be a topological space, \(B := \sigma(\tau)\) be the Borel \(\sigma\)–algebra, and \(\mu\) be a probability measure on \((S, B)\). We say that \(A \subset B\) is a **continuity set** if \(\mu(B(a)) = 0\). Notice that this is equivalent to saying that \(\mu(A^c) = \mu(A)\).

**Theorem 24.86 (Skorohod Theorem).** Let \((S, \rho)\) be a separable metric space and \(\{\mu_n\}_{n=0}^\infty\) be probability measures on \((S, B_S)\) such that \(\lim_{n \to \infty} \mu_n(A) = \mu(A)\) for all \(A \subset B\) such that \(\mu(bd(A)) = 0\)\(^7\). Then there exists a probability space \((\Omega, \mathcal{B}, \mathbb{P})\) and measurable functions, \(Y_n : \Omega \to S\), such that \(\mu_n = \mathbb{P} \circ Y_n^{-1}\) for all \(n \in \mathbb{N} := \mathbb{N} \cup \{0\}\) and \(\lim_{n \to \infty} Y_n = Y\ a.s.

**Proposition 24.87 (The Portmanteau Theorem).** Suppose that \(S\) is a complete separable metric space and \(\mu\) is a probability measure on \((S, B)\). Then the following are equivalent:

1. \(\mu_n \Rightarrow \mu\) 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 \(F \subset S\).
4. \(\liminf_{n \to \infty} \mu_n(G) \geq \mu(G)\) for all \(G \subset S\).
5. \(\lim_{n \to \infty} \mu_n(A) = \mu(A)\) for all \(A \subset B\) such that \(\mu(bd(A)) = 0\).

**Definition 24.88.** Let \(S\) be a topological space. A collection of probability measures \(\lambda\) on \((S, B_S)\) is said to be **tight** if for every \(\varepsilon > 0\) there exists a compact subset \(K_\varepsilon \subset B_S\) such that \(\mu(K_\varepsilon) \geq 1 - \varepsilon\) for all \(\mu \in \Lambda\).

**Theorem 24.89 (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 \(B_S\). Then there exists a subsequence \(\{\mu_{n_k}\}_{k=1}^\infty\) which is weakly convergent to a probability measure \(\mu\) on \(B_S\).

Conversely, if we further assume that \((S, \rho)\) is a complete and \(\lambda\) is a sequentially compact subset of the probability measures on \((S, B_S)\) 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 \(B_{S_1}\) and \(B_{S_2}\) 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

\[
\rho((x_1, y_1), (x_2, y_2)) = \rho_1(x_1, y_1) \lor \rho_2(x_2, y_2)
\]

and let \(B_{S_1 \times S_2}\) 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

\[
B_{S_i} \otimes B_{S_2} = \sigma(\pi_1, \pi_2) = \sigma(\pi_1^{-1}(B_{S_1}) \cup \pi_2^{-1}(B_{S_2})).
\]

**Exercise 24.10 (Continuous Mapping Theorem).** Let \((S_1, \rho_1)\) and \((S_2, \rho_2)\) be separable metric spaces and \(B_{S_1}\) and \(B_{S_2}\) be the Borel \(\sigma\)–algebras on \(S_1\) and \(S_2\) respectively. Let further suppose that \(\{\mu_n\} \cup \{\mu\}\) are probability measures on \((S_1, B_{S_1})\) such that \(\mu_n \Rightarrow \mu\). If \(f : S_1 \to S_2\) is a Borel measurable function such that \(\mu(D(f)) = 0\) (see Notation 24.17), then \(f_*\mu_n \Rightarrow f_*\mu\) where \(f_*\mu := \mu \circ f^{-1}\).

**Exercise 24.11.** Prove the analogue of Lemma 9.25, namely show \(B_{S_1 \times S_2} = B_{S_1} \otimes B_{S_2}\). Hint: you may find Exercise 9.7 helpful.

**Exercise 24.12.** Let \((S_1, \rho_1)\) and \((S_2, \rho_2)\) be separable metric spaces and \(B_{S_1}\) and \(B_{S_2}\) 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 \((S_1, B_{S_1})\) and \((S_2, B_{S_2})\) respectively. If \(\mu_n \Rightarrow \mu\) and \(\nu_n \Rightarrow \nu\), then \(\mu_n \otimes \nu_n \Rightarrow \mu \otimes \nu\).

**Exercise 24.11** and **24.12** have obvious generalizations to finite product spaces. In particular, if \(\{X_n(i)\}_{n=0}^\infty\) are sequences of random variables for \(1 \leq i \leq K\) such that for each \(n\), \(\{X_n(i)\}_{i=1}^K\) are independent random variables with \(X_n(i) \Rightarrow X_0(i)\) as \(n \to \infty\) for each \(1 \leq i \leq K\), then

\[
\left( X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(K)} \right) \Rightarrow \left( X_0^{(1)}, X_0^{(2)}, \ldots, X_0^{(K)} \right) \text{ as } n \to \infty.
\]

These comments will be useful for Exercise 24.13 below.
Definition 24.90 (Convergence of finite dimensional distributions). Let \( \{X_n(t) : t \geq 0\}_{n=0}^{\infty} \) be a collection of random processes, \( X_n(t) : \Omega \rightarrow \mathbb{R} \). We say that \( X_n \) converges to \( X_0 \) in finite dimensional distributions and write \( X_n \xrightarrow{fd} X_0 \) provided for every finite subset \( \Lambda := \{0 = t_0 < t_1 < t_2 < \cdots < t_K\} \) of \( \mathbb{R}_+ \) we have

\[
(X_n(t_0), \ldots, X_n(t_K)) \xrightarrow{d} (X_0(t_0), \ldots, X_0(t_K)) \text{ as } n \to \infty.
\]

Exercise 24.13. Let \( \{X_n\}_{n=1}^{\infty} \) be an i.i.d. sequence of random variables with zero mean and \( \text{Var}(X_n) = 1 \). 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 \( S_m := \sum_{k=1}^{m} X_k \) where \( S_0 = 0 \) by definition. Show that \( W_n \xrightarrow{d} B \) where \( \{B(t) : t \geq 0\} \) is a Brownian motion as defined in definition 20.24. You might use the following outline.

1. For any \( 0 \leq s < t < \infty \), explain why \( W_n(t) - W_n(s) \xrightarrow{d} N(0, (t-s)) \).
2. Given \( A := \{0 = t_0 < t_1 < t_2 < \cdots < t_K\} \subset \mathbb{R}_+ \) argue that \( \{W_n(t_i) - W_n(t_{i-1})\}_{i=1}^{K} \) are independent and then show

\[
\{W_n(t_i) - W_n(t_{i-1})\}_{i=1}^{K} \xrightarrow{d} \{B(t_i) - B(t_{i-1})\}_{i=1}^{K} \text{ as } n \to \infty.
\]
3. Now show that \( \{W_n(t_i)\}_{i=1}^{K} \xrightarrow{d} \{B(t_i)\}_{i=1}^{K} \text{ as } n \to \infty. \)

The rest of this section is devoted to the proofs of the results stated at the beginning of the section. (These proofs may safely be skipped on first reading.)

24.10.1 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 \( A \subset S \), the closure of \( A \), is defined by

\[
\bar{A} := \bigcap \{C : A \subset C \subset S\} \quad \text{and} \quad A^c := \bigcup \{V : \tau \ni V \subset A\}
\]

and the interior of \( A \) is defined by

\[
A^o = \bigcup \{V : \tau \ni V \subset A\}.
\]

Thus \( \bar{A} \) is the smallest closed set containing \( A \) and \( A^o \) is the largest open set contained in \( A \). The relationship between the interior and closure operations is;

\[
(A^o)^c = \bigcap \{V^c : \tau \ni V \subset A\} = \bar{A}^c.
\]

Finally recall that the topological boundary of a set \( A \subset S \) is defined by

\[
\text{bd}(A) = \bar{A} \cap A^c = \bar{A} \cap \text{bd}(A) \quad (= \text{bd}(A^o)).
\]

In the case of a metric space we may describe \( \bar{A} \) and \( \text{bd}(A) \) as

\[
\bar{A} = \{x \in S : \exists \{x_n\} \subset A \ni x = \lim_{n \to \infty} x_n\} \quad \text{and} \quad \text{bd}(A) = \{x \in S : \exists \{x_n\} \subset A \ni \lim_{n \to \infty} y_n = x = \lim_{n \to \infty} x_n\}.
\]

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 24.91. For any subsets, \( A \) and \( B \), of \( S \) we have \( \text{bd}(A \cap B) \subset \text{bd}(A) \cup \text{bd}(B) \), \( \text{bd}(A \setminus B) \subset \text{bd}(A) \cup \text{bd}(B) \), and \( \text{bd}(A \cup B) \subset \text{bd}(A) \cup \text{bd}(B) \).

Proof. We begin by observing that \( A^o \cap B^o \subset A \cap B \subset \bar{A} \cap \bar{B} \) from which it follows that

\[
A^o \cap B^o \subset [A \cap B]^o \subset A \cap B \subset \bar{A} \cap \bar{B}
\]

and hence,

\[
\text{bd}(A \cap B) \subset [\bar{A} \cap \bar{B}] \setminus [A^o \cap B^o].
\]

Combining this inclusion with

\[
[\bar{A} \cap \bar{B}] \setminus [A^o \cap B^o] = [\bar{A} \cap \bar{B}] \cap [(A^o)^c \cup (B^o)^c] = [\bar{A} \cap \bar{B}] \cap [(A^c \cup (B^c))^c] = [\bar{A} \cap (A^c)^c] \cup [\bar{B} \cap (B^c)^c] \subset [\bar{A} \cap (A^c)^c] \cup [\bar{B} \cap (B^c)^c] = \text{bd}(A) \cup \text{bd}(B)
\]

completes the proof of the first assertion. The second and third assertions are easy consequence of the first because;

\[
\text{bd}(A \setminus B) = \text{bd}(A \cap B^c) \subset \text{bd}(A) \cup \text{bd}(B^c) = \text{bd}(A) \cup \text{bd}(B)
\]

and

\[
\text{bd}(A \cup B) = \text{bd}([A \cup B]^o) = \text{bd}(A^c \cap B^c) \subset \text{bd}(A^c) \cup \text{bd}(B^c) = \text{bd}(A) \cup \text{bd}(B).
\]

\[\blacksquare\]
24.10.2 Proof of Skorohod’s Theorem \[24.86\]

Lemma 24.92. Let \((S, \rho)\) be a separable metric space, \(B\) be the Borel \(\sigma\)-algebra on \(S\), and \(\mu\) be a probability measure on \(B\). Then for every \(\varepsilon > 0\) there exists a countable partition, \(\{B_n\}_{n=1}^\infty\), of \(S\) such that \(B_n \in B\), \(\text{diam}(B_n) \leq \varepsilon\) and \(B_n\) is a \(\mu\)-continuity set (i.e., \(\mu(\text{bd}(B_n)) = 0\)) for all \(n\).

Proof. For \(x \in S\) and \(r \geq 0\) let \(S(x,r) := \{y \in S : \rho(x,y) = r\}\). For any finite subset, \(\Gamma \subset [0, \infty)\), we have \(\sum_{r \in \Gamma} \mu(S(x,r)) \leq \mu(S) = 1\).

As \(\Gamma \subset \{f \in [0, \infty)\) was arbitrary we may conclude that \(\sum_{r \geq 0} \mu(S(x,r)) \leq 1 < \infty\) and therefore the set \(Q_x := \{r \geq 0 : \mu(S(x,r)) > 0\}\) is at most countable.

If \(B(x,r) := \{y \in S : \rho(x,y) < r\}\) and \(C(x,r) := \{y \in S : \rho(x,y) \leq r\}\) are the open and closed \(r\) balls about \(x\) respectively, we have \(S(x,r) = C(x,r) \setminus B(x,r)\).

Thus \(\text{bd}(B(x,r)) = \overline{B(x,r)} \setminus B(x,r) \subset C(x,r) \setminus B(x,r) = S(x,r),\) 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 \(Q := \bigcup_{n=1}^\infty Q_{x_n} - \) a countable subset of \([0, \infty)\). Choose \(r \in [0, \infty) \setminus Q\) such that \(r \leq \varepsilon/2\) and then define \(B_n := B(x_n,r) \setminus \bigcup_{k=0}^r B(x_n-k,r)\).

It is clear that \(\{B_n\}_{n=1}^\infty \subset B\) is a partition of \(S\) with \(\text{diam}(B_n) \leq 2r \leq \varepsilon\). Moreover, we know that

\[
\text{bd}(B_n) \subset \text{bd}(B(x_n,r)) \cup \text{bd}(B(x_1,r)) \cup \cdots \cup \text{bd}(B(x_{n-1},r)) \\
\subset \bigcup_{k=1}^n \text{bd}(B(x_k,r)) \subset \bigcup_{k=1}^n S(x_k,r)
\]

and therefore as \(r \notin Q\) we have

\[
\mu(\text{bd}(B_n)) \leq \sum_{k=1}^n \mu(S(x_k,r)) = 0
\]

so that \(B_n\) is a \(\mu\)-continuity set for each \(n \in \mathbb{N}\).

We are now ready to prove Skorohod’s Theorem \[24.86\].

Proof. (of Skorohod’s Theorem \[24.86\].) 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 \((\Omega_0, B_0, P_0)\) which is sufficiently rich so as to support the collection of independent random variables needed in the proof. The final probability space will then be given by \((\Omega, B, P) = (\Omega_0 \times S, B_0 \otimes B_2, P_0 \otimes \mu)\) 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 24.92 to construct a partition, \(\{B_n\}_{n=1}^\infty\), of \(S\) such that \(\text{diam}(B_n) < 2^{-p}\) and \(\mu(\text{bd}(B_n)) = 0\) for all \(n\). Choose \(m\) sufficiently large so that \(\mu(\bigcup_{n=m+1}^\infty B_n) < 2^{-p}\) and let \(B_0 := \bigcup_{n=m+1}^\infty B_n\) so that \(\{B_k\}_{k=0}^m\) is a partition of \(S\). Now define

\[
\kappa := \sum_{k=0}^m k 1_{B_k}(Y) = \sum_{k=0}^m k 1_{Y \in B_k}
\]

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 \[24.31\] implies there exists \(\tilde{\kappa}_n : (0, 1) \times \{0, \ldots, m\} \to \{0, \ldots, m\}\) such that \(\tilde{\kappa}_n(\Theta, k) = \tilde{k}\) when \(0 \leq \tilde{k} \leq m\) and \(\text{Law}_{\text{m}}(\mu(B_k))_{k=0}^m(\kappa_n) = \{\mu(B_k)\}_{k=0}^m\). Now let \(\kappa_n := \tilde{\kappa}_n(\Theta, k)\) so that \(P(\kappa_n = k) = \mu(B_k)\) for all \(n \in \mathbb{N}\) and \(0 \leq k \leq m\) and \(\kappa_n = k\) when \(\Theta \leq \mu(B_k)/\mu(B_k)\). Since \(\mu(\text{bd}(B_k)) = 0\) for all \(k\) it follows \(\mu(B_k) \to \mu(B_k)\) for all \(0 \leq k \leq m\) and therefore \(\lim_{n \to \infty} \kappa_n = \kappa\), \(P\)-a.s.

Now choose \(\xi^k_n\) independent of everything such that \(P(\xi^k_n \in A) = \mu_n(A|B_k)\) for all \(n\) and \(0 \leq k \leq m\). Then define

\[
Y^p_n := \xi^k_n(\Theta, k) \equiv \sum_{k=0}^m 1_{\kappa_n(\Theta, k) = k} \cdot \xi^k_n,
\]

Notice that

\[
P(Y^p_n \in A) = \sum_{k=0}^m P(\xi^k_n \in A \& \kappa_n(\Theta, k) = k) = \sum_{k=0}^m \mu_n(A) \mu_n(B_k) = \sum_{k=0}^m \mu_n(A \cap B_k) = \mu_n(A),
\]

and

\[
\{\rho(Y^p_n, Y) > 2^{-p}\} \subset \{Y \in B_0\} \cup \{\kappa \neq \kappa_n\}
\]

so that

\[
P(\bigcup_{n \geq N} \{\rho(Y^p,Y) > 2^{-p}\}) \leq P(Y \in B_0) + P(\bigcup_{n \geq N} \{\kappa_n \neq \kappa\}) < 2^{-p} + P(\bigcup_{n \geq N} \{\kappa_n \neq \kappa\}).
\]

An examination of the proof will show that \(\Omega_0\) can be taken to be \((0, 1) \times S^n\) equipped with a well-chosen infinite product measure.
Since $\kappa_n \to \kappa$ a.s. it follows that
\[
0 = P(\kappa_n \neq \kappa \text{ i.o. } n) = \lim_{N \to \infty} P(\cup_{n \geq N} \{\kappa_n \neq \kappa\})
\]
and so there exists $n_p < \infty$ such that
\[
P(\cup_{n \geq n_p} \{\rho(Y_n, Y) > 2^{-p}\}) < 2^{-p}.
\]
To finish the proof, construct $\{Y_n\}_{n=1}^\infty$ 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 $n_1 < n_2 < n_3 < \ldots$. As
\[
\sum_p P(\cup_{n \geq n_p} \{\rho(Y_n, Y) > 2^{-p}\}) < \sum_p 2^{-p} < \infty
\]
it follows from the first Borel Cantelli lemma that $P(N) = 0$ where
\[
N := \left\{ \left[ \cup_{n \geq n_p} \{\rho(Y_n, Y) > 2^{-p}\} \right] \text{ i.o. } p \right\}.
\]
So off the null set $N$ we have $\rho(Y_n, Y) \leq 2^{-p}$ for all $n \geq n_p$ and a.a. $p$. We now define $\{Y_n\}_{n=1}^\infty$ by
\[
Y_n := Y_n^p \text{ for } n \leq n < n_{p+1} \text{ and } p \in \mathbb{N}.
\]
Then by construction we have $\text{Law}(Y_n) = \mu_n$ for all $n$ and $\rho(Y_n, Y) \to 0$ a.s. \(\blacksquare\)

### 24.10.3 Proof of Proposition – The Portmanteau Theorem \[24.87\]

**Proof.** (of Proposition \[24.87\]) 1. \(\implies\) 2. is obvious.

For 2. \(\implies\) 3., let
\[
\varphi(t) := \begin{cases} 
1 & \text{if} \ t \leq 0 \\
1 - t & \text{if} \ 0 \leq t \leq 1 \\
0 & \text{if} \ t \geq 1
\end{cases}
\]
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
\[
1 - \liminf_{n \to \infty} \mu_n(G) = \limsup_{n \to \infty} (1 - \mu_n(G)) = \limsup_{n \to \infty} \mu_n(G^c) \leq \mu(G^c) = 1 - \mu(G)
\]
which implies 4. Similarly 4. \(\implies\) 3.

3. \(\iff\) 5. Recall that $\text{bd}(A) = \overline{A} \setminus A^\circ$, so if $\mu(\text{bd}(A)) = 0$ and 3. (and hence also 4.) holds we have
\[
\limsup_{n \to \infty} \mu_n(A) \leq \limsup_{n \to \infty} \mu_n(\overline{A}) = \mu(\overline{A}) = \mu(A) \text{ and }
\liminf_{n \to \infty} \mu_n(A^\circ) \geq \liminf_{n \to \infty} \mu_n(A^\circ) = \mu(A^\circ) = \mu(A)
\]
from which it follows that $\lim_{n \to \infty} \mu_n(A) = \mu(A)$. Conversely, let $F \subset S$ and set $F_\delta := \{x \in S : \rho(x, F) \leq \delta\}$. Then
\[
\text{bd}(F_\delta) \subset F_\delta \setminus \{x \in S : \rho(x, F) < \delta\} = A_\delta
\]
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 $A := \{\delta > 0 : \mu(A_\delta) > 0\}$ is at most countable. Let $\delta_n \notin A$ be chosen so that $\delta_n \downarrow 0$ as $n \to \infty$, then
\[
\mu(F_{\delta_n}) = \lim_{n \to \infty} \mu_n(F_{\delta_n}) \geq \limsup_{n \to \infty} \mu_n(F).
\]
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 \[24.86\] just as was done in the proof of Theorem \[24.32\]. 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
\[
\sum_{i=1}^k \frac{(i-1)}{k} \left\{ \frac{1}{n+1} \leq f < \frac{i}{n+1} \right\} \leq f \leq \sum_{i=1}^k \frac{i-1}{k} \left\{ \frac{1}{n+1} \leq f < \frac{i}{n+1} \right\},
\]
(24.60)
Let $F_i := \{ \frac{i}{k} \leq f \}$ and notice that $F_k = \emptyset$. Then for any probability $\mu$,
\[
\sum_{i=1}^k \frac{(i-1)}{k} [\mu(F_{i-1}) - \mu(F_i)] \leq \mu(f) \leq \sum_{i=1}^k \frac{i}{k} [\mu(F_{i-1}) - \mu(F_i)].
\]
(24.61)
Since
\[
\sum_{i=1}^{k} \frac{(i-1)}{k} [\mu(F_{i-1}) - \mu(F_i)]
= \sum_{i=1}^{k} \frac{(i-1)}{k} \mu(F_{i-1}) - \sum_{i=1}^{k} \frac{(i-1)}{k} \mu(F_i)
= \sum_{i=1}^{k-1} \frac{i}{k} \mu(F_i) - \sum_{i=1}^{k-1} \frac{i-1}{k} \mu(F_i) = \frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i)
\]
and
\[
\sum_{i=1}^{k} [\mu(F_{i-1}) - \mu(F_i)]
= \sum_{i=1}^{k} \frac{i-1}{k} [\mu(F_{i-1}) - \mu(F_i)] + \sum_{i=1}^{k} \frac{1}{k} [\mu(F_{i-1}) - \mu(F_i)]
= \sum_{i=1}^{k-1} \mu(F_i) + \frac{1}{k},
\]
Eq. (24.61) becomes,
\[
\frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i) \leq \mu(f) \leq \frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i) + 1/k.
\]
Using this equation with \(\mu = \mu_n\) and then with \(\mu = \mu\) we find
\[
\limsup_{n \to \infty} \mu_n(f) \leq \limsup_{n \to \infty} \left[ \frac{1}{k} \sum_{i=1}^{k-1} \mu_n(F_i) + 1/k \right]
\leq \frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i) + 1/k \leq \mu(f) + 1/k.
\]
Since \(k\) is arbitrary, \(\limsup_{n \to \infty} \mu_n(f) \leq \mu(f)\). Replacing \(f\) by \(1-f\) in this inequality also gives \(\liminf_{n \to \infty} \mu_n(f) \geq \mu(f)\) and hence we have shown \(\lim_{n \to \infty} \mu_n(f) = \mu(f)\) as claimed.

### Proof of Prokhorov’s Compactness Theorem

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


**Proof.** (of Prokhorov’s compactness Theorem) First suppose that \(S\) is compact. In this case \(C(S)\) is a Banach space which is separable by the Stone – Weierstrass 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 \((S,B_S)\). 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 \(\{\mu_k\}_{k=1}^{\infty}\) 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 \(\{\mu_k\}_{k=1}^{\infty}\) such that \(A(f) := \lim_{k \to \infty} \mu_k(f)\) exists for all \(f \in \Gamma\). Then for \(g \in C(S)\) and \(f \in \Gamma\), we have
\[
|\mu_k(g) - \mu_n(g)| \leq |\mu_k(g) - \mu_k(f)| + |\mu_k(f) - \mu_n(f)|
+ |\mu_n(f) - \mu_n(g)|
\leq 2 \|g - f\|_\infty + |\mu_n(f) - \mu_n(g)|
\]
which shows
\[
\limsup_{n \to \infty} |\mu_n(g) - \mu_n(g)| \leq 2 \|g - f\|_\infty .
\]

Letting \(f \in A\) tend to \(g\) in \(C(S)\) shows \(\limsup_{n \to \infty} |\mu_n(g) - \mu_n(g)| = 0\) and hence \(A(g) := \lim_{k \to \infty} \mu_k(g)\) for all \(g \in C(S)\). It is now clear that \(A(g) \geq 0\) for all \(g \geq 0\) so that \(A\) is a positive linear functional on \(S\) and thus there is a probability measure \(\mu\) such that \(A(g) = \mu(g)\).

**General case.** By Theorem 24.77 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 \(\tilde{\mu}_n(A) := \mu_n(A \cap S)\) for all \(A \in B_S\). By what we have just proved, there is a subsequence \(\{\tilde{\mu}_n\}_{k=1}^{\infty}\) such that \(\tilde{\mu}_k\) 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 \(\tilde{\mu}_n(K_\varepsilon) \geq 1 - \varepsilon\) for all \(n\). Since \(K_\varepsilon\) is compact in \(\bar{S}\) it is compact in \(\bar{S}\) as well and in particular a closed subset of \(\bar{S}\). Therefore by Proposition 24.87
\[
\tilde{\mu}(K_\varepsilon) \geq \limsup_{k \to \infty} \tilde{\mu}_k(K_\varepsilon) = 1 - \varepsilon.
\]
Since \(\varepsilon > 0\) is arbitrary, this shows with \(S_0 := \cup_{n=1}^{\infty} K_{1/n}\) satisfies \(\tilde{\mu}(S_0) = 1\). Because \(S_0 \subset B_S\), we may view \(\tilde{\mu}\) as a measure on \(\bar{S}\) by letting \(\mu(A) := \tilde{\mu}(A \cap S_0)\) for all \(A \in B_S\). Given a closed subset \(F \subset S\), choose \(\bar{F} \subset \bar{S}\) such that \(F = \bar{F} \cap S\). Then
\[
\limsup_{k \to \infty} \tilde{\mu}_k(F) = \limsup_{k \to \infty} \tilde{\mu}_k(\bar{F}) \leq \tilde{\mu}(\bar{F} \cap S_0) = \mu(F),
\]
which shows \(\tilde{\mu}_k \Rightarrow \mu\).
Converse direction. Suppose now that \((S, \rho)\) is complete and \(A\) is a sequentially compact subset of the probability measures on \((S, B_S)\). 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

\[
c := \sup_n \inf_{\mu \in A} \mu(G_n) = \lim_{n \to \infty} \inf_{\mu \in A} \mu(G_n) = 1.
\]

Suppose for sake of contradiction that \(c < 1\) and let \(c' \in (c, 1)\). By our assumption we have \(\inf_{\mu \in A} \mu(G_n) \leq c\) for all \(n\) therefore there exists \(\mu_n \in A\) such that \(\mu_n(G_n) \leq c'\) 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 \(\nu_n := \mu_k \Rightarrow \mu\) for some probability measure \(\mu\) on \(S\) and \(\nu_n(G'_n) \leq c'\) for all \(n\) where \(G'_n \uparrow S\) as \(n \uparrow \infty\). For fixed \(N \in \mathbb{N}\) we have \(\nu_n(G'_N) \leq \nu_n(G'_n) \leq c'\) for \(n \geq N\). Passing to the limit as \(n \to \infty\) in these inequalities then implies

\[
\mu(G'_N) \leq \liminf_{n \to \infty} \nu_n(G'_n) \leq c' < 1.
\]

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 \(G_n := \bigcup_{k=1}^m B(x_k, \frac{1}{m}) \uparrow S\) and so by the above claim there exists \(n_m\) such \(V_m := G_{n_m}\) satisfies \(\inf_k \mu_k(V_m) \geq 1 - \varepsilon 2^{-m}\). We now let \(A := \cap_m V_m\) so that \(\mu_k(A) \geq 1 - \varepsilon\) for all \(k\). As \(A\) is totally bounded and \(S\) is complete, \(K_\varepsilon := A\) is the desired compact subset of \(S\) such that \(\mu_k(K_\varepsilon) \geq 1 - \varepsilon\) for all \(k\).
Characteristic Functions (Fourier Transform)

**Notation 25.1** 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 11.12 here.

**Definition 25.2.** 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 \( f (\lambda) := f_X (\lambda) := \mathbb{E} [e^{i\lambda \cdot X}] \). Of course, if \( \mu := P \circ X^{-1} \), then \( f_X (\lambda) = \hat{\mu} (\lambda) \).

From Corollary 11.13 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.

### 25.1 Basic Properties of the Characteristic Function

**Definition 25.3.** A function \( f : \mathbb{R}^n \to \mathbb{C} \) is said to be positive definite, iff \( f (-\lambda) = f (\lambda) \) 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, \( \{f (\lambda_j - \lambda_k)\}_{j,k=1}^m \) is non-negative. More explicitly we require,

\[
\sum_{j,k=1}^m f (\lambda_j - \lambda_k) \xi_j \xi_k \geq 0 \text{ for all } (\xi_1, \ldots, \xi_m) \in \mathbb{C}^m.
\]

**Notation 25.4** 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 all \( k = 1, \ldots, l \) and all \( j_1, \ldots, j_k \in \{1, 2, \ldots, n\} \).

**Proposition 25.5 (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. \( \hat{\mu} (\lambda) = \hat{\mu} (-\lambda) \) for all \( \lambda \in \mathbb{R}^n \) and in particular, \( \hat{\mu} \) is real valued iff \( \mu \) is symmetric, i.e. iff \( \mu (\lambda) = \mu (A) \) 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 25.46 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 \( \int_{\mathbb{R}^n} \|x\|^l \, d\nu (x) < \infty \), 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

\[
\hat{\mu} \star \hat{\nu} (\lambda) = \hat{\mu} (\lambda) \hat{\nu} (\lambda) \text{ for all } \lambda \in \mathbb{R}^n.
\]

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 \( \hat{\mu} (\lambda) = \hat{\mu} (-\lambda) \) 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 11.13 then implies \( \mu = \nu \), i.e. \( \mu \) is symmetric. This proves item 3.
Item 5. follows by induction using Corollary \[\text{10.30}\] For item 4. let \( m \in \mathbb{N}, \) \( \{\lambda_j\}_{j=1}^m \subseteq \mathbb{R}^n \) and \((\xi_1, \ldots, \xi_m) \in \mathbb{C}^m.\) Then
\[
\begin{align*}
\sum_{j,k=1}^m \hat{\mu}(\lambda_j - \lambda_k) \xi_j \xi_k &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i(\lambda_j - \lambda_k) \cdot x} \xi_j \xi_k d\mu(x) \\
&= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i\lambda_j \cdot x} \xi_j e^{i\lambda_k \cdot x} d\mu(x) \\
&= \int_{\mathbb{R}^n} \left| \sum_{j=1}^m e^{i\lambda_j \cdot x} \right|^2 d\mu(x) = 0.
\end{align*}
\]

Example 25.6 (Example \[\text{24.3 continued}\].) Let \( d\mu(x) = 1_{[0,1]}(x) dx \) and \( \nu(A) = \mu(-A). \) Then
\[
\hat{\mu}(\lambda) = \int_0^1 e^{i\lambda x} dx = \frac{e^{i\lambda} - 1}{i\lambda},
\]
\[
\hat{\nu}(\lambda) = \hat{\mu}(-\lambda) = \frac{e^{-i\lambda} - 1}{-i\lambda},
\]
and
\[
\hat{\mu} \ast \hat{\nu}(\lambda) = \hat{\mu}(\lambda) \hat{\nu}(\lambda) = |\hat{\mu}(\lambda)|^2 = \frac{2}{\lambda^2} [1 - \cos \lambda]. \quad (25.1)
\]

According to Example \[\text{24.3}\] we also have \( d(\mu \ast \nu)(x) = (1 - |x|)_+ dx \) and so we may directly verify Eq. \[\text{(25.1)}\] as follows;
\[
\hat{\mu} \ast \hat{\nu}(\lambda) = \int \lambda (1 - |x|)_+ dx = \int \lambda (\cos(\lambda x) (1 - |x|)_+ dx \\
= 2 \int_0^1 \lambda (1 - x) \cos \lambda x dx \\
= 2 \int_0^1 (1 - x) \sin \lambda x dx \\
= -2 \int_0^1 d(1 - x) \sin \lambda x dx = 2 \int_0^1 \sin \lambda x dx \\
= 2 \frac{1 - \cos \lambda}{\lambda^2}.
\]

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 \[\text{25.5}\].

**Lemma 25.7.** 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 \( \mu(\lambda) := \mathbb{E}[e^{i\lambda X}] \) is \( C^n \) differentiable and
\[
\left( \frac{d\mu}{d\lambda} \right)^{(l)}(\lambda) = \mathbb{E}[(ix)^l e^{i\lambda X}] = \int_{\mathbb{R}} (ix)^l e^{i\lambda x} d\mu(x) \text{ for } l = 0, 1, 2, \ldots, n.
\]

In particular it follows that
\[
\mathbb{E}[(X^l)] = \frac{\hat{\mu}^{(l)}(0)}{l!}.
\]

The following theorem is a partial converse to this lemma. Hence the combination of Lemma \[\text{25.7, Theorem 25.8, and Corollary 25.19}\] (see also Corollary \[\text{25.38}\] below) shows that there is a correspondence between the number of moments of \( X \) and the differentiability of \( f_X. \)

**Theorem 25.8 (Smoothness implies integrability I).** Let \( X \) be a random variable, \( m \in \{0, 1, 2, \ldots, \}, \) \( f(\lambda) = \mathbb{E}[e^{i\lambda X}]. \) 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 \( \mathbb{E}[(X^{2m+2})] < \infty \) 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
\[
\hat{u}(\lambda) = \mathbb{E}[X^{2m} \cos(\lambda X)] = \text{Re} \mathbb{E}[X^{2m} e^{i\lambda X}]
\]
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
\[
\frac{u(0) - u(\lambda)}{\lambda c_\lambda} = \frac{u'(c_\lambda) - u'(0)}{c_\lambda} \to -u''(0) \text{ as } \lambda \downarrow 0. \quad (25.2)
\]
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,
\[
\mathbb{E}[(X^{2m} \frac{1 - \cos(\lambda X)}{\lambda^2})] \leq \mathbb{E}[(X^{2m} \frac{1 - \cos(\lambda X)}{\lambda c_\lambda})] = \frac{u(0) - u(\lambda)}{\lambda c_\lambda},
\]
to find
\[
\frac{1}{2} \mathbb{E}[(X^{2m+2})] \leq \lim_{\lambda \downarrow 0} \frac{u(0) - u(\lambda)}{\lambda c_\lambda} = -u''(0) < \infty,
\]
where the last equality is a consequence of Eq. \[\text{(25.2)}\].
With this result in hand, the theorem is now easily proved by induction. We start with \( m = 0 \) and recall from Proposition \[25.5\] that \( f \in C^1 (\mathbb{R}, \mathbb{C}) \). Assuming \( f \) is differentiable in a neighborhood of \( 0 \) and \( f'' (0) \) exists we may apply the above result with \( u = \text{Re} \, f \) in order to learn \( \mathbb{E} [X^2] < \infty \). An application of Lemma \[25.7\] then implies that \( f \in C^2 (\mathbb{R}, \mathbb{C}) \). The induction step is handled in much the same way upon noting,

\[
f^{(2m)} (\lambda) = (-1)^m \mathbb{E} [X^{2m} e^{i\lambda X}]
\]

so that

\[
u (\lambda) := (-1)^m \text{Re} \, f^{(2m)} (\lambda) = \mathbb{E} [X^{2m} \cos (\lambda X)].
\]

**Corollary 25.9.** Suppose that \( X \) is a \( \mathbb{R}^d \) – valued random vector such that for all \( \lambda \in \mathbb{R}^d \) the function

\[
f_\lambda (t) := f_X (t\lambda) = \mathbb{E} [e^{it\lambda \cdot X}]
\]

is \( 2m \) – times differentiable in a neighborhood of \( t = 0 \), then \( \mathbb{E} \|X\|^{2m} < \infty \) and \( f_X \in C^{2m} (\mathbb{R}^d, \mathbb{C}) \).

**Proof.** Applying Theorem \[25.8\] with \( X \) replaced by \( \lambda \cdot X \) shows that \( \mathbb{E} [\|\lambda \cdot X\|^{2m}] < \infty \) for all \( \lambda \in \mathbb{R}^d \). In particular, taking \( \lambda = e_i \) (the \( i^{\text{th}} \) standard basis vector) implies that \( \mathbb{E} [\|X_i\|^{2m}] < \infty \) for \( 1 \leq i \leq d \). So by Minkowski’s inequality;

\[
\|X\|_{L^{2m}(P)} = \left\| \sum_{i=1}^d X_i e_i \right\|_{L^{2m}(P)} \leq \left\| \sum_{i=1}^d |X_i| \right\|_{L^2(P)} \leq \sum_{i=1}^d \|X_i\|_{L^{2m}(P)} < \infty,
\]

i.e. \( \mathbb{E} \|X\|^{2m} < \infty \). The fact that \( f_X \in C^{2m} (\mathbb{R}^d, \mathbb{C}) \) now follows from Proposition \[25.5\].

### 25.2 Examples

**Example 25.10.** If \( -\infty < a < b < \infty \) and \( d\mu (x) = \frac{1}{b-a} 1_{[a,b]} (x) \, dx \) then

\[
\hat{\mu} (\lambda) = \frac{1}{b-a} \int_a^b e^{i\lambda x} \, dx = \frac{e^{ib\lambda} - e^{ia\lambda}}{i\lambda (b-a)}.
\]

If \( a = -c \) and \( b = c \) with \( c > 0 \), then

\[
\hat{\mu} (\lambda) = \frac{\sin \lambda c}{\lambda c}.
\]

Observe that

\[
\hat{\mu}' (\lambda) = 1 - \frac{1}{3!} \lambda^2 c^2 + \ldots
\]

and therefore, \( \hat{\mu}' (0) = 0 \) and \( \hat{\mu}'' (0) = -\frac{1}{3} c^2 \) and hence it follows that

\[
\int_{\mathbb{R}} x d\mu (x) = 0 \quad \text{and} \quad \int_{\mathbb{R}} x^2 d\mu (x) = \frac{1}{3} c^2.
\]

**Example 25.11.** Suppose \( Z \) is a Poisson random variable with mean \( a > 0 \), i.e. \( P (Z = n) = e^{-a} \frac{a^n}{n!} \). Then

\[
f_Z (\lambda) = \mathbb{E} [e^{i\lambda Z}] = e^{-a} \sum_{n=0}^\infty e^{i\lambda n} \frac{a^n}{n!} = e^{-a} \sum_{n=0}^\infty \left( \frac{ae^{i\lambda}}{n!} \right)^n = \exp \left( a (e^{i\lambda} - 1) \right).
\]

Differentiating this result gives,

\[
f'_Z (\lambda) = iae^{i\lambda} \exp \left( a (e^{i\lambda} - 1) \right) \quad \text{and} \quad f''_Z (\lambda) = (-a^2 e^{2i\lambda} - ae^{i\lambda}) \exp \left( a (e^{i\lambda} - 1) \right),
\]

from which we conclude,

\[
\mathbb{E} Z = \frac{1}{i} f'_Z (0) = a \quad \text{and} \quad \mathbb{E} Z^2 = -f''_Z (0) = a^2 + a.
\]

Therefore, \( \mathbb{E} Z = a = \text{Var} (Z) \).

**Example 25.12.** Suppose \( T \overset{d}{=} \exp (a) \), i.e. \( T \geq 0 \) a.s. and \( P (T \geq t) = e^{-at} \) for all \( t \geq 0 \). Recall that \( \mu = \text{Law} (T) \) is given by

\[
d\mu (t) = F'_T (t) \, dt = ae^{-at} 1_{t \geq 0} \, dt.
\]

Therefore,

\[
\mathbb{E} [e^{iaT}] = \int_0^\infty ae^{-at} e^{ia \lambda} \, dt = \frac{a}{a - i\lambda} = \hat{\mu} (\lambda).
\]

Since

\[
\hat{\mu}' (\lambda) = i \frac{a}{(a - i\lambda)^2} \quad \text{and} \quad \hat{\mu}'' (\lambda) = -2 \frac{a}{(a - i\lambda)^3}
\]

it follows that

\[
\mathbb{E} T = \frac{\hat{\mu}' (0)}{i} = a^{-1} \quad \text{and} \quad \mathbb{E} T^2 = \frac{\hat{\mu}'' (0)}{i^2} = \frac{2}{a^2}
\]

and hence \( \text{Var} (T) = \frac{2}{a^2} - \left( \frac{1}{a} \right)^2 = a^{-2} \).
Example 25.13. From Exercise [10.15] 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
\[
\int_{\mathbb{R}} x d\mu (x) = 0 \quad \text{and} \quad \int_{\mathbb{R}} x^2 d\mu (x) = 1.
\]

Recall from Section [12.9] that we have defined a random vector, \( X \in \mathbb{R}^d \), to be Gaussian iff
\[
\mathbb{E} \left[ e^{i\lambda \cdot X} \right] = \exp \left( -\frac{1}{2} \text{Var} (\lambda \cdot X) + i\lambda \cdot m \right)
\]
where \( m := \int_{\mathbb{R}^d} x d\mu (x) \) and \( q (\lambda, \lambda) := \int_{\mathbb{R}^d} (\lambda \cdot x)^2 d\mu (x) - (\lambda \cdot m)^2 \).

Example 25.14. 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
\[
\mu^n := \hat{\mu} \ast \cdots \ast \hat{\mu}.
\]
(25.3)
Alternatively, put, if \( \{x_k\}_{k=1}^n \) are i.i.d. random variables with \( \mu = P \circ X_k^{-1} \), then
\[
f_{X_1, \ldots, X_n} (\lambda) = f^n_{X_1} (\lambda).
\]

Example 25.15. 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,
\[
\mu := \sum_{n=0}^\infty p_n \hat{\mu}_n.
\]
Here is a more interesting interpretation of \( \mu \). Let \( \{X_n\}_{n=0}^\infty \cup \{T\} \) be independent random variables with \( P \circ X^{-1} = \mu_n \) and \( P (T = n) = p_n \) for all \( n \in \mathbb{N} \). Then \( \mu (A) = P (X_T \in A) \), where \( X_T (\omega) := X_T (\omega) \). 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 25.16. 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, \( \nu = \sum_{n=0}^\infty p_n \hat{\mu}_n \) where \( \hat{\mu}_n \) is defined in Eq. (25.3). 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,
\[
f_{X_T} (\lambda) = \mathbb{E} \left[ e^{i\lambda X_T} \right] = \exp \left( a (f_{X_1} (\lambda) - 1) \right).
\]

25.3 Tail Estimates

Lemma 25.17 (Tail Estimate). Let \( X : (\Omega, \mathcal{B}, P) \rightarrow \mathbb{R} \) be a random variable and \( f_{X_\lambda} (\lambda) := \mathbb{E} \left[ e^{i\lambda X} \right] \) be its characteristic function. Then for \( a > 0 \),
\[
P (|X| \geq a) \leq \frac{a}{2} \int_{-2/a}^{2/a} (1 - f_{X_\lambda} (\lambda)) d\lambda = \frac{a}{2} \int_{-2/a}^{2/a} (1 - \text{Re} f_{X_\lambda} (\lambda)) d\lambda = 2 \int_{-1}^1 \left( 1 - \text{Re} f_{X_\lambda} \right)
\]
(25.4)

Proof. Recall that the Fourier transform of the uniform distribution on \([-c, c] \) is \( \frac{\sin \lambda c}{\lambda c} \) and hence
\[
\frac{1}{2c} \int_{-c}^c f_{X_\lambda} (\lambda) d\lambda = \frac{1}{2c} \int_{-c}^c \mathbb{E} \left[ e^{i\lambda X} \right] d\lambda = \mathbb{E} \left[ \frac{\sin cX}{cX} \right].
\]

Therefore,
\[
\frac{1}{2c} \int_{-c}^{c} (1 - f_X(\lambda)) \, d\lambda = 1 - \mathbb{E} \left[ \frac{\sin cX}{cX} \right] = \mathbb{E}[Y_c]
\] (25.5)
where
\[
Y_c := 1 - \frac{\sin cX}{cX}.
\]

Notice that \( Y_c \geq 0 \) (see Eq. (25.51)) and moreover, \( Y_c \geq 1/2 \) if \( |cX| \geq 2 \left( |\sin cX| / |cX| \leq |\sin cX| / 2 \leq 1/2 \text{ if } |cX| \geq 2 \right) \). 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. (25.5) 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. (25.4). \( \blacksquare \)

**Remark 25.18.** 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,
\[
\int_{\mathbb{R}^d} \varphi(k) \, \hat{\mu}(k) \, dk = \int_{\mathbb{R}^d} dk \varphi(k) \int_{\mathbb{R}^d} d\mu(x) \, e^{ik \cdot x} = \int_{\mathbb{R}^d} d\mu(x) \int_{\mathbb{R}^d} dk \varphi(k) \, e^{ik \cdot x} = \int_{\mathbb{R}^d} \hat{\varphi}(x) \, d\mu(x)
\]
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
\[
\varphi(x) = \delta_x(x) := \frac{1}{\varepsilon^d} \delta \left( \frac{x}{\varepsilon} \right)
\]
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
\[
\hat{\delta}_\varepsilon(k) = \int_{\mathbb{R}^d} \delta_x(x) \, e^{ik \cdot x} \, dx = \int_{\mathbb{R}^d} \frac{1}{\varepsilon^d} \delta \left( \frac{x}{\varepsilon} \right) \, e^{ik \cdot x} \, dx = \int_{\mathbb{R}^d} \delta(y) \, e^{ik \cdot \varepsilon y} \, dy = \hat{\delta}(\varepsilon k).
\]
Thus the duality relation now becomes,
\[
\int_{\mathbb{R}^d} \hat{\delta}(\varepsilon k) \, d\mu(k) = \int_{\mathbb{R}^d} \delta_x(k) \, \hat{\mu}(k) \, dk = \int_{\mathbb{R}^d} \delta(k) \, \hat{\mu}(\varepsilon k) \, dk.
\]
Subtracting this equation from the simple identity,
\[
\int_{\mathbb{R}^d} 1 \, d\mu(k) = \mu(\mathbb{R}^d) = 1 = \int_{\mathbb{R}^d} \delta(k) \, dk,
\]
then shows,
\[
\int_{\mathbb{R}^d} \left[ 1 - \delta(\varepsilon k) \right] \, d\mu(k) = \int_{\mathbb{R}^d} \left[ 1 - \hat{\mu}(\varepsilon k) \right] \delta(k) \, dk
\]
and by taking the real part of the identity,
\[
\int_{\mathbb{R}^d} \left[ 1 - \Re \delta(\varepsilon k) \right] \, d\mu(k) = \int_{\mathbb{R}^d} \left[ 1 - \Re \hat{\mu}(\varepsilon k) \right] \delta(k) \, dk.
\]
Since (by the Riemann Lebesgue lemma) \( \lim_{|k| \to \infty} \hat{\delta}(k) = 0 \), there exists \( M = M(\delta) < \infty \) such that \( |\hat{\delta}(k)| \leq \frac{1}{2} \) if \( |k| \geq M \) and hence
\[
1 - \Re \delta(\varepsilon k) \geq \frac{1}{2} \text{ if } |\varepsilon k| \geq M, \text{ i.e. if } |k| \geq \frac{M}{\varepsilon}.
\]
Since \( |\hat{\delta}(k)| \leq 1 \) for all \( k \) it also follows that \( 1 - \Re \hat{\delta}(\varepsilon k) \geq 0 \) for all \( k \). From these remarks we may conclude that
\[
\int_{\mathbb{R}^d} \left[ 1 - \Re \delta(\varepsilon k) \right] \, d\mu(k) \geq \int_{\mathbb{R}^d} \left[ 1 - \Re \hat{\delta}(\varepsilon k) \right] \, 1_{|k| \geq M/\varepsilon} \, d\mu(k) \geq \frac{1}{2} \mu \left( |k| \geq \frac{M}{\varepsilon} \right).
\]
Thus we arrive at the tail estimate
\[
\mu \left( |k| \geq \frac{M}{\varepsilon} \right) \leq 2 \int_{\mathbb{R}^d} \left[ 1 - \Re \hat{\mu}(\varepsilon k) \right] \delta(k) \, dk = 2 \int_{\mathbb{R}^d} \left[ 1 - \Re \hat{\mu}(k) \right] \delta_x(k) \, dk
\]
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
\[ \mu \left( |k| \geq \frac{\sqrt{2 \ln 2}}{\varepsilon} \right) \leq 2 \int_{\mathbb{R}^d} \left[ 1 - \text{Re} \hat{\mu}(\varepsilon k) \right] p_1(k) \, dk \]
\[ = 2 \int_{\mathbb{R}^d} \left[ 1 - \text{Re} \hat{\mu}(k) \right] p_{2\ln 2 - 2}(k) \, dk. \]

or by setting \( a := \frac{\sqrt{2 \ln 2}}{\varepsilon} \) so that \( \varepsilon = \frac{\sqrt{2 \ln 2}}{a} \) we find,
\[ \mu \left( |k| \geq a \right) \leq 2 \int_{\mathbb{R}^d} \left[ 1 - \text{Re} \left( \frac{\sqrt{2 \ln 2}}{a} k \right) \right] p_1(k) \, dk \]
\[ = 2 \int_{\mathbb{R}^d} \left[ 1 - \text{Re} \hat{\mu}(k) \right] p_{2\ln 2 - 2}(k) \, dk. \]

Let us further note that if \( \hat{\mu}(k) \) is continuously differentiable at 0, then \( |\hat{\mu}(k)| \leq C (|k| + 1) \) and we conclude that
\[ \mu \left( |k| \geq \frac{M}{\varepsilon} \right) \leq 2C \int_{\mathbb{R}^d} |\varepsilon k| \geq 1 \cdot \delta(k) \, dk \]
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,
\[ \mu \left( |k| \geq a \right) \leq \frac{K}{a} \text{ for some } K < \infty. \]

Exercise 25.1. Suppose now \( X : (\Omega, \mathcal{F}, P) \to \mathbb{R}^d \) is a random vector and \( f_X(\lambda) := \mathbb{E}[e^{i\lambda \cdot X}] \) is its characteristic function. Show for \( a > 0 \),
\[ P \left( |X|_{\infty} \geq a \right) \leq 2 \left( \frac{a}{4} \right)^d \int_{[-a/2,a/2]^d} (1 - f_X(\lambda)) \, d\lambda \]
\[ = 2 \left( \frac{a}{4} \right)^d \int_{[-a/2,a/2]^d} (1 - \text{Re} f_X(\lambda)) \, d\lambda \] (25.6)
where \( |X|_{\infty} = \max_i |X_i| \) and \( d\lambda = d\lambda_1, \ldots, d\lambda_d. \)

Exercise 25.2 (Smoothness implies integrability II). Keeping the notation in Lemma 25.17 and letting \( u(\lambda) = \text{Re} f_X(\lambda) = \mathbb{E}\cos(\lambda X) \). 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
\[ P \left( |X| \geq a \right) \leq K(C, \delta) \left( \frac{1}{a} \right)^{1+\delta} \text{ for all } a > 2/\varepsilon. \] (25.7)

Use this estimate to show \( \mathbb{E}|X| < \infty. \)

Corollary 25.19 (Smoothness implies integrability III). Suppose that \( X \) is a random variable such that \( u(\lambda) = \text{Re} f_X(\lambda) = \mathbb{E}\cos(\lambda X) \) is a \( C^{2m+1}(\mathbb{R}) \) function of \( \lambda \). If there exists \( \varepsilon, \delta, C > 0 \) so that
\[ |u^{(2m+1)}(\lambda)| \leq C |\lambda|^\delta \text{ for all } |\lambda| < \varepsilon, \]
then \( \mathbb{E}|X|^{2m+1} < \infty. \)

Proof. By Theorem 25.8, we already know \( \mathbb{E}|X|^{2m} < \infty \) and therefore,
\[ u^{(2m)}(\lambda) = (-1)^m \frac{\mathbb{E}[X^{2m} \cos \lambda X]}{\mathbb{E}[|X|^{2m}]} \]
Now let \( Y \) be a random variable whose distribution is determined by
\[ \mathbb{E} h(Y) = \frac{\mathbb{E}[X^{2m} h(X)]}{\mathbb{E}[|X|^{2m}]} \; \forall \text{ bounded } h. \]

Then
\[ u_Y(\lambda) := \text{Re} f_Y(\lambda) = \mathbb{E}[\cos \lambda Y] = \frac{\mathbb{E}[X^{2m} \cos \lambda X]}{\mathbb{E}[X^{2m}]} = (-1)^m \frac{u^{(2m)}(\lambda)}{\mathbb{E}[X^{2m}]} \]
and so we see that \( u_Y(\lambda) \) satisfies the assumptions of Exercise 25.2. This allows us to conclude that \( \mathbb{E}|Y| < \infty \) which completes the proof since,
\[ \mathbb{E}|Y| = \frac{\mathbb{E}[X^{2m} |X|]}{\mathbb{E}[X^{2m}]} = \mathbb{E}|X|^{2m+1}. \]

\[ \square \]

25.4 Continuity Theorem

Theorem 25.20 (Continuity Theorem). Suppose that \( \{\mu_n\}_{n=1}^\infty \) is a sequence of probability measure on \( (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \) and suppose that \( f(\lambda) := \lim_{n \to \infty} \mu_n(\lambda) \) 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 \( \mu_n \Rightarrow \mu \) as \( n \to \infty. \)
The condition of a function being positive definite is preserved under taking \( f, \) \( g, \) \( \lim\) \( \lambda \) \( \rightarrow\infty \), and continuous function, it follows that \( \mu \) \( \lambda \) \( \rightarrow\infty \) again, there is a further subsequence, \( \mu_k'' := \mu_k' \) such that \( \mu_k'' \Rightarrow \nu \) for some probability measure \( \nu \). Since \( \hat{\mu} (\lambda) = \lim_{n \rightarrow\infty} \mu_n'' (\lambda) = f (\lambda) = \hat{\mu} (\lambda) \), it follows that \( \mu = \nu \). This leads to a contradiction since, 
\[
\varepsilon \leq \lim_{n \rightarrow\infty} |\mu (g) - \mu_n'' (g)| = |\mu (g) - \nu (g)| = 0.
\]

Remark 25.21. One could also use Proposition 25.43 and Bochner’s Theorem 25.46 below to conclude; if \( f (\lambda) := \lim_{n \rightarrow\infty} \hat{\mu}_n (\lambda) \) 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 25.22 (Example 25.10 continued). For \( c > 0 \), let \( d\mu_c (x) = \frac{1}{2c} \chi_{[-c,c]} (x) \, dx \). As in Example 25.10, we know \( \hat{\mu}_c (\lambda) = \frac{\sin \lambda c}{\lambda c} \). In this case, 
\[
\mu_\infty (\lambda) := \lim_{c \rightarrow\infty} \hat{\mu}_c (\lambda) = \begin{cases} 
1 & \text{if } \lambda = 0 \\
0 & \text{if } \lambda \neq 0 
\end{cases}.
\]

Noise that the limiting function is discontinuous and is not the characteristic function of a measure. Moreover, since 
\[
F_c (x) := \mu_c ((-\infty, x]) = \begin{cases} 
0 & \text{if } x \leq -c \\
1 & \text{if } x \geq c 
\end{cases} \Rightarrow F_c (x) = \frac{1}{2} \text{ as } c \uparrow \infty.
\]

Corollary 25.23. Suppose that \( \{X_n\}_{n=1}^\infty \cup \{X\} \) are random vectors in \( \mathbb{R}^d \), then \( X_n \Rightarrow X \) iff \( \lim_{n \rightarrow\infty} \mathbb{E} [e^{i\lambda \cdot X_n}] = \mathbb{E} [e^{i\lambda \cdot X}] \) for all \( \lambda \in \mathbb{R}^d \).

Proof. Since \( f (x) := e^{i\lambda \cdot x} \) is in \( BC (\mathbb{R}^d) \) for all \( \lambda \in \mathbb{R}^d \), \( X_n \Rightarrow X \) then \( \lim_{n \rightarrow\infty} \mathbb{E} [e^{i\lambda \cdot X_n}] = \mathbb{E} [e^{i\lambda \cdot X}] \). Conversely if \( \lim_{n \rightarrow\infty} \mathbb{E} [e^{i\lambda \cdot X_n}] = \mathbb{E} [e^{i\lambda \cdot X}] \) for all \( \lambda \in \mathbb{R}^d \) and \( \mu_n := \text{Law} (X_n) \) and which is equivalent to \( X_n \Rightarrow X \).

The proof of the next corollary is a straightforward consequence of Corollary 25.23 used for dimension \( d \) and dimension 1.

Corollary 25.24. Suppose that \( \{X_n\}_{n=1}^\infty \cup \{X\} \) are random vectors in \( \mathbb{R}^d \), then \( X_n \Rightarrow X \) if \( \lambda \cdot X_n \Rightarrow \lambda \cdot X \) for all \( \lambda \in \mathbb{R}^d \).

Lemma 25.25. 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 \( \mu_n (\mathbb{R}^d \setminus [-M_\varepsilon, M_\varepsilon]^d) \leq \varepsilon \) for all \( n \). Let \( \lambda, h, \in \mathbb{R}^d \), then 
\[
\begin{align*}
|\hat{\mu}_n (\lambda + h) - \hat{\mu}_n (\lambda)| &\leq \int_{\mathbb{R}^d} |e^{ix(\lambda + h)} - e^{ix\lambda}| \, d\mu_n (x) \\
&= \int_{\mathbb{R}^d} |e^{ixh} - 1| \, d\mu_n (x) \\
&\leq 2\varepsilon + \sup_{x \in [-M_\varepsilon, M_\varepsilon]^d} |e^{ixh} - 1|.
\end{align*}
\]

Therefore it follows that 
\[
\limsup_{h \rightarrow 0} \sup_{\lambda \in \mathbb{R}^d} |\hat{\mu}_n (\lambda + h) - \hat{\mu}_n (\lambda)| \leq 2\varepsilon
\]
and as \( \varepsilon > 0 \) was arbitrary the result follows.
Corollary 25.26 (Uniform Convergence). If \( \mu_n \xrightarrow{\ast} \mu \) as \( n \to \infty \) then \( \mu_n (\lambda) \to \mu (\lambda) \) uniformly on compact subsets of \( \mathbb{R}^n \).

Proof. This is a consequence of Theorem 25.20, Lemma 25.25 and the Arzela - Ascoli Theorem ??. For completeness here is a sketch of the proof.

Let \( K \) be a compact subset of \( \mathbb{R}^n \) and \( \varepsilon > 0 \) be given. Applying Lemma 25.25 to \( \{ \mu \} \cup \{ \mu_n \} \) we know that there exists \( \delta > 0 \) such that
\[
\sup_{\lambda} |\hat{\mu}_n (\lambda + h) - \hat{\mu}_n (\lambda)| \leq \varepsilon \text{ and } \sup_{\lambda} |\hat{\mu} (\lambda + h) - \hat{\mu} (\lambda)| \leq \varepsilon
\]
whenever \( ||h|| \leq \delta \). Let \( F \subset K \) be a finite set such that \( K \subset \cup_{\xi \in F} B (\xi, \delta) \). Since we already know that \( \mu_n \to \mu \) 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. (25.8) to conclude that
\[
\sup_{\lambda \in K} |\hat{\mu}_n (\lambda) - \hat{\mu} (\lambda)| \leq 2\varepsilon + \max_{\xi \in F} |\hat{\mu}_n (\xi) - \hat{\mu} (\xi)|
\]
and therefore, \( \limsup_{n \to \infty} \sup_{\lambda \in K} |\hat{\mu}_n (\lambda) - \hat{\mu} (\lambda)| \leq 2\varepsilon \). As \( \varepsilon > 0 \) was arbitrary the result follows.

The following lemma will be needed before giving our first applications of the continuity theorem.

Lemma 25.27. 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 25.48 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,
\[
(1 + z_n)^n = \left[ e^{\ln(1+z_n)} \right]^n = e^{n \ln(1+z_n)} = e^{n(z_n + O(\frac{1}{n^2}))} \to e^{\xi} \text{ as } n \to \infty.
\]

Proposition 25.28 (Weak Law of Large Numbers revisited). Suppose that \( \{ X_n \}_{n=1}^{\infty} \) are i.i.d. integrable random variables. Then \( \frac{S_n}{n} \overset{P}{\to} E X_1 =: c \).

Proof. Let \( f (\lambda) := f_{X_1} (\lambda) = E \left[ e^{i\lambda X_1} \right] \) in which case
\[
f_{\frac{S_n}{n}} (\lambda) = \left[ f \left( \frac{\lambda}{n} \right) \right]^n.
\]
By Taylor’s theorem (see Appendix 25.8), \( f (\lambda) = 1 + k (\lambda) \lambda \) where
\[
\lim_{\lambda \to 0} k (\lambda) = k (0) = f' (0) = i E [X_1].
\]
It now follows from Lemma 25.27 that
\[
f_{\frac{S_n}{n}} (\lambda) = \left[ 1 + k \left( \frac{\lambda}{n} \right) \frac{\lambda}{n} \right]^n \to e^{i c \lambda} \text{ as } n \to \infty
\]
which is the characteristic function of the constant random variable, \( c \). By the continuity Theorem 25.20 it follows that \( \frac{S_n}{\sqrt{n}} \overset{P}{\to} c \) and since \( c \) is constant we may apply Lemma 24.28 to conclude \( \frac{S_n}{\sqrt{n}} \overset{P}{\to} c = E X_1 \).

We are now ready to continue are investigation of central limit theorems that was begun with Theorem 13.42 above.

Theorem 25.29 (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 \( \frac{S_n}{\sqrt{n}} \overset{P}{\to} N (0, 1) \).

Proof. If \( f (\lambda) := E \left[ e^{i\lambda X_1} \right] \), then by Taylor’s theorem (see Appendix 25.8),
\[
f (\lambda) = f (0) + f' (0) \lambda + \frac{1}{2} k (\lambda) \lambda^2 = 1 + \frac{1}{2} k (\lambda) \lambda^2 \quad (25.9)
\]
where
\[
\lim_{\lambda \to 0} k (\lambda) = k (0) = f'' (0) = -E [X_1^2] = -1.
\]
Hence, using Lemma 25.27 we find
\[
E \left[ e^{i \lambda \frac{S_n}{\sqrt{n}}} \right] = \left[ f \left( \frac{\lambda}{\sqrt{n}} \right) \right]^n = \left[ 1 + \frac{1}{2} k \left( \frac{\lambda}{\sqrt{n}} \right) \frac{\lambda^2}{n} \right]^n \to e^{-\lambda^2/2}.
\]
Since \( e^{-\lambda^2/2} \) is the characteristic function of \( N (0, 1) \) (Example 25.13), the result now follows from the continuity Theorem 25.20.

Alternative proof. Again it suffices to show
\[
\lim_{n \to \infty} E \left[ e^{i \lambda \frac{S_n}{\sqrt{n}}} \right] = e^{-\lambda^2/2} \text{ for all } \lambda \in \mathbb{R}.
\]
We do this using Lemma 26.14 below as follows:

\[ |f_{\frac{Sn}{\sqrt{n}}} (\lambda) - e^{-\lambda^2/2}| = \left| \left[f \left( \frac{\lambda}{\sqrt{n}} \right) \right]^n - [e^{-\lambda^2/2n}]^n \right| \]
\[ \leq n \left| f \left( \frac{\lambda}{\sqrt{n}} \right) - e^{-\lambda^2/2n} \right| \]
\[ = n \left( 1 - \frac{1}{2} \left( 1 + \varepsilon \left( \frac{\lambda}{\sqrt{n}} \right) \right) \lambda^2 n - \left( 1 - \frac{\lambda^2}{2n} + O \left( \frac{1}{n^2} \right) \right) \right| \]
\[ \to 0 \text{ as } n \to \infty. \]

Corollary 25.30. 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

\[ \sup_{\lambda \in \mathbb{R}} \left| P \left( \frac{S_n}{\sqrt{n}} \leq y \right) - P(N(0,1) \leq y) \right| \to 0 \text{ as } n \to \infty. \quad (25.10) \]

**Proof.** This is a direct consequence of Theorem 25.29 and Exercise 24.7.

Remark 25.31 (Why normal?). It is now a reasonable question to ask “why” is the limiting random variable normal in Theorem 25.29. One way to understand this is, if under the assumptions of Theorem 25.29, we know \( S_n \xrightarrow{d} L \) where \( L \) is some random variable with \( \mathbb{E}L = 0 \) and \( \mathbb{E}L^2 = 1 \), then

\[ \frac{S_{2n}}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left( \sum_{k=1, \text{odd}}^{2n} \frac{X_j}{\sqrt{n}} + \sum_{k=1, \text{even}}^{2n} \frac{X_j}{\sqrt{n}} \right) \]
\[ \xrightarrow{d} \frac{1}{\sqrt{2}} (L_1 + L_2) \]

where \( L_1 \overset{d}{=} L \overset{d}{=} L_2 \) and \( L_1 \) and \( L_2 \) are independent – see Exercise 24.12. In particular this implies that

\[ f(\lambda) = \left[ f \left( \frac{\lambda}{\sqrt{2}} \right) \right]^2 \text{ for all } \lambda \in \mathbb{R}. \quad (25.12) \]

We could also arrive at Eq. (25.12) by passing to the limit in the identity,

\[ f_{\frac{Sn}{\sqrt{n}}} (\lambda) = f_{\frac{Sn}{\sqrt{n}}} \left( \lambda \right) f_{\frac{Sn}{\sqrt{n}}} \left( \frac{\lambda}{\sqrt{2}} \right). \]

Iterating Eq. (25.12) and then Eq. (25.9) and Lemma 25.27 above we again deduce that,

\[ f(\lambda) = \left[ f \left( \frac{\lambda}{\sqrt{2}} \right) \right]^{2n} \]
\[ = \left[ 1 + \frac{1}{2} k \left( \frac{\lambda}{2n/2} \right) \lambda^2 \right]^{2n} \to e^{-\frac{1}{2}\lambda^2} = f_{N(0,1)}(\lambda). \]

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 25.32 (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 \( Q = \mathbb{E} \left( (X_1 - m) (X_1 - m)^{tr} \right) \), that is \( m \in \mathbb{R}^d \) and \( Q \) is the \( d \times d \) matrix defined by

\[ m_j := \mathbb{E} (X_1)_j \text{ and } Q_{ij} := \mathbb{E} \left( (X_1 - m)_i (X_1 - m)_j \right) = \text{Cov} \left( (X_1)_i, (X_1)_j \right) \]

for all \( 1 \leq i, j \leq d \). Then

\[ \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \xrightarrow{d} Z \quad (25.13) \]

where \( Z \overset{d}{=} N(0,Q) \), i.e. \( Z \) is a random vector such that

\[ \mathbb{E} [e^{i\lambda \cdot Z}] = \exp \left( -\frac{1}{2} Q \lambda \cdot \lambda \right) \text{ for all } \lambda \in \mathbb{R}^d. \]

**Proof.** Let \( \lambda \in \mathbb{R}^d \), then
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 \( \text{Var}(\lambda \cdot X_k) = Q\lambda \cdot \lambda \). 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
\[
\lambda \cdot \left( \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m) \right) = 0 \quad \Rightarrow \quad 0 = \lambda \cdot Z. \tag{25.14}
\]

If \( Q\lambda \cdot \lambda > 0 \) then \( \left\{ \lambda \cdot X_k - \lambda \cdot m \right\}_{k=1}^\infty \) satisfy the hypothesis of Theorem 25.29 and therefore,
\[
\frac{1}{\sqrt{Q\lambda \cdot \lambda}} \lambda \cdot \left[ \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m) \right] = \frac{1}{\sqrt{n}} \sum_{k=1}^n \lambda \cdot X_k - \lambda \cdot m \quad \Rightarrow \quad N(0,1)
\]
which combined with Eq. (25.14) implies, for all \( \lambda \in \mathbb{R}^d \), we have
\[
\lambda \cdot \left[ \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - m) \right] \quad \Rightarrow \quad \sqrt{Q\lambda \cdot \lambda} \cdot N(0,1) \quad \Rightarrow \quad \lambda \cdot Z.
\]

We may now apply Corollary 25.24 to conclude that Eq. (25.13) holds.

### 25.5 A Fourier Transform Inversion Formula

Corollary 11.13 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 \( e_n^L (x) := e^{-i\frac{x}{L}} \) and let
\[
(f, g)_L := \frac{1}{2\pi L} \int_{-\pi L}^{\pi L} f(x) g(x) \, dx
\]
for \( f, g \in L^2([-\pi L, \pi L], dx) \). Then it is well known (and fairly elementary to prove) that \( \{e_n^L : n \in \mathbb{Z}\} \) is an orthonormal basis for \( L^2([-\pi L, \pi L], dx) \). In particular, if \( f \in C_c(\mathbb{R}) \) with \( \text{supp}(f) \subset [-\pi L, \pi L] \), then for \( x \in [-\pi L, \pi L] \),
\[
\hat{f}(\lambda) = \int_{-\pi L}^{\pi L} f(x) e^{-i\frac{\pi L}{\lambda} x} \, dx
\]
where
\[
\hat{f}(\lambda) = \int_{-\pi L}^{\pi L} f(y) e^{i\lambda y} \, dy.
\]
Letting \( L \to \infty \) in Eq. (25.15) then suggests that
\[
\hat{f}(\lambda) = \int_{-\infty}^{\infty} f(y) e^{-i\lambda y} \, dy
\]
and we are lead to expect,
\[
f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\lambda) e^{-i\pi L x} \, d\lambda. \tag{25.16}
\]

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. (25.16) we expect that
\[
\mu([a,b]) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(\lambda) \left( \int_{a}^{b} e^{-i\pi L x} \, dx \right) d\lambda
\]
\[
= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(\lambda) \left( \frac{\sin \pi L (b-a)}{i\lambda} \right) d\lambda
\]
\[
= \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \hat{\mu}(\lambda) \left( \frac{\sin \pi L c - \sin \pi L (-c)}{i\lambda} \right) d\lambda. \tag{25.17}
\]

We will prove this formula is essentially correct in Theorem 25.34 below. The following lemma is the key to computing the limit appearing in Eq. (25.17) which will be the heart of the proof of the inversion formula.

**Lemma 25.33.** For \( c > 0 \), let
\[
S(c) := \int_{-c}^{c} \frac{\sin \lambda y}{\lambda} \, d\lambda. \tag{25.18}
\]
Then \( S(c) \) is a continuous function such that \( S(c) \to \pi \) boundedly as \( c \to \infty \), see Figure 25.1. Moreover for any \( y \in \mathbb{R} \) we have
\[
\int_{-c}^{c} \frac{\sin \lambda y}{\lambda} d\lambda = \text{sgn}(y) S\left( |c| \right) \tag{25.19}
\]
where
\[
\text{sgn}(y) = \begin{cases} 
1 & \text{if } y > 0 \\
-1 & \text{if } y < 0 \\
0 & \text{if } y = 0
\end{cases}
\]
Theorem 25.34 (Fourier Inversion Formula). Let \( f \) be a probability measure on \((\mathbb{R}, B_\mathbb{R})\) such that \( f \in L^1(m)\), then \( df = \rho dm\) where \( \rho \) is the continuous probability density on \( \mathbb{R} \) given by

\[
\rho(x) := \frac{1}{2\pi} \int \hat{f}(\lambda) e^{-i\lambda x} d\lambda.
\]  

(At the end points, the limit picks up only half of the mass.)

Proof. Let \( I(c) \) denote the integral appearing in Eq. (25.21). By Fubini’s theorem and Lemma 25.33, we have

\[
I(c) := \int_{-c}^{c} \hat{f}(\lambda) \left( \frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda
\]

(25.22)

It follows that \( \text{Im} \left( \frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right) \) is an odd function of \( \lambda \) and

\[
\text{Re} \left( \frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right) = \frac{1}{\lambda} \left[ \sin(\lambda(x-a)) - \sin(\lambda(x-b)) \right],
\]

(25.23)

and therefore (using Lemma 25.33)

\[
I(c) = \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda \text{Re} \left( \frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right)
\]

\[
= \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda \frac{\sin(\lambda(x-a)) - \sin(\lambda(x-b))}{\lambda}
\]

\[
= \int_{\mathbb{R}} d\mu(x) \left[ \text{sgn}(x-a) \text{sgn}(c|x-a|) - \text{sgn}(x-b) \text{sgn}(c|x-b|) \right].
\]

Using Lemma 25.33 again along with the DCT we may pass to the limit as \( c \to \infty \) in the previous identity to get the result;

\[
\lim_{c \to \infty} \frac{1}{2\pi} I(c) = \frac{1}{2} \int_{\mathbb{R}} d\mu(x) \left[ \text{sgn}(x-a) - \text{sgn}(x-b) \right]
\]

\[
= \frac{1}{2} \int_{\mathbb{R}} d\mu(x) \left[ 2 \cdot 1_{(a,b)}(x) + 1_{(a]}(x) + 1_{[b)}(x) \right]
\]

\[
= \mu((a,b)) + \frac{1}{2} [\mu\{a\}] + \mu\{b\}.
\]
Proof. The function $\rho$ defined in Eq. (25.23) is continuous by the dominated convergence theorem. Moreover for any $-\infty < a < b < \infty$ we have

$$\int_a^b \rho(x) \, dx = \frac{1}{2\pi} \int_a^b dx \int_{-\infty}^{\infty} d\lambda \, \hat{\mu}(\lambda) e^{-i\lambda x}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \, \hat{\mu}(\lambda) \int_a^b dx e^{-i\lambda x}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \, \hat{\mu}(\lambda) \left( e^{-i\lambda a} - e^{-i\lambda b} \right)$$

$$= \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} d\lambda \, \left( e^{-i\lambda a} - e^{-i\lambda b} \right)$$

$$= \mu((a,b)) + \frac{1}{2} \mu(\{a\}) + \mu(\{b\}),$$

wherein we have used Theorem 25.34 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

$$\mu((a,b)) = \int_a^b \rho(x) \, dx$$

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 $\int h d\mu = \int h \rho d\mu$ for all bounded measurable functions $h : \mathbb{R} \to \mathbb{R}$. This then implies that $\rho \geq 0$, $m - \alpha e_1$\footnote{Since $\rho$ is continuous we may further conclude that $\rho(x) \geq 0$ for every $x \in \mathbb{R}$.} and the $d\mu = \rho dm$.

Example 25.36. Let $\rho(x) = (1 - |x|)_+$ be the triangle density in Figure 25.2

Fig. 25.2. The triangular density function.

Recall from Example 25.6 that

\[ \int_{-\infty}^{\infty} e^{i\lambda x} (1 - |x|)_+ \, dx = \frac{2(1 - \cos \lambda)}{\lambda^2}. \]

Alternatively by direct calculation,

\[ \int_{-\infty}^{\infty} e^{i\lambda x} (1 - |x|)_+ \, dx = 2 \text{Re} \int_{0}^{1} e^{i\lambda x} (1 - x) \, dx \]

\[ = 2 \text{Re} \left[ \left( 1 - \frac{1}{i} \frac{d}{d\lambda} \right) \int_{0}^{1} e^{i\lambda x} \, dx \right] \]

\[ = 2 \text{Re} \left[ \left( 1 - \frac{1}{i} \frac{d}{d\lambda} \right) \frac{e^{i\lambda} - 1}{i\lambda} \right] \]

\[ = 2 \frac{1 - \cos \lambda}{\lambda^2}. \]

Hence it follows\footnote{This identity could also be verified directly using residue calculus techniques from complex variables.} from Corollary 25.35 that

\[ (1 - |x|)_+ = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos \lambda}{\lambda^2} e^{-i\lambda x} \, d\lambda. \]

Evaluating Eq. (25.24) at $x = 0$ gives the identity

\[ 1 = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos \lambda}{\lambda^2} \, d\lambda. \]

from which we deduce that

\[ d\mu(x) := \frac{1}{\pi} \frac{1 - \cos x}{x^2} \, dx \]

is a probability measure such that (from Eq. (25.24)) has characteristic function,

\[ \hat{\mu}(\lambda) = (1 - |\lambda|)_+. \]

Corollary 25.37. For all random variables, $X$, we have

\[ \mathbb{E} |X| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \text{Re} f_X(\lambda)}{\lambda^2} \, d\lambda. \]

Proof. For $M \in \mathbb{R} \setminus \{0\}$, make the change of variables, $\lambda \to M\lambda$ in Eq. (25.25) for find

\[ |M| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos (\lambda M)}{\lambda^2} \, d\lambda. \]
Observe the identity holds for $M = 0$ as well. Taking $M = X$ in Eq. (25.29) and then taking expectations implies,

$$E\ |X| = \frac{1}{\pi} \int_R \frac{1 - \cos \lambda X}{\lambda^2} \, d\lambda = \frac{1}{\pi} \int_R \frac{1 - Re f_X(\lambda)}{\lambda^2} \, d\lambda.$$

Suppose that we did not know the value of $c := \int_{-\infty}^{\infty} \frac{1 - \cos \lambda X}{\lambda^2} \, d\lambda$ is $\pi$, we could still proceed as above to learn

$$E\ |X| = \frac{1}{c} \int_R \frac{1 - Re f_X(\lambda)}{\lambda^2} \, d\lambda.$$

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

$$E\ |X| = \frac{1}{\sqrt{2\pi}} \int_R |x| e^{-x^2/2} \, dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} \, dx = \frac{\sqrt{2\pi}}{2}.$$

On the other hand, $f_X(\lambda) = e^{-\lambda^2/2}$ and so

$$\sqrt{\frac{2}{\pi}} = \frac{1}{c} \int_R \left(1 - e^{-\lambda^2/2}\right) d\left(\lambda^{-1}\right) = \frac{1}{c} \int_R d\left(1 - e^{-\lambda^2/2}\right) (\lambda^{-1})$$

$$= \frac{1}{c} \int_R e^{-\lambda^2/2} \, d\lambda = \frac{\sqrt{2\pi}}{c}$$

from which it follows, again, that $c = \pi$.

**Corollary 25.38.** Suppose $X$ is a random variable and there exists $\varepsilon > 0$ such that $u(\lambda) := Re f_X(\lambda) = E[\cos \lambda X]$ is continuously differentiable for $\lambda \in (-2\varepsilon, 2\varepsilon)$. If we further assume that

$$\int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda} \, d\lambda < \infty, \quad (25.30)$$

then $E\ |X| < \infty$ and $f_X \in C^1 (R, C)$. (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. (25.30) would hold.)

**Proof.** According to Eq. (25.28)

$$\pi \cdot E\ |X| = \int_R \frac{1 - u(\lambda)}{\lambda^2} \, d\lambda = \int_{|\lambda| \leq \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} \, d\lambda + \int_{|\lambda| > \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} \, d\lambda.$$

Since $0 \leq 1 - u(\lambda) \leq 2$ and $2/\lambda^2$ is integrable for $|\lambda| > \varepsilon$, to show $E\ |X| < \infty$ we must show

$$\int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda} \, d\lambda < \infty.$$
25 Characteristic Functions (Fourier Transform)

\[ f_Y(\lambda) := \mathbb{E}[e^{\lambda Y}] = \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x) \, d\nu(x)\right) \]

where \( \nu \) is the discrete measure on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) given by

\[ \nu = \alpha^2 \delta_0 + \sum_{k=1}^{n} a_k \lambda_k^2 \delta_{\lambda_k}. \]  

(25.32)

**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

\[ \varphi(\lambda, x) = -\lambda^2 \int_{0}^{1} e^{it\lambda x} (1 - t) \, dt. \]  

(25.33)

From this formula it is clear that \( \varphi \) is a smooth function of \((\lambda, x)\).

**Exercise 25.4.** To each finite and compactly supported measure, \( \nu \), on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\), there exists a sequence \( \{\nu_n\}_{n=1}^{\infty} \) of finitely supported finite measures on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) such that \( \nu_n \rightarrow \nu \). 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, \( A \subset \mathbb{R} \) such that \( \nu(\mathbb{R} \setminus A) = 0 \).

**Exercise 25.5.** Show that if \( \nu \) is a finite measure on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\), then

\[ f(\lambda) := \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x) \, d\nu(x)\right) \]  

(25.34)

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 25.8 to be of help.)

1. Show \( f(\lambda) \) is continuous.
2. Suppose \( \nu \) is compactly supported. Show, using Exercises 25.3 and the continuity Theorem 25.20 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 25.6 (Exercise 2.3 in [44].)** Let \( \mu \) be the probability measure on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\), such that \( \mu(\{n\}) = p(n) = c \frac{1}{n^{1+\ln|n|}} 1_{|n| \geq 2} \) with \( c \) chosen so that \( \sum_{n \in \mathbb{Z}} p(n) = 1 \). Show that \( \mu \in C^1(\mathbb{R}, \mathbb{C}) \) 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 25.7 (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 \((\mathbb{R}, \mathcal{B}_\mathbb{R})\).

**Exercise 25.8.** Let \( d\nu(x) = \frac{1}{2} e^{-|x|} dx \). Find \( \hat{\nu}(\lambda) \) and use your result to conclude,

\[ \frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{1 + \lambda^2} e^{-i\lambda x} d\lambda = e^{-|x|} \quad \forall \, x \in \mathbb{R}. \]

(25.35)

**Remark:** Another standard way to compute this integral is to use residue calculus from complex variable theory.

**Exercise 25.9.** Let

\[ c := \int_{-\infty}^{\infty} \frac{(1 - \cos \lambda)^2}{\lambda^4} d\lambda, \]

(25.36)

and \( u(\lambda) := \text{Re} f_X(\lambda) \) where \( f_X \) is the characteristic function of a random variable \( X \). Show, using the ideas in and after Corollary 25.37 that

\[ \mathbb{E}|X|^3 = \frac{1}{c} \int_{\mathbb{R}} \frac{3}{2} - 2u(\lambda) + \frac{1}{2} u(2\lambda) \, d\lambda. \]

(25.37)

[If you are adventurous you might try to find a similar (but more complicated formula) to compute \( \mathbb{E}|X|^{2k+1} \) for all \( k \in \mathbb{N}_0 \).]
25.7 Appendix: Bochner’s Theorem

Definition 25.39. 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 $|f(x)| \leq C_N (1 + |x|)^{-N}$ 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 that

$$\sup_{x \in \mathbb{R}^n} (1 + |x|)^{-N} |f(x)| < \infty,$$

i.e. there exists $N \in \mathbb{N}$ and $C < \infty$ such that $|f(x)| \leq C (1 + |x|)^N$ for all $x \in \mathbb{R}^n$.

Definition 25.40 (Schwartz Test Functions). Let $S$ 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

$$\|f\|_{N, \alpha} = \sup_{x \in \mathbb{R}^n} |(1 + |x|)^N \partial^\alpha f(x)|$$

so that

$$S = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N, \alpha} < \infty \text{ for all } N \text{ and } \alpha \right\}.$$

Also let $P$ 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 $P$ iff for all multi-indices $\alpha$, there exists $N_\alpha < \infty$ such that

$$\sup_{x \in \mathbb{R}^n} |(1 + |x|)^{-N_\alpha} \partial^\alpha g(x)| < \infty.$$

(Notice that any polynomial function on $\mathbb{R}^n$ is in $P$.)

Definition 25.41. A function $\chi : \mathbb{R}^n \rightarrow \mathbb{C}$ is said to be positive (semi) definite if 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 25.42. If $\mu$ is a finite positive measure on $\mathcal{B}_{\mathbb{R}^n}$, then $\chi := \mu \in C(\mathbb{R}^n, \mathbb{C})$ is a positive definite function.

Proof. The dominated convergence theorem implies $\mu \in C(\mathbb{R}^n, \mathbb{C})$. Since $\mu$ is a positive measure (and hence real),

$$\hat{\mu}(-\xi) = \int_{\mathbb{R}^n} e^{-i\xi \cdot x} \mu(x) = \int_{\mathbb{R}^n} e^{-i\xi \cdot x} d\mu(x) = \overline{\hat{\mu}(\xi)}.$$

From this it follows that for any $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$, the matrix $A := \{\mu(\xi_k - \xi_j)\}_{k,j=1}^m$ is self-adjoint. Moreover if $\lambda \in \mathbb{C}^m$,

$$\sum_{k,j=1}^m \mu(\xi_k - \xi_j) \lambda_k \bar{\lambda_j} = \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i(\xi_k - \xi_j) \cdot x} \lambda_k \bar{\lambda_j} d\mu(x)$$

$$= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i\xi_k \cdot x} \lambda_k e^{-i\xi_j \cdot x} \bar{\lambda_j} d\mu(x)$$

$$= \int_{\mathbb{R}^n} \sum_{k=1}^m e^{-i\xi_k \cdot x} \lambda_k \bar{\lambda_k}^2 d\mu(x) \geq 0$$

showing $A$ is positive definite.

Proposition 25.43. Suppose that $\chi : \mathbb{R}^n \rightarrow \mathbb{C}$ 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 25.41 we conclude that

$$A := \begin{bmatrix} 1 & \chi(x - y) & \chi(y) \\ \chi(y - x) & 1 & \chi(y) \\ \chi(-y) & \chi(y) & 1 \end{bmatrix} = \begin{bmatrix} 1 & \chi(x - y) & \chi(y) \\ \chi(x - y) & 1 & \chi(y) \\ \chi(x) & \chi(y) & 1 \end{bmatrix}$$

is positive definite. In particular,

$$0 \leq \det A = 1 + \chi(x - y) \chi(y) \bar{\chi}(x) + \chi(x) \bar{\chi}(x - y) \bar{\chi}(y)$$

$$- |\chi(x)|^2 - |\chi(y)|^2 - |\chi(x - y)|^2.$$

Combining this inequality with the identity,

$$|\chi(x) - \chi(y)|^2 = |\chi(x)|^2 + |\chi(y)|^2 - \chi(x) \chi(y) - (\chi(x) \bar{\chi}(y) + \chi(y) \bar{\chi}(x)),$$

gives

$$0 \leq 1 - |\chi(x - y)|^2 + \chi(x - y) \chi(y) \bar{\chi}(x) + \chi(x) \bar{\chi}(x - y) \bar{\chi}(y)$$

$$- \left\{ |\chi(x) - \chi(y)|^2 + \chi(x) \bar{\chi}(y) + \chi(y) \bar{\chi}(x) \right\}$$

$$= 1 - |\chi(x - y)|^2 - |\chi(x) - \chi(y)|^2$$

$$+ \chi(x - y) \chi(y) \bar{\chi}(x) - \chi(x) \bar{\chi}(y) + \chi(x) \bar{\chi}(x - y) \bar{\chi}(y) - \chi(x) \chi(y)$$

$$= 1 - |\chi(x - y)|^2 - |\chi(x) - \chi(y)|^2 + 2 \Re((\chi(x - y) - 1)(\chi(y) \bar{\chi}(x)))$$

$$\leq 1 - |\chi(x - y)|^2 - |\chi(x) - \chi(y)|^2 + 2 |\chi(x - y) - 1|.$$
Hence we have

\[ |\chi(x) - \chi(y)|^2 \leq 1 - |\chi(x-y)|^2 + 2|\chi(x-y)| - 1 \]
\[ = (1 - |\chi(x-y)|)(1 + |\chi(x-y)|) + 2|\chi(x-y)| - 1 \]
\[ \leq 4|1 - \chi(x-y)| \]

which completes the proof. \[\blacksquare\]

**Remark 25.44.** 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 if \( f \) is not continuous at \( \lambda = 0 \).

**Lemma 25.45.** If \( \chi \in C(\mathbb{R}^n, \mathbb{C}) \) is a positive definite function, then

1. \( \chi(0) \geq 0 \).
2. \( \chi(-\xi) = \chi(\xi) \) for all \( \xi \in \mathbb{R}^n \).
3. \( |\chi(\xi)| \leq \chi(0) \) for all \( \xi \in \mathbb{R}^n \).
4. If we further assume that \( \chi \) is continuous, then

\[
\int_{\mathbb{R}^n \times \mathbb{R}^n} \chi(\xi - \eta) f(\xi)\overline{f(\eta)} d\xi d\eta \geq 0 \tag{25.38}
\]

for all \( f \in S(\mathbb{R}^d) \).

**Proof.** Taking \( m = 1 \) and \( \xi_1 = 0 \) we learn \( \chi(0)|\lambda|^2 \geq 0 \) 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 \( \chi(\xi - \eta) = \chi(\eta - \xi) \) (since \( A = A^* \) by definition) and

\[
0 \leq \det \begin{bmatrix}
\chi(0) & \chi(\xi - \eta) \\
\chi(\eta - \xi) & \chi(0)
\end{bmatrix} = |\chi(0)|^2 - |\chi(\xi - \eta)|^2.
\]

and hence \( |\chi(\xi)| \leq \chi(0) \) for all \( \xi \). This proves items 2. and 3. Item 4. follows by approximating the integral in Eq. \(25.38\) by Riemann sums,

\[
\int_{\mathbb{R}^n \times \mathbb{R}^n} \chi(\xi - \eta) f(\xi)\overline{f(\eta)} d\xi d\eta = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}^n \times \mathbb{R}^n} \chi(\xi - \eta) f(\xi)\overline{f(\eta)} d\xi d\eta \geq 0.
\]

The details are left to the reader keeping in mind this is where we must use the assumption that \( \chi \) is continuous. \[\blacksquare\]
which coupled with the dominated convergence theorem shows

$$\langle I \star \rho_t, \psi \rangle \to \int_{\mathbb{R}^n} \chi(\xi) \psi^\prime(\xi) d\xi = I(\psi) \text{ as } t \downarrow 0.$$  

Hence if $\psi \geq 0$, then $I(\psi) = \lim_{t \downarrow 0} \langle I_t, \psi \rangle \geq 0$.

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 $f \in C^\infty_c(\mathbb{R}, \mathbb{R})$ is a smooth function with $\text{supp}(f) \subset K$, then $0 \leq |f|_{\infty} \psi - f \in S$ and hence

$$0 \leq \langle I, |f|_{\infty} \psi - f \rangle = |f|_{\infty} \langle I, \psi \rangle - \langle I, f \rangle$$

and therefore $\langle I, f \rangle \leq |f|_{\infty} \langle I, \psi \rangle$. Replacing $f$ by $-f$ implies, $-\langle I, f \rangle \leq |f|_{\infty} \langle I, \psi \rangle$ and hence we have proved

$$|\langle I, f \rangle| \leq C(\text{supp}(f)) |f|_{\infty}$$  \hspace{1cm} (25.40)

for all $f \in D_{\mathbb{R}^n} := C^\infty_c(\mathbb{R}^n, \mathbb{R})$ where $C(K)$ is a finite constant for each compact subset of $\mathbb{R}^n$. Because of the estimate in Eq. (25.40), it follows that $I|_{D_{\mathbb{R}^n}} \equiv 0$ has a unique extension $I$ to $C_c(\mathbb{R}^n, \mathbb{R})$ still satisfying the estimates in Eq. (25.40) 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 $\hat{\mu}(\eta) = \hat{\chi}(\eta)$ for all $\eta \in \mathbb{R}^n$ given

$$\mu(f) = \int_{\mathbb{R}^n} \chi(\xi) f^\prime(\xi) d\xi \text{ for all } f \in C^\infty_c(\mathbb{R}^n, \mathbb{R}).$$  \hspace{1cm} (25.41)

Let $f \in C^\infty_c(\mathbb{R}^n, \mathbb{R})$ 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 ??,

$$\mathcal{F}^{-1} \left[ e^{-i\eta x} f_\varepsilon(x) \right] (\xi) = e^{-n \varepsilon f^\prime(\frac{\xi - \eta}{\varepsilon})}$$

and therefore, from Eq. (25.41),

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) = \int_{\mathbb{R}^n} \chi(\xi) e^{-n f^\prime(\frac{\xi - \eta}{\varepsilon})} d\xi.$$  \hspace{1cm} (25.42)

Because $\int_{\mathbb{R}^n} f^\prime(\xi) d\xi = \mathcal{F} f^\prime(0) = f(0) = 1$, we may apply the approximate $\delta$–function Theorem (25.47) below to Eq. (25.42) to find (using the continuity of $\chi$ here!)

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) \to \chi(\eta) \text{ as } \varepsilon \downarrow 0.$$  \hspace{1cm} (25.43)

On 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 conclude

$$\mu(e^{-i\eta x} f_\varepsilon(x)) \to \mu(e^{-i\eta x}) = \hat{\mu}(\eta) \text{ as } \varepsilon \downarrow 0$$

for all $\eta$. Combining this equation with Eq. (25.43) shows $\hat{\mu}(\eta) = \hat{\chi}(\eta)$ for all $\eta \in \mathbb{R}^n$.

**Better proof** is to use continuity Theorem 24.63 from the probability notes which is based on Helly’s selection theorem. To this end, let

$$\rho_t(x) := (\chi \cdot \hat{\rho}_t) \chi(x) = \int_{\mathbb{R}^n} \chi(\xi) \hat{\rho}_t(\xi) e^{i\xi \cdot x} d\xi.$$  

Notice that

$$\hat{\rho}_t(\xi) e^{i\xi \cdot x} = e^{i\xi \cdot x} \int_{\mathbb{R}^n} p_t(y) e^{-iy \cdot \xi} dy$$

$$= \int_{\mathbb{R}^n} p_t(y) e^{i(x-y) \cdot \xi} dy$$

$$= \int_{\mathbb{R}^n} p_t(x+y) e^{-iy \cdot \xi} dy$$

$$= [p_t(x+\cdot)]^\wedge$$

so that

$$\rho_t(x) = \langle \chi, [p_t(x+\cdot)]^\wedge \rangle = \chi, \left[ \left( \int \sqrt{p_t(x+\cdot)} \right)^2 \right]^\wedge \geq 0.$$  

This shows that $\rho_t \geq 0$.

**Claim:** For all $t > 0$,

$$\int_{\mathbb{R}^n} \rho_t(x) dx = (\chi \cdot \hat{\rho}_t)(0) = \chi(0) < \infty.$$  \hspace{1cm} (25.44)

Given the claim, we have that $d\mu_t(x) = \rho_t(x) dx$ is a finite measure such that $\hat{\mu}_t = \chi \cdot \hat{\rho}_t \to \chi$ as $t \downarrow 0$ so by the continuity Theorem (25.20) 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,

$$\int_{\mathbb{R}^n} \rho_t(x) \hat{\rho}_s(x) dx = \int_{\mathbb{R}^n} (\chi \cdot \hat{\rho}_t) \chi(x) \hat{\rho}_s(x) dx$$

$$= \int_{\mathbb{R}^n} (\chi \cdot \hat{\rho}_s)(\xi) p_s(\xi) d\xi.$$  

We then make use of Fatou’s lemma (using $\rho_t \geq 0$) to show,
which again tends to zero by the dominated convergence theorem because 
\( \lim_{t \downarrow 0} \| \tau_t f - f \|_\infty = 0 \) 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 \( K \subset \subset U \), then
\[
\sup_{x \in K} \| \varphi_t * f(x) - af(x) \|
\leq \int_{B_R} \| f(x - tz) - f(x) \| \varphi(z) dz + \int_{B_R^c} \| f(x - tz) - f(x) \| \varphi(z) dz
\]
\[
\leq \int_{B_R} |\varphi(z)| dz \cdot \sup_{x \in K, z \in B_R} |f(x - tz) - f(x)| + 2 \|f\|_\infty \int_{B_R^c} |\varphi(z)| dz
\leq \|\varphi\|_1 \cdot \sup_{x \in K, z \in B_R} |f(x - tz) - f(x)| + 2 \|f\|_\infty \int_{|z| > R} |\varphi(z)| dz
\]
so that using the uniform continuity of \( f \) on compact subsets of \( U \),
\[
\limsup_{t \downarrow 0} \sup_{x \in K} \| \varphi_t * f(x) - af(x) \| \leq 2 \|f\|_\infty \int_{|z| > R} |\varphi(z)| dz \to 0 \text{ as } R \to \infty.
\]

### 25.8 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
\[
f(z + \Delta) = \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k r_k(z, \Delta)
\]
\[
= \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k \left[ \frac{1}{k!} f^{(k)}(z) + \varepsilon(z, \Delta) \right]
\]
where
\[
r_k(z, \Delta) = \frac{1}{(k - 1)!} \int_0^1 f^{(k)}(z + t\Delta) (1 - t)^{k-1} dt
\]
and
\[
f^{(k)}(z) + \varepsilon(z, \Delta)
\]
\[ \varepsilon(z, \Delta) = \frac{1}{(k-1)!} \int_0^1 \left[ f^{(k)}(z + t\Delta) - f^{(k)}(z) \right] (1-t)^{k-1} \, dt \to 0 \text{ as } \Delta \to 0. \]  

To prove this, use integration by parts to show,

\[ r_k(z, \Delta) = \frac{1}{k!} \int_0^1 f^{(k)}(z + t\Delta) \left( -\frac{d}{dt} \right) (1-t)^k \, dt \]

\[ = -\frac{1}{k!} \left[ f^{(k)}(z + t\Delta) (1-t)^k \right]_{t=0}^{t=1} + \frac{\Delta}{k!} \int_0^1 f^{(k+1)}(z + t\Delta) (1-t)^k \, dt \]

\[ = \frac{1}{k!} f^{(k)}(z) \Delta r_k(z, \Delta), \]

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

\[ |\sin y| \leq |y|. \]  \hfill (25.51)

3. For \( y \in \mathbb{R} \) we have

\[ \cos y = 1 + y^2 \int_0^1 -\cos (ty) (1-t) \, dt \geq 1 + y^2 \int_0^1 -(1-t) \, dt = 1 - \frac{y^2}{2}. \]

Equivalently put

\[ g(y) := \cos y - 1 + \frac{y^2}{2} \geq 0 \text{ for all } y \in \mathbb{R}. \]  \hfill (25.52)

4. Since

\[ |\sin y| = \left| \int_0^y \cos x \, dx \right| \leq \int_0^y |\cos x| \, dx \leq |y| \]

and for \( y \geq 0 \) we have,

\[ \cos y - 1 = \int_0^y -\sin x \, dx \geq \int_0^y -x \, dx = -\frac{y^2}{2}. \]

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 \( y = 0 \).

5. Since \( e^{iy} - 1 = iy \int_0^1 e^{iy} \, dt \), \( |e^{iy} - 1| \leq |y| \) and hence

\[ |e^{iy} - 1| \leq 2 \land |y| \text{ for all } y \in \mathbb{R}. \]  \hfill (25.55)

**Lemma 25.48.** 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 \( e^{\ln z} = z \) and if \( |z| < 1 \) then

\[ |\ln (1+z) - z| \leq |z|^2 \frac{1}{2(1-|z|^2)} \text{ for } |z| < 1. \]  \hfill (25.56)

**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,

\[ \frac{d}{dz} \ln (1+z) = \sum_{n=1}^{\infty} (-z)^n / n \]

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 \( e^{\ln(1+z)} = 1+z \).
We may now conclude that

\[ \ln (1 + z) = z - z^2 \int_0^1 \frac{1}{(1 + tz)^2} (1 - t) \, dt. \]

Eq. (25.57)

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,

\[ \int_0^1 \frac{1}{(1 + tz)^2} (1 - t) \, dt \leq \frac{1}{2 (1 - |z|)^2}. \]  

Eq. (25.58)

Eq. (25.56) is now a consequence of Eq. (25.57) and Eq. (25.58).

**Alternative formulation and proof.** For the purposes of this lemma we could have defined \( \ln (1 + z) \) by

\[ \ln (1 + z) = -\sum_{n=1}^{\infty} (-z)^n / n \]  

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,

\[ f'(z) = \frac{1}{1 + z} f(z) \text{ with } f(0) = 1. \]  

Eq. (25.59)

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. (25.59) 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. (25.56).

Lemma 25.49. For all \( y \in \mathbb{R} \) and \( n \in \mathbb{N} \cup \{0\} \),

\[ |e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \leq \frac{|y|^{n+1}}{(n+1)!} \]

and in particular,

\[ |e^{iy} - 1| \leq |y| \wedge 2 \]

and

\[ e^{iy} - \left( 1 + iy - \frac{y^2}{2!} \right) \leq y^2 \wedge \frac{|y|^3}{3!}. \]

More generally for all \( n \in \mathbb{N} \) we have

\[ |e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \leq \frac{|y|^{n+1}}{(n+1)!} \wedge \frac{2 |y|^n}{n!}. \]

Lemma 25.50. If \( X \) is a square integrable random variable, then

\[ \left| f(\lambda) - \left( 1 + i\lambda EX - \frac{\lambda^2}{2!} E[X^2] \right) \right| \leq E \left| e^{\lambda X} - \left( 1 + i\lambda X - \frac{\lambda^2}{2!} X^2 \right) \right| \leq \lambda^2 \varepsilon (\lambda) \]  

Lemma 25.49
where
\[ \varepsilon(\lambda) := E \left[ X^2 \wedge \frac{|\lambda| |X|^3}{3!} \right] \to 0 \text{ as } \lambda \to 0. \] (25.65)

**Proof.** Using Eq. (25.62) with \( y = \lambda X \) and taking expectations gives Eq. (25.64). 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

\[
S_n := \sum_{k=1}^{n} X_{n,k} \quad \text{and} \quad f_n(k) := \mathbb{E}[e^{i\lambda X_{n,k}}] \quad \text{(characteristic function of } X_{n,k}).
\]  

(26.1)

The goal of this chapter is to discuss some of the possible weak limits of such \( S_n \) 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 26.1.** A probability distribution, \( \mu \), on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) is **infinitely divisible** if for all \( n \in \mathbb{N} \) there exists i.i.d. nondegenerate random variables, \( \{X_{n,k}\}_{k=1}^{n} \), such that \( X_{n,1} + \cdots + X_{n,n} \overset{d}{=} \mu \). This can be formulated in the following two equivalent ways. For all \( n \in \mathbb{N} \) there should exist a non-degenerate probability measure, \( \mu_n \), on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) such that \( \mu_n^n = \mu \). For all \( n \in \mathbb{N} \), \( \mu(\lambda) = [g_n(\lambda)]^n \) for some non-constant characteristic function, \( g_n \).

**Remark 26.2.** 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 26.3.** Suppose that \( \mu \) is a probability measure on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) and \( \text{Law}(X) = \mu \). Then \( \mu \) is infinitely divisible if there exist an array, \( \{X_{n,k} : 1 \leq k \leq m_n\} \) with \( \{X_{n,k}\}_{k=1}^{m_n} \) being i.i.d. such that \( \sum_{k=1}^{m_n} X_{n,k} \rightarrow X \) and \( m_n \rightarrow \infty \) as \( n \rightarrow \infty \).

**Proof.** The only non-trivial direction is \((\Rightarrow)\). 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, \( S_{nk} = \sum_{i=1}^{k} S_i^{nk} \), where

\[
S_i^{nk} = \sum_{j=k(i-1)+1}^{ki} X_{n,j}.
\]

Notice that \( \{S_i^{nk}\}_{i=1}^{k} \) are i.i.d. for each \( n \in \mathbb{N} \) and since \( S_{nk} \rightarrow X \) as \( n \rightarrow \infty \) we know that \( \{S_{nk}\}_{n=1}^{\infty} \) is tight and there exists \( \varepsilon(r) \downarrow 0 \) as \( r \uparrow \infty \) such that

\[
P(|S_{nk}| > r) \leq \varepsilon(r).
\]

Since

\[
P(S_i^{nk} > r) \leq P(S_{nk} > kr) \leq P(|S_{nk}| > kr) \leq \varepsilon(kr)
\]

and similarly,

\[
P(-S_i^{nk} > r) \leq P(-S_{nk} > kr) \leq P(|S_{nk}| > kr) \leq \varepsilon(kr)
\]

we see that \( P(|S_i^{nk}| > r) \leq 2\varepsilon(kr)^{1/k} \rightarrow 0 \) as \( r \uparrow \infty \) which shows that \( \{S_i^{nk}\}_{i=1}^{\infty} \)

has tight distributions as well.

Thus there exists a subsequence \( \{n_l\} \) such that \( S_{ni} \rightarrow Y \) as \( l \rightarrow \infty \). Let \( \{Y_i\}_{i=1}^{k} \) be i.i.d. random variables with \( Y_i \overset{d}{=} Y \). Then by Exercise \[24.12\] it follows that

\[
S_{kn_l} = \sum_{i=1}^{k} S_i^{nk_l} \Rightarrow Y_1 + \cdots + Y_k
\]

from which we conclude that \( X \overset{d}{=} Y_1 + \cdots + Y_k \). Since \( k \) was arbitrary we have shown \( X \) is infinitely divisible.

**Remark 26.4.** [See Theorem 26.25 and Corollary 26.27 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 Var(\( X_{n,k} \)) of condition (M) of Definition 26.9 below.]

The Lévy Kintchine formula of Theorem 26.7 below asserts that \( \mu \in \mathcal{P}_1(\mathbb{R}, \mathcal{B}_\mathbb{R}) \) is infinitely divisible iff \( \mu(\lambda) = e^{\psi(\lambda)} \) where \( \psi(\lambda) \) has the form given
in Eq. (26.4) below. [A more restrictive class of distributions are the so called stable distributions, see Definition 26.29, Lemma 26.33 and Theorem 26.34]

Before stating Lévy Kintchine formula it is worth recording a couple of examples.

**Example 26.5 (Following Theorem 26.33).** Let $\mu \in \mathcal{P}_1 (\mathbb{R}, \mathcal{B}_\mathbb{R})$, \{\(Z_n\)\}_{n=1}^{\infty} be i.i.d. random variables with \(Z_n \overset{d}{=} \mu\), \(N_\alpha \overset{d}{=} \text{Poi} (\alpha)\), and \(Y \overset{d}{=} \sigma N (0, 1) + c\) all be chosen so that \(\{Z_n\}_{n=1}^{\infty} \cup \{N_\alpha, Y\}\) are independent. Then \(S := Y + \sum_{n \leq N_\alpha} Z_n\) is infinitely divisible. Indeed we have

\[
\begin{align*}
\hat{f}_S (\lambda) &= \mathbb{E} \left[ e^{i\lambda S} \right] = \mathbb{E} \left[ e^{i\lambda Y} \cdot e^{i\lambda \sum_{k \leq N_\alpha} Z_k} \right] \\
&= \mathbb{E} \left[ e^{i\lambda Y} \right] \sum_{k=0}^{\infty} \mathbb{E} \left[ e^{i\lambda \sum_{k \leq N_\alpha} Z_k} | N_\alpha = n \right] P (N_\alpha = n) \\
&= \mathbb{E} \left[ e^{i\lambda Y} \right] \sum_{n=0}^{\infty} \mathbb{E} \left[ e^{i\lambda (Z_1 + \ldots + Z_n)} \right] P (N_\alpha = n) \\
&= \exp \left( -\frac{1}{2} \sigma^2 \lambda^2 + i\lambda \mu \right) \sum_{n=0}^{\infty} e^{-n} \frac{\alpha^n}{n!} \hat{\mu} (\lambda)^n \\
&= \exp \left( -\frac{1}{2} \sigma^2 \lambda^2 + i\lambda \mu + \alpha (\hat{\mu} (\lambda) - 1) \right) = e^{\psi (\lambda)}
\end{align*}
\]

where

\[
\psi (\lambda) = -\frac{1}{2} \sigma^2 \lambda^2 + i\lambda \mu + \int_{\mathbb{R}} \left( e^{i\lambda x} - 1 \right) d\nu (x)
\]

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 \(e^{\psi (\lambda)}\) 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 \(X_\alpha := \sum_{n \leq N_\alpha} Z_n\) 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 26.6 (Exercise 25.7).** Recall from Exercise 25.5 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

\[
\psi (\lambda) = \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu (x).
\]

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. (26.3) as

\[
\psi (\lambda) = i\lambda b + \int_{\mathbb{R}} \left( e^{i\lambda x} - 1 - i\lambda x \mathbf{1}_{|x| \leq 1} \right) d\tilde{\nu} (x)
\]

where

\[
b = -\int_{|x| > 1} x \mathbf{1}_{|x| > 1} d\tilde{\nu} (x) = -\int_{|x| > 1} \frac{1}{|x|} d\nu (x) \in \mathbb{R}
\]

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 26.7 (Lévy Kintchine formula).** A probability measure \(\mu\) on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) is infinitely divisible if \(\hat{\mu} (\lambda) = e^{\psi (\lambda)}\) where

\[
\psi (\lambda) = i\lambda b - \frac{1}{2} a \lambda^2 + \int_{\mathbb{R} \setminus \{0\}} \left( e^{i\lambda x} - 1 - i\lambda x \mathbf{1}_{|x| \leq 1} \right) d\nu (x)
\]

for some \(b \in \mathbb{R}\), \(a \geq 0\), and some measure \(\nu\) on \(\mathbb{R} \setminus \{0\}\) such that

\[
\int_{\mathbb{R} \setminus \{0\}} (x^2 \wedge 1) d\nu (x) < \infty.
\]

[Note that the term \(-\frac{1}{2} a \lambda^2\) in Eq. (26.4) the logarithm of the characteristic function of \(N (0, \sqrt{a})\).]

**Proof.** We will give the easy direction of this proof, namely the implication \((\Leftarrow).\) To summarize we want to show if \(\psi\) is of the form in Eq. (26.4), 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. (26.4) 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

\[
b \equiv b - \int_{|x| \leq 1} x d\nu (x).
\]

Thus we may use Example 26.5 in order to construct a random variable with distribution given by \(\mu\).

For general \(\nu\) satisfying Eq. (26.5) let, for \(\varepsilon > 0\),

\[
d\nu^{\varepsilon} (x) := 1_{|x| \geq \varepsilon} d\nu (x) - \text{a finite measure on } \mathbb{R}
\]

Thus by Example 26.5 there exits a probability measure, \(\mu,\) on \((\mathbb{R}, \mathcal{B}_\mathbb{R})\) such that \(\hat{\mu} = e^{\psi}\) where
Let \( \psi_\varepsilon (\lambda) := \varepsilon a - \frac{1}{2} a \varepsilon^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\leq 1}) 1_{|x|\geq \varepsilon} d\nu (x) . \)

As \( e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\leq 1} \) is bounded and less than \( C(\lambda) x^2 \) for \( |x| \leq 1 \), we may use DCT to show \( \psi_\varepsilon (\lambda) \rightarrow \psi (\lambda) \). Furthermore the DCT also shows \( \psi (\lambda) \) is continuous and therefore \( e^{\psi (\lambda)} \) is continuous. Thus we have shown that \( \mu_{\varepsilon} \rightarrow e^{\psi} \) where the limit is continuous and therefore by the continuity Theorem 25.20 there exists a probability measure \( \mu \) on \( (\mathbb{R}, B(\mathbb{R})) \) such that \( \bar{\mu} = e^{\psi} \).

The proof of the other implication \( (\Rightarrow) \) will be discussed in Appendix 26.4 below. For more information about Poisson processes and Lévy processes, see Protter [26, Chapter 1], [10] Chapter 9.5, and [15] Chapter XVII.2, p. 558–562 for analytic proofs. Also see http://www.math.ucsd.edu/~bass/scldp.pdf, Kallenberg [29, Theorem 15.13, p. 254], [1], [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. (26.1) in the limit as \( n \rightarrow \infty \).

**Assumption 2** Assume \( \mathbb{E} [X_{n,k}] = 0, \sigma^2_{n,k} = \mathbb{E} [X_{n,k}^2] < \infty \), and \( \text{Var} (S_n) = \sum_{k=1}^n \sigma^2_{n,k} = 1 \).

**Example 26.8.** Suppose \( \{X_n\}_{n=1}^\infty \) are mean zero square integrable random variables with \( \sigma^2_n = \text{Var} (X_n) \). If we let \( s^2_n := \sum_{k=1}^n \text{Var} (X_k) = \sum_{k=1}^n \sigma^2_k \), \( \sigma^2_{n,k} := \sigma^2_k / s^2_n \), and \( X_{n,k} := X_k / s_n \), then \( \{X_{n,k}\}_{k=1}^n \) satisfy the above hypothesis and \( S_n = \frac{1}{s_n} \sum_{k=1}^n X_{n,k} \).

The next definition records some possible meanings of small and comparable in size.

**Definition 26.9.** Let \( \{X_{n,k}\} \) be as above.

1. \( \{X_{n,k}\} \) satisfies the **Lindeberg Condition (LC)** iff

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} [X_{n,k}^2 : |X_{n,k}| > t] = 0 \quad \text{for all} \ t > 0.
\]

   [Since \( \sum_{k=1}^n \mathbb{E} [X_{n,k}^2 : |X_{n,k}| > t] \) 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

\[
D_n := \max \{\sigma^2_{n,k} : k \leq n \} \to 0 \quad \text{as} \ n \to \infty.
\]

3. \( \{X_{n,k}\} \) is uniformly asymptotic negligibility (UAN) if for all \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \max_{k \leq n} P (|X_{n,k}| > \varepsilon) = 0.
\]

Each of these conditions imposes constraints on the size of the tails of the \( \{X_{n,k}\} \), see Lemma 26.13 below where it is shown \( (LC) \implies (M) \implies (UAN) \). Condition (M) asserts that all of the terms in the sum \( \sum_{k=1}^n \sigma^2_{n,k} = \text{Var} (S_n) = 1 \) are small so that no one term is contributing by itself.

**Remark 26.10.** The reader should observe that in order for condition (M) to hold in the setup of Example 26.8 it is necessary that \( \lim_{n \to \infty} s_n^2 = \infty \).

**Example 26.11.** Suppose \( \{X_n\}_{n=1}^\infty \) are i.i.d. with \( \mathbb{E} X_n = 0 \) and \( \text{Var} (X_n) = \sigma^2 \). Then \( \left\{X_{n,k} := \frac{1}{\sqrt{n\sigma}} X_k\right\}_{k=1}^n \) satisfy (LC). Indeed,

\[
\frac{1}{n} \sum_{k=1}^n \mathbb{E} [X_{n,k}^2 : |X_{n,k}| > t] = \frac{1}{n\sigma^2} \sum_{k=1}^n \mathbb{E} \left[ X_k^2 : \left| \frac{X_k}{\sqrt{n\sigma}} \right| > t \right]
\]

\[
= \frac{1}{\sigma^2} \mathbb{E} [X_1^2 : |X_1| > \sqrt{n\sigma} t]
\]

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 3 the asymptotic normality results (proved in the next section) may be summarized as follows.

1. The Lindeberg-Feller CLT Theorem 26.15 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 26.20.

### 26.1 Lindeberg-Feller CLT

In this section we will assume that Assumption 2 is in place, i.e. that

\[
\mathbb{E} [X_{n,k}] = 0, \sigma^2_{n,k} = \mathbb{E} [X_{n,k}^2] < \infty, \quad \text{and}
\]

\[
\text{Var} (S_n) = \sum_{k=1}^n \sigma^2_{n,k} = 1.
\]
Lemma 26.12. Let \( \{X_{n,k}\}_{k=1}^{n} \) for \( n \in \mathbb{N} \) be as in Assumption 3. If \( \{X_{n,k}\}_{k=1}^{n} \) satisfy the Liapunov condition (LiapC):

\[
\lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E} \left| X_{n,k} \right|^\alpha = 0 \quad \text{for some } \alpha > 2 \tag{26.9}
\]

then (LC) holds. More generally, if \( \{X_{n,k}\} \) satisfies the Liapunov condition,

\[
\lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 \varphi (|X_{n,k}|) \right] = 0
\]

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. (26.9), then for any \( t > 0 \),

\[
\sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 : |X_{n,k}| > t \right] \leq \sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 \left| \frac{X_{n,k}}{t} \right|^{\alpha - 2} : |X_{n,k}| > t \right] \\
\leq \frac{1}{t^{\alpha - 2}} \sum_{k=1}^{n} \mathbb{E} \left[ |X_{n,k}|^\alpha \right] \\
= \frac{1}{t^{\alpha - 2}} \sum_{k=1}^{n} \mathbb{E} \left[ |X_{n,k}|^\alpha \right] \to 0 \quad \text{as } n \to \infty.
\]

The generalization is proved similarly;

\[
\sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 : |X_{n,k}| > t \right] \leq \sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 \varphi (|X_{n,k}|) : |X_{n,k}| > t \right] \\
\leq \frac{1}{\varphi (t)} \sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^2 \varphi (|X_{n,k}|) \right] \to 0 \quad \text{as } n \to \infty.
\]

\[\blacksquare\]

Lemma 26.13. Let \( \{X_{n,k} : 1 \leq k \leq n < \infty \} \) be as above, then (LiapC) \( \implies \) (LC) \( \implies \) (M) \( \implies \) (UAN). Moreover the Lindeberg Condition (LC) implies the following strong form of (UAN),

\[
\sum_{k=1}^{n} P \left[ |X_{n,k}| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \mathbb{E} \left[ |X_{n,k}|^2 : |X_{n,k}| > \varepsilon \right] \to 0. \tag{26.10}
\]

Proof. The assertion, (LiapC) \( \implies \) (LC), was proved in Lemma 26.12.

For \( k \leq n \),

\[
\sigma_{n,k}^2 = \mathbb{E} \left[ X_{n,k}^2 \right] = \mathbb{E} \left[ X_{n,k}^2 1_{|X_{n,k}| \leq t} \right] + \mathbb{E} \left[ X_{n,k}^2 1_{|X_{n,k}| > t} \right] \\
\leq t^2 + \mathbb{E} \left[ X_{n,k}^2 1_{|X_{n,k}| > t} \right] \leq t^2 + \sum_{m=1}^{n} \mathbb{E} \left[ X_{n,m}^2 1_{|X_{n,m}| > t} \right]
\]

and therefore using (LC) we find

\[
\lim_{n \to \infty} D_n := \lim_{n \to \infty} \max_{1 \leq k \leq n} \sigma_{n,k}^2 \leq t^2 \quad \text{for all } t > 0.
\]

This clearly implies (M) holds. For \( \varepsilon > 0 \) we have by Chebyshev’s inequality that

\[
P (|X_{n,k}| > \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E} \left[ |X_{n,k}|^2 : |X_{n,k}| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sigma_{n,k}^2 \tag{26.11}
\]

and therefore,

\[
\max_{1 \leq k \leq n} P (|X_{n,k}| > \varepsilon) \leq \frac{1}{\varepsilon^2} \max_{k \leq n} \sigma_{n,k}^2 = \frac{1}{\varepsilon^2} D_n \to 0 \quad \text{as } n \to \infty
\]

which shows (M) \( \implies \) (UAN). Summing Eq. (26.11) on \( k \) gives Eq. (26.10) and the right member of this equation tends to zero as \( n \to \infty \) if (LC) holds. \[\blacksquare\]

We will need the following lemma for our subsequent applications of the continuity theorem.

Lemma 26.14. Suppose that \( a_i, b_i \in \mathbb{C} \) with \( |a_i|, |b_i| \leq 1 \) for \( i = 1, 2, \ldots, n \).

Then

\[
\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i \leq \sum_{i=1}^{n} |a_i - b_i|.
\]

Proof. Let \( a := \prod_{i=1}^{n-1} a_i \) and \( b := \prod_{i=1}^{n-1} b_i \) and observe that \( |a|, |b| \leq 1 \) and that

\[
|a_n a - b_n b| \leq |a_n a - a_n b| + |a_n b - b_n b| \\
= |a_n| |a - b| + |a_n - b_n| |b| \\
\leq |a - b| + |a_n - b_n|.
\]

The proof is now easily completed by induction on \( n \). \[\blacksquare\]

Theorem 26.15 (Lindeberg-Feller CLT (I)). Suppose \( \{X_{n,k}\} \) satisfies (LC) and the hypothesis in Assumption 3 then

\[
S_n \implies N (0, 1).
\tag{26.12}
\]

(See Theorem 26.20 for a converse to this theorem.)
To prove this theorem we must show
\[ \mathbb{E} [e^{i\lambda S_n}] \to e^{-\lambda^2/2} \text{ as } n \to \infty. \] (26.13)

Before starting the formal proof, let me give an informal explanation for Eq. (26.13). Using
\[ \sum_{k=1}^{n} \ln f_n(k) - (\sum_{k=1}^{n} (f_n(k) - 1)) \]
we might expect
\[ \mathbb{E} [e^{i\lambda S_n}] = \prod_{k=1}^{n} f_n(k) = e^{\sum_{k=1}^{n} \ln f_n(k)} = e^{\sum_{k=1}^{n} \ln(1 + f_n(k) - 1)} \approx e^{\sum_{k=1}^{n} (f_n(k) - 1)} \approx e^{-\frac{\lambda^2}{2} \sigma_{n,k}^2}. \]

The question then becomes under what conditions are these approximations valid. It turns out that approximation (A), namely that
\[ \lim_{n \to \infty} \left| \prod_{k=1}^{n} f_n(k) - \exp\left(\sum_{k=1}^{n} (f_n(k) - 1)\right) \right| = 0, \] (26.14)
is valid if condition (M) holds, see Lemma 26.18 below and the approximation (B) is valid, i.e.
\[ \lim_{n \to \infty} \sum_{k=1}^{n} (f_n(k) - 1) = -\frac{1}{2} \lambda^2, \]
if (LC) is satisfied, see Lemma 26.16. These observations would then constitute a proof of Theorem 26.15. The proof we give below of Theorem 26.15 will not quite follow this route and will not use Lemma 26.18 directly. However, this lemma will be used in the proofs of Theorems 26.20 and 26.25.

**Proof.** (Proof of Theorem 26.15) Since
\[ \mathbb{E} [e^{i\lambda S_n}] = \prod_{k=1}^{n} f_n(k) \] and \( e^{-\lambda^2/2} = \prod_{k=1}^{n} e^{-\lambda^2 \sigma_{n,k}^2/2}, \)
we may use Lemma 26.14 to conclude,
\[ \left| \mathbb{E} [e^{i\lambda S_n}] - e^{-\lambda^2/2} \right| \leq \sum_{k=1}^{n} \left| f_n(k) - e^{-\lambda^2 \sigma_{n,k}^2/2} \right| = \sum_{k=1}^{n} (A_{n,k} + B_{n,k}) \]
where
\[ A_{n,k} := \left| f_n(k) - 1 + \frac{\lambda^2 \sigma_{n,k}^2}{2} \right| \text{ and } \]
\[ B_{n,k} := \left| 1 - \frac{\lambda^2 \sigma_{n,k}^2}{2} - e^{-\lambda^2 \sigma_{n,k}^2/2} \right|. \] (26.16)

Because of Lemma 26.16 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, \( |e^{-u} - 1 + u| \leq u^2/2 \) valid for \( u \geq 0 \) (see Eq. (25.54) with \( z = -u \)). With this estimate we find,
\[ \sum_{k=1}^{n} B_{n,k} \leq \sum_{k=1}^{n} \left[ 1 - \frac{\lambda^2 \sigma_{n,k}^2}{2} \right] - e^{-\lambda^2 \sigma_{n,k}^2/2} \]
\[ \leq \sum_{k=1}^{n} \frac{1}{2} \left( \frac{\lambda^2 \sigma_{n,k}^2}{2} \right)^2 = \frac{\lambda^4}{8} \sum_{k=1}^{n} \sigma_{n,k}^2 \]
\[ \leq \frac{\lambda^4}{8} \max_{k \leq n} \sigma_{n,k}^2 \sum_{k=1}^{n} \sigma_{n,k}^2 = \frac{\lambda^4}{8} \max \sigma_{n,k}^2 \to 0, \]
wheren we have used (M) (which is implied by (LC)) in taking the limit as \( n \to \infty. \)

**Lemma 26.16.** Let \( A_{n,k} \) be as in Eq. (26.15). If \( \{X_{n,k}\}_{k=1}^{n} \) satisfies (LC), then
\[ \lim_{n \to \infty} \sup_{k} A_{n,k} = 0 \] and
\[ \lim_{n \to \infty} \sum_{k=1}^{n} (f_n(k) - 1) = -\lambda^2/2 \text{ for all } \lambda \in \mathbb{R}. \]

**Proof.** Rewriting \( A_{n,k} \) using \( \mathbb{E} X_{n,k} = 0 \) and then using Lemma 25.50 implies for every \( \varepsilon > 0 \) that,
The second limit follows from the first and the simple estimate;
\[
23.42. 
\]

Summing this equation on \(k\) and making use of (LC) gives;
\[
\limsup_{n \to \infty} \sum_{k=1}^{n} A_{n,k} \leq \frac{\lambda^3 \varepsilon}{6} \to 0 \quad \text{as} \quad \varepsilon \downarrow 0. \tag{26.17}
\]

The second limit follows from the first and the simple estimate;
\[
\left| \sum_{k=1}^{n} (f_{nk}(\lambda) - 1) + \lambda^2 / 2 \right| = \sum_{k=1}^{n} \left( f_{nk}(\lambda) - 1 + \frac{\lambda^2 \sigma_{n,k}^2}{2} \right) \leq \sum_{k=1}^{n} A_{n,k}.
\]

As an application of Theorem 26.15 we can give half of the proof of Theorem 23.42

**Theorem 26.17 (Converse assertion in Theorem 23.42).** 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. \( \sum_{n=1}^{\infty} P\left( |X_n| > c \right) < \infty \),
2. \( \sum_{n=1}^{\infty} \text{Var} \left( X_n 1_{|X_n| \leq c} \right) < \infty \), and
3. \( \sum_{n=1}^{\infty} \text{Cov} \left( X_n 1_{|X_n| \leq c} \right) \) 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 \( c > 0 \), \( P\left( |X_n| \geq c \right) \to 0 \). According to the Borel zero one law (Lemma 13.47) this implies for every \( c > 0 \) that \( \sum_{n=1}^{\infty} P\left( |X_n| > c \right) < \infty \). Since \( X_n \to 0 \) a.s., \( \{X_n\} \) and \( \{X_n \mid X_n \leq c\} \) 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 \( s_n^2 := \text{Var}(Y_1 + \cdots + Y_n) = \sum_{k=1}^{n} \text{Var}(Y_k) = \sum_{k=1}^{n} \text{Var}(X_k^c) = \sum_{k=1}^{n} \text{Var}(X_k 1_{|X_k| \leq c}) \).

For the sake of contradictions, suppose \( s_n^2 \to \infty \) as \( n \to \infty \). Since \( |Y_k| \leq 2c \), it follows that \( \sum_{k=1}^{n} \mathbb{E}[|Y_k|^2 | |Y_k| > s_n t] = 0 \) for all sufficiently large \( n \) and hence
\[
\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[|Y_k|^2 | |Y_k| > s_n t] = 0,
\]

i.e. \( \{Y_{n,k} := Y_k / s_n\}_{n=1}^{\infty} \) satisfies (LC) – see Examples 26.8 and Remark 26.10. So by the central limit Theorem 26.15 it follows that
\[
\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} (X_n^c - \mathbb{E}[X_n^c]) = \frac{1}{s_n^2} \sum_{k=1}^{n} Y_k \implies N(0,1).
\]

On the other hand we know
\[
\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} X_n^c = \frac{\sum_{n=1}^{\infty} X_k^c}{\lim_{n \to \infty} s_n^2} = 0 \text{ a.s.}
\]

and so by Slutsky’s theorem,
\[
\frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[X_n^c] = \frac{1}{s_n^2} \sum_{k=1}^{n} X_k^c - \frac{1}{s_n^2} \sum_{k=1}^{n} Y_k \implies N(0,1).
\]

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
\[
\sum_{n=1}^{\infty} \text{Var}(X_n 1_{|X_n| \leq c}) = \sum_{n=1}^{\infty} \text{Var}(X_n^c) = \lim_{n \to \infty} s_n^2 < \infty.
\]

An application of Kolmogorov’s convergence criteria (Theorem 23.11) 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 \( \sum_{n=1}^{\infty} \mathbb{E}[X_n 1_{|X_n| \leq c}] \) is convergent.

Let us now turn to the converse of Theorem 26.15 see Theorem 26.20 below.
Lemma 26.18. Suppose that \{X_{n,k}\} satisfies property (M), i.e. \(D_n := \max_{k \leq n} \sigma_{n,k} \to 0\). If we define,

\[ \varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1 = \mathbb{E}[e^{i\lambda X_{n,k}} - 1], \]

then:

1. \( \lim_{n \to \infty} \max_{k \leq n} |\varphi_{n,k}(\lambda)| = 0 \) and
2. \( f_{n}(\lambda) - \prod_{k=1}^{\infty} e^{\varphi_{n,k}(\lambda)} \to 0 \) as \( n \to \infty \), where

\[ f_{n}(\lambda) = \mathbb{E}[e^{i\lambda S_n}] = \prod_{k=1}^{n} f_{n,k}(\lambda). \]

Proof. For any \( \varepsilon > 0 \) we have, making use of Eq. (25.61) and Chebyshev’s inequality, that

\[ |\varphi_{n,k}(\lambda)| = |f_{n,k}(\lambda) - 1| \leq \mathbb{E}|e^{i\lambda X_{n,k}} - 1| \leq \mathbb{E}[2 \wedge |\lambda X_{n,k}|] \leq \mathbb{E}[2 \wedge |\lambda X_{n,k}| : |X_{n,k}| \geq \varepsilon] + \mathbb{E}[2 \wedge |\lambda X_{n,k}| : |X_{n,k}| < \varepsilon] \leq 2P[|X_{n,k}| \geq \varepsilon] + |\lambda| \varepsilon \leq \frac{2\sigma_{n,k}^2}{\varepsilon^2} + |\lambda| \varepsilon. \]

Therefore,

\[ \limsup_{n \to \infty} \max_{k \leq n} |\varphi_{n,k}(\lambda)| \leq \limsup_{n \to \infty} \left[ \frac{2D_n}{\varepsilon^2} + |\lambda| \varepsilon \right] = |\lambda| \varepsilon \to 0 \text{ as } \varepsilon \downarrow 0. \]

For the second item, observe that \( \text{Re} \varphi_{n,k}(\lambda) = \text{Re} f_{n,k}(\lambda) - 1 \leq 0 \) and hence \( e^{\varphi_{n,k}(\lambda)} = e^{\text{Re} \varphi_{n,k}(\lambda)} \leq 1 \). Therefore by Lemma 26.14 and the estimate (25.54) we find:

\[ \left| \prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)} - \prod_{k=1}^{n} f_{n,k}(\lambda) \right| \leq \sum_{k=1}^{n} |e^{\varphi_{n,k}(\lambda)} - f_{n,k}(\lambda)| \leq \frac{1}{2} \sum_{k=1}^{n} |\varphi_{n,k}(\lambda)|^2 \leq \frac{1}{2} \max_{k \leq n} |\varphi_{n,k}(\lambda)| \sum_{k=1}^{n} |\varphi_{n,k}(\lambda)|. \]

Since \( \mathbb{E}X_{n,k} = 0 \) we may write express \( \varphi_{n,k}(\lambda) \)

\[ \varphi_{n,k}(\lambda) = \mathbb{E}[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}] \]

and then using estimate in Eq. (25.54) again shows

\[ \sum_{k=1}^{n} |\varphi_{n,k}(\lambda)| = \sum_{k=1}^{n} |\mathbb{E}[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}]| \leq \sum_{k=1}^{n} \left| \mathbb{E}\left[ \frac{1}{2} |\lambda X_{n,k}|^2 \right] \right| \leq \frac{\lambda^2}{2} \sum_{k=1}^{n} \sigma_{n,k}^2 = \frac{\lambda^2}{2}. \]

Thus we have shown,

\[ \prod_{k=1}^{n} f_{n,k}(\lambda) - \prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)} \leq \frac{\lambda^2}{4} \max_{k \leq n} |\varphi_{n,k}(\lambda)| \]

and the latter expression tends to zero by item 1.

Lemma 26.19. Let \( X \) be a random variable such that \( \mathbb{E}X^2 < \infty \) and \( \mathbb{E}X = 0 \). Further let \( f(\lambda) := \mathbb{E}[e^{i\lambda X}] \) and \( u(\lambda) := \text{Re}(f(\lambda)) - 1 \). Then for all \( c > 0 \),

\[ u(\lambda) + \frac{\lambda^2}{2} \mathbb{E}[X^2] \geq \mathbb{E}\left[ X^2 \left( \frac{\lambda^2}{2} - \frac{2c^2}{c^2} \right) : |X| > c \right]. \]

or equivalently

\[ \mathbb{E}\left[ \cos \lambda X - 1 + \frac{\lambda^2}{2} X^2 \right] \geq \mathbb{E}\left[ X^2 \left( \frac{\lambda^2}{2} - \frac{2c^2}{c^2} \right) : |X| > c \right]. \]

In particular if we choose \( |\lambda| \geq \sqrt{6}/|c| \), then

\[ \mathbb{E}\left[ \cos \lambda X - 1 + \frac{\lambda^2}{2} X^2 \right] \geq \frac{1}{c^2} \mathbb{E}[X^2 : |X| > c]. \]

Proof. For all \( \lambda \in \mathbb{R} \), we have (see Eq. (25.52)) \( \cos \lambda X - 1 + \frac{\lambda^2}{2} X^2 \geq 0 \) and \( \cos \lambda X - 1 \geq -2. \) Therefore,

\[ u(\lambda) + \frac{\lambda^2}{2} \mathbb{E}[X^2] = \mathbb{E}\left[ \cos \lambda X - 1 + \frac{\lambda^2}{2} X^2 \right] \geq \mathbb{E}\left[ \cos \lambda X - 1 + \frac{\lambda^2}{2} X^2 \right] : |X| > c \]

\[ \geq \mathbb{E}\left[ -2 + \frac{\lambda^2}{2} X^2 : |X| > c \right] \geq \mathbb{E}\left[ -2 \frac{|X|^2}{c^2} + \frac{\lambda^2}{2} X^2 : |X| > c \right] \]

which gives Eq. (26.18).
Proposition 26.21. Suppose \( \{X_{n,k}\} \) satisfies (M) and \( S_n \Rightarrow N(0,1) \) (i.e. the central limit theorem in Eq. (26.12) 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

\[
\lim_{n \to \infty} \max_{0 \leq k \leq n} \sigma_{n,k}^2 = 0 \quad \text{and} \quad \lim_{n \to \infty} \prod_{k=1}^{n} f_n(k) = e^{-\lambda^2/2}.
\]

The second inequality combined with Lemma 26.18 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,

\[
\lim_{n \to \infty} e^{\sum_{k=1}^{n} \Re \varphi_{n,k}(\lambda)} \leq \lim_{n \to \infty} \left| e^{\sum_{k=1}^{n} \varphi_{n,k}(\lambda)} \right| = e^{-\lambda^2/2}
\]

from which we may conclude

\[
\lim_{n \to \infty} \sum_{k=1}^{n} \Re \varphi_{n,k}(\lambda) = -\lambda^2/2.
\]

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 26.19 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 26.15 let us see what it has to say about Brownian motion. In what follows we say that \( \{B_t\}_{t \geq 0} \) 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 27.

Proposition 26.21. Suppose \( \{B_t\}_{t \geq 0} \) is a stochastic process on some probability space, \( (\Omega, \mathcal{B}, \mathbb{P}) \) such that;

1. \( B_0 = 0 \) a.s., \( EB_t = 0 \) for all \( t \geq 0 \),
2. \( \mathbb{E}(B_t - B_s)^2 = t - s \) for all \( 0 \leq s \leq t < \infty \),
3. \( B \) has independent increments, i.e. if \( 0 = t_0 < t_1 < \cdots < t_n < \infty \), then \( \{B_{t_j} - B_{t_{j-1}}\}_{j=1}^{n} \) are independent random variables.
4. (Moment Condition) There exists \( p > 2, q > 1 \) and \( c < \infty \) such that \( \mathbb{E}|B_t - B_s|^p \leq c|t-s|^q \) for all \( s, t \in \mathbb{R}^+ \).

Then \( B_t - B_s \overset{d}{=} N(0, t-s) \) for all \( 0 \leq s < t < \infty \). We call such a process satisfying these conditions a pre-Brownian motion.

Proof. Let \( 0 \leq s < t \) and for each \( n \in \mathbb{N} \) and \( 1 \leq k \leq n \) 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,

\[
\sum_{k=1}^{n} \mathbb{E} \left[ X_{n,k}^p \right] \leq c \sum_{k=1}^{n} \left( \frac{t-s}{n} \right)^q = c(t-s)^q \frac{n}{n^q} \to 0 \quad \text{as} \quad n \to \infty.
\]

Thus we have shown that \( \{X_{n,k}\} \) satisfies a Liapunov condition which by Lemma 26.12 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 26.15.

Remark 26.22 (Poisson Process). There certainly are other processes satisfying items 1.-3. other than a pre-Brownian motion. Indeed, if \( \{N_t\}_{t \geq 0} \) is a Poisson process with intensity \( \lambda \) (see Example 28.9), then \( B_t := \lambda^{-1} N_t - t \) satisfies the items 1.-3. above. Recall that \( \mathbb{V}(N_t - N_s) = (\lambda t - \lambda s)^2 \) and \( \mathbb{E}(N_t - N_s) = \lambda (t-s) \) so \( \mathbb{E}B_t = 0 \) and

\[
\mathbb{E}(B_t - B_s)^2 = \mathbb{V}(B_t - B_s) = \mathbb{V}(\lambda^{-1} (N_t - N_s)) = \lambda^{-2} \mathbb{V}(N_t - N_s) = \lambda^{-2} (\lambda t - \lambda s)^2 = t-s.
\]

In this case one can show that \( \mathbb{E} \|B_t - B_s\|^p \sim |t-s| \) for all \( 1 \leq p < \infty \).

This last remark leads us to our next topic.

26.2 More on Infinitely Divisible Distributions

In 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 26.23 (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. \( \lim_{n \to \infty} \max_{1 \leq k \leq n} p_{n,k} = 0 \). (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 24.10 while the second uses Fourier transform methods.

**First proof.** Let \( Z_n \sim \text{Poi} \left( \sum_{k=1}^{n} p_{n,k} \right) \), then by Theorem 24.10 we know that

\[
\text{TV}(Z_n, S_n) \leq \sum_{k=1}^{n} p_{n,k} \leq \max_{1 \leq k \leq n} p_{n,k} \cdot \sum_{k=1}^{n} p_{n,k}.
\]

From the assumptions it follows that \( \lim_{n \to \infty} \text{TV}(Z_n, S_n) = 0 \) and from part 3. of Exercise 24.6 we know that \( \lim_{n \to \infty} \text{TV}(Z_n, Z) = 0 \). Therefore, \( \lim_{n \to \infty} \text{TV}(Z, S_n) = 0 \).

**Second proof.** Recall from Example 25.11 that for any \( a > 0 \),

\[
\mathbb{E}[e^{i\lambda Z}] = \exp \left( a \left( e^{i\lambda} - 1 \right) \right).
\]

Since

\[
\mathbb{E}[e^{i\lambda Y_{n,k}}] = e^{i\lambda} p_{n,k} + (1 - p_{n,k}) = 1 + p_{n,k} \left( e^{i\lambda} - 1 \right),
\]

it follows that

\[
\mathbb{E}[e^{i\lambda S_n}] = \prod_{k=1}^{n} \left[ 1 + p_{n,k} \left( e^{i\lambda} - 1 \right) \right].
\]

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 26.1) that

\[
\left| 1 + p_{n,k} \left( e^{i\lambda} - 1 \right) \right| \leq 1.
\]

Hence we may apply Lemma 26.14 to find

\[
\prod_{k=1}^{n} \exp \left( p_{n,k} \left( e^{i\lambda} - 1 \right) \right) \to 0 \text{ as } n \to \infty.
\]

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

\[
\lim_{n \to \infty} \mathbb{E}[e^{i\lambda S_n}] = \lim_{n \to \infty} \prod_{k=1}^{n} \left[ 1 + p_{n,k} \left( e^{i\lambda} - 1 \right) \right]
\]

\[
= \lim_{n \to \infty} \prod_{k=1}^{n} \exp \left( p_{n,k} \left( e^{i\lambda} - 1 \right) \right) = \exp \left( a \left( e^{i\lambda} - 1 \right) \right).
\]

The result now follows by an application of the continuity Theorem 25.20.
Remark 26.24. Keeping the notation in Theorem 26.23, we have
\[ \mathbb{E}[Y_{n,k}] = p_{n,k} \text{ and } \text{Var}(Y_{n,k}) = p_{n,k}(1 - p_{n,k}) \]
and
\[ s_n^2 := \sum_{k=1}^{n} \text{Var}(Y_{n,k}) = \sum_{k=1}^{n} p_{n,k}(1 - p_{n,k}). \]
Under the assumptions of Theorem 26.23, 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
\[ \sigma_{n,k}^2 := \text{Var}(X_{n,k}) = \frac{1}{s_n^2} \text{Var}(Y_{n,k}) = \frac{1}{s_n^2} p_{n,k}(1 - p_{n,k}), \]
\[ \mathbb{E}[X_{n,k}] = 0, \text{Var}(\sum_{k=1}^{n} X_{n,k}) = 1, \text{and the } \{X_{n,k}\} \text{ satisfy condition } (M). \]
On the other hand for small \( t \) and large \( n \) we have
\[ \mathbb{E}[X_{n,k}^2 : |X_{n,k}| > t] = \mathbb{E}\left[X_{n,k}^2 : \frac{Y_{n,k} - p_{n,k}}{s_n} > \frac{t}{s_n}\right] \]
\[ = \mathbb{E}\left[X_{n,k}^2 : |Y_{n,k} - p_{n,k}| > ts_n\right] \]
\[ \geq \mathbb{E}\left[X_{n,k}^2 : |Y_{n,k} - p_{n,k}| > 2at\right] \]
\[ = \mathbb{E}\left[X_{n,k}^2 : Y_{n,k} = 1\right] = p_{n,k}\left(1 - \frac{p_{n,k}}{s_n}\right)^2 \]
from which it follows that
\[ \lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E}[X_{n,k}^2 : |X_{n,k}| > t] = \lim_{n \to \infty} \sum_{k=1}^{n} p_{n,k}\left(1 - \frac{p_{n,k}}{s_n}\right)^2 = a. \]
Therefore \( \{X_{n,k}\} \) do not satisfy (LC). Nevertheless we have by Theorem 26.23 along with Slutsky’s Theorem 24.43 that
\[ \sum_{k=1}^{n} X_{n,k} = \frac{\sum_{k=1}^{n} Y_{n,k} - \sum_{k=1}^{n} p_{n,k}}{s_n} \implies Z - a \sim a \]
where \( Z \) is a Poisson random variable with mean \( a \). Notice that the limit is not a normal random variable in agreement with Theorem 26.20.

We are now going to 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 26.25 (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
\[ f_L(\lambda) := \mathbb{E}[e^{i\lambda L}] = \exp\left(\int_{\mathbb{R}}\frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right) \]
for some finite positive measure, \( \nu \), on \( (\mathbb{R}, B_{\mathbb{R}}) \) with \( \nu(\mathbb{R}) \leq 1 \).

**Proof.** As before, let \( f_{n,k}(\lambda) = \mathbb{E}[e^{i\lambda X_{n,k}}] \) and \( \varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1 \). By the continuity theorem we are assuming
\[ \lim_{n \to \infty} f_{n,k}(\lambda) = \lim_{n \to \infty} \prod_{k=1}^{n} f_{n,k}(\lambda) = f(\lambda) \]
where \( f(\lambda) \) is continuous at \( \lambda = 0 \). We are also assuming property \( (M) \), i.e.
\[ \lim_{n \to \infty} \max_{1 \leq n \leq k} \sigma_{n,k}^2 = 0. \]
Under condition \( (M) \), we expect \( f_{n,k}(\lambda) \equiv 1 \) for \( n \) large. Therefore we expect
\[ f_{n,k}(\lambda) = e^{\ln f_{n,k}(\lambda)} = e^{\ln[1 + (f_{n,k}(\lambda) - 1)]} \approx e(f_{n,k}(\lambda) - 1) \]
and hence that
\[ \mathbb{E}[e^{i\lambda S_n}] = \prod_{k=1}^{n} f_{n,k}(\lambda) \equiv \prod_{k=1}^{n} e(f_{n,k}(\lambda) - 1) = \exp\left(\sum_{n=1}^{n} (f_{n,k}(\lambda) - 1)\right). \]
This is in fact correct, since Lemma 26.18 indeed implies
\[ \lim_{n \to \infty} \left[ \mathbb{E}[e^{i\lambda S_n}] - \exp\left(\sum_{k=1}^{n} (f_{n,k}(\lambda) - 1)\right) \right] = 0. \]
Since \( \mathbb{E}[X_{n,k}] = 0 \),
\[ f_{n,k}(\lambda) - 1 = \mathbb{E}[e^{i\lambda X_{n,k}} - 1] = \mathbb{E}[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}] \]
\[ = \int_{\mathbb{R}} (e^{i\lambda x} - 1 - i\lambda x) d\mu_{n,k}(x) \]
where \( \mu_{n,k} := P \circ X_{n,k}^{-1} \) is the law of \( X_{n,k} \). Therefore we have
\[ \exp\left(\sum_{k=1}^{n} (f_{n,k}(\lambda) - 1)\right) = \exp\left(\sum_{k=1}^{n} \int_{\mathbb{R}} (e^{i\lambda x} - 1 - i\lambda x) d\mu_{n,k}(x)\right) \]
\[ = \exp\left(\int_{\mathbb{R}} (e^{i\lambda x} - 1 - i\lambda x) \sum_{k=1}^{n} d\mu_{n,k}(x)\right) \]
\[ = \exp\left(\int_{\mathbb{R}} (e^{i\lambda x} - 1 - i\lambda x) d\nu_n(x)\right) \]
(26.23)
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. (26.22) and Eq. (26.23) that
\[
\left| f_{S_n} (\lambda) - \exp \left( \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_n (x) \right) \right| \to 0. \tag{26.24}
\]

Let
\[
\varphi (\lambda, x) := \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} = -\frac{\lambda^2}{2} \int_0^1 e^{it\lambda x} 2(1-t) \, dt \tag{26.25}
\]
(the second equality is from Taylor’s theorem) and extend \( \varphi (\lambda, \cdot) \) to \( \mathbb{R} \) by setting \( \varphi (\lambda, \pm \infty) = 0 \). Then \( \{ \varphi (\lambda, \cdot) \}_{\lambda \in \mathbb{R}} \subset C (\mathbb{R}) \) and therefore by Helly’s selection Theorem [24.61] there is a probability measure \( \bar{\nu} \) on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) 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 \( h \in C (\mathbb{R}) \)). Combining this with Eq. (26.24) allows us to conclude,
\[
\frac{n}{L/s} \left( \int \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_n (x) \right) \to 0.
\]
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, \( \nu (\mathbb{R}) = \bar{\nu} (\mathbb{R}) \leq \bar{\nu} (\mathbb{R}) = 1 \).

We are now going to drop the assumption that \( \text{Var} (S_n) = 1 \) for all \( n \) and replace it with the following property.

**Definition 26.26.** We say that \( \{ X_{n,k} \}_{k=1}^n \) has bounded variation (BV) iff
\[
\sup_n \text{Var} (S_n) = \sup_n \sum_{k=1}^n \sigma_{n,k}^2 < \infty. \tag{26.26}
\]

**Corollary 26.27 (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 \( S_n := \sum_{k=1}^n X_{n,k} \Rightarrow L \) for some random variable \( L \), then
\[
f_L (\lambda) = \exp \left( \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu (x) \right) \tag{26.27}
\]
where \( \nu \) is a finite positive measure on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \).

**Proof.** Let \( s_n^2 := \text{Var} (S_n) \). If \( \lim_{n \to \infty} s_n = 0 \), then \( S_n \to 0 \) in \( L^2 \) and hence weakly, therefore Eq. (26.27) 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 [24.43] that \( S_n/s_n \Rightarrow L/s \). Hence by an application of Theorem 26.25 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 \( \nu (\mathbb{R}) \leq 1 \). Letting \( \lambda \to s\lambda \) in this expression then implies
\[
f_L (\lambda) = \exp \left( \int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{x^2} d\nu (x) \right)
\]
\[
= \exp \left( \int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{sx^2} s^2 d\nu (x) \right)
\]
\[
= \exp \left( \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_s (x) \right)
\]
where \( \nu_s \) is the finite measure on \( (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) defined by
\[
\nu_s (A) := s^2 \nu (s^{-1} A) \text{ for all } A \in \mathcal{B}_\mathbb{R}.
\]

From Eq. (26.25) 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,
\[
\frac{d}{d\lambda} \varphi (\lambda, x) = \frac{ixe^{i\lambda x} - ix}{x^2} = i e^{i\lambda x} - \frac{1}{x}
\]
and
\[
\frac{d^2}{d\lambda^2} \varphi (\lambda, x) = \frac{ixe^{i\lambda x}}{x} = -e^{i\lambda x}.
\]
Using these remarks and the fact that \( \nu (\mathbb{R}) < \infty \), it is easy to see that
\[
f_L' (\lambda) = \left( \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1}{x} d\nu_s (x) \right) f_L (\lambda)
\]
and
\[
f_L'' (\lambda) = \left( \int_{\mathbb{R}} -e^{i\lambda x} d\nu_s (x) + \left( \int_{\mathbb{R}} \frac{e^{i\lambda x} - 1}{x} d\nu_s (x) \right)^2 \right) f_L (\lambda).
\]
and in particular, \( f_L'' (0) = 0 \) and \( f_L'' (0) = -\nu_s (R) \). Therefore by Theorem 26.28 the probability measure \( \mu \) on \((R, B_R)\) such that \( \hat{\mu} (\lambda) = f_L (\lambda) \) has mean zero and variance, \( \nu_s (R) < \infty \). This later condition reflects the \((BV)\) assumption that we made.

**Theorem 26.28.** The following class of **symmetric** distributions on \((R, B_R)\) 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 26.27.
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 26.27. For \( C_2 \subset C_3 \), observe that

\[
\hat{\mu} (\lambda) = \exp \left( \int_R \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} \, d\nu (x) \right)
\]

then \( \hat{\mu} (\lambda) = [\hat{\mu}_n (\lambda)]^n \) where \( \mu_n \) is the unique probability measure on \((R, B_R)\) such that

\[
\hat{\mu}_n (\lambda) = \exp \left( \int_R \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} \, d\nu (x) \right).
\]

For \( C_3 \subset C_1 \), simply define \( \{X_{n,k}\}_{k=1}^n \) to be i.i.d with \( \mathbb{E} [e^{i\lambda X_{n,k}}] = \hat{\mu}_n (\lambda) \). In this case \( S_n = \sum_{k=1}^n X_{n,k} \overset{d}{=} \mu \).

### 26.3 Stable Distributions

**Definition 26.29.** A non-degenerate distribution \( \mu = \text{Law} (X) \) on \( R \) is **stable** if whenever \( X_1 \) and \( X_2 \) are independent copies of \( X \), then for all \( a, b \in R \) there exists \( c, d \in R \) such that \( aX_1 + bX_2 \overset{d}{=} cX + d \) with some constants \( c \) and \( d \).

**Example 26.30.** Any Gaussian random variable is stable. Indeed if \( X \overset{d}{=} \sigma N + \mu \) where \( \sigma > 0 \) and \( \mu \in R \) and \( N = N(0,1) \), then \( X_1 = \sigma N_1 + \mu \) where \( N_1 \) and \( N_2 \) are independent with \( N_1 \overset{d}{=} N \) we will have \( aX_1 + bX_2 \) is Gaussian mean \( (a+b)\mu \) and variance \((a^2 + b^2)\sigma^2 \) so that

\[
aX_1 + bX_2 \overset{d}{=} \sqrt{(a^2 + b^2)\sigma^2} + (a+b)\mu
\]

\[
\overset{d}{=} \sqrt{(a^2 + b^2)} (X - \mu) + (a+b)\mu
\]

\[
= \sqrt{(a^2 + b^2)} X + \left( a + b - \sqrt{(a^2 + b^2)} \right) \mu.
\]

**Example 26.31.** Poisson random variables are not stable. For suppose that \( Z = \text{Pois} (\rho) \) and \( Z_1 \overset{d}{=} Z_2 \overset{d}{=} Z \), then \( Z_1 + Z_2 \overset{d}{=} \text{Pois} (2\rho) \). If we could find \( a, b \) such that

\[
\text{Pois} (2\rho) \overset{d}{=} Z_1 + Z_2 \overset{d}{=} aZ + b
\]

we would have

\[
e^{-2\rho (2\rho)^n \over n!} = P(aZ + b = n) = P \left( Z = \frac{n - b}{a} \right)
\]

In particular this implies that \( {2\rho \over n} = k_n \in N_0 \) for all \( n \in N_0 \) 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 \( Z \overset{d}{=} \text{Pois} (2\rho) \) which is absurd.

**Lemma 26.32.** 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 26.33.** If \( \mu \) is a stable distribution then it is infinitely divisible.

**Proof.** Let \( \{X_n\}_{i=1}^N \) be i.i.d random variables with \( \text{Law} (X_n) = \mu = \text{Law} (X) \). As \( \mu \) is stable we know that

\[
X_1 + \cdots + X_N \overset{d}{=} aX + b. \tag{26.28}
\]

As \( \mu \) is non-degenerate, it follows from Lemma 26.32 that \( a \neq 0 \). Therefore form Eq. (26.28) 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 26.31. 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 26.34. 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 \( aX_1 + bX_2 = cX + d \), we must have that \( d = 0 \) and may take \( c > 0 \) by the symmetry assumption. Letting \( f(\lambda) = E[e^{i\lambda X}] \) we may now conclude that
\[
f(a\lambda) \cdot f(b\lambda) = E[e^{i(\alpha a + b)X}] = E[e^{i\lambda X}] = f(c\lambda).
\]
It turns out that the solution to these functional equation are of the form \( f(\lambda) = e^{-k|\lambda|^{\alpha}} \). If \( f(\lambda) \) is of this form then
\[
f(a\lambda) \cdot f(b\lambda) = \exp(-k(|a|^\alpha + |b|^\alpha)|\lambda|^\alpha) = f(c\lambda)
\]
where \( c = (|a|^\alpha + |b|^\alpha)^{1/\alpha} \). 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
\[
f'(\lambda) = -k|\lambda|^\alpha \lambda f(\lambda) \leq 0 \quad \text{and}
\]
\[
f''(\lambda) = \left(k^2|\lambda|^{2\alpha - 2} - k(\alpha - 1)|\lambda|^{\alpha - 2}\right) f(\lambda) \geq 0
\]
so that \( f \) is a decreasing convex symmetric function for \( \alpha \geq 0 \). Therefore by Polya’s criteria of Exercise [25.7] 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.

26.4 *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 26.35. 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 \( (\mathbb{R}, B_\mathbb{R}) \). The convention here is that
\[
e^{i\lambda x} \frac{1 - i\lambda x}{x^2} \bigg|_{x=0} := \lim_{x \to 0} e^{i\lambda x} \frac{1 - i\lambda x}{x^2} = -\frac{1}{2} \lambda^2.
\]

Proof. This is the content of Exercise 25.5.

1. If \( \{X_t\}_{t \geq 0} \) is a right continuous process with stationary and independent increments, then let \( f_t(\lambda) := E[e^{i\lambda(X_t-X_0)}] \) for any \( \sigma \geq 0 \). It then follows that
\[
\begin{align*}
\hat{f}_{t+s}(\lambda) &= E[e^{i\lambda(X_{t+s}-X_0)}] = E[e^{i\lambda(X_{t+s}-X_t+X_t-X_0)}] \\
&= E[e^{i\lambda(X_{t+s}-X_t)}] \cdot E[e^{i\lambda(X_t-X_0)}] \\
&= f_s(\lambda) \cdot f_t(\lambda).
\end{align*}
\]
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
\[
e^{t \text{Re } \psi(\lambda)} = |f_t(\lambda)| \leq 1
\]
it follows that \( \text{Re } \psi(\lambda) \leq 0 \). 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 \( \lambda \) 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,
\[
g(t+s) = g(t) + g(s) \text{ for all } 0 \leq s, t \leq \varepsilon.
\]
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 \( g_\varepsilon(1) = g_\varepsilon(n/1/n) = n \cdot g_\varepsilon(1/n) \) 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 \( g_\varepsilon(t) = tg_\varepsilon(1) \) for all \( 0 \leq t < 1 \). Thus we have shown \( g_\varepsilon(\varepsilon t) = tg_\varepsilon(t) \) for \( 0 \leq t < \varepsilon \) 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
Thus we have shown that \( f_t(\lambda) = e^{\psi(\lambda)} \) for some function \( \psi(\lambda) \). Let us further observe that
\[
\psi(\lambda) = \lim_{t \to 0} \frac{f_t(\lambda) - 1}{t}
\]
from which it follows that \( \psi \) must be measurable. Furthermore,
\[
\psi(-\lambda) = \lim_{t \to 0} \frac{f_t(-\lambda) - 1}{t} = \lim_{t \to 0} \frac{f_t(\lambda) - 1}{t} = \psi(\lambda).
\]

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
\[
\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j = \lim_{t \to 0} \sum_{i,j=1}^n \frac{f_t(\lambda_i - \lambda_j) - 1}{t} z_i \bar{z}_j
\]
while for any \( \{z_i\}_{i=1}^n \subset \mathbb{C} \) we have
\[
\sum_{i,j=1}^n f_t(\lambda_i - \lambda_j) z_i \bar{z}_j = \sum_{i,j=1}^n \mathbb{E} \left[ e^{i(\lambda_i - \lambda_j) X_i} \right] z_i \bar{z}_j
\]
\[
= \sum_{i,j=1}^n \mathbb{E} \left[ e^{i\lambda_i X_i z_i} \cdot e^{-i\lambda_j X_i \bar{z}_j} \right]
\]
\[
= \mathbb{E} \left[ \sum_{i=1}^n e^{i\lambda_i X_i z_i} \cdot \sum_{j=1}^n e^{-i\lambda_j X_i \bar{z}_j} \right]
\]
\[
= \mathbb{E} \left[ \sum_{i=1}^n e^{i\lambda_i X_i z_i} \right]^2 \geq 0.
\]

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, \( \psi(-\lambda) = \psi(\lambda) \) and \( \psi \) is conditionally positive definite, then \( e^{i\psi(\lambda)} \) is a characteristic function. We will prove this below using Bochner’s Theorem [25, 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 \( \psi(-\lambda) = -i\lambda a - b\lambda^2 = \psi(\lambda) \) 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 \( \sum_{i,j=1}^n [\lambda_j^2] z_i \bar{z}_j = 0 \), it follows that
\[
\sum_{i,j=1}^n \psi(\lambda_i - \lambda_j) z_i \bar{z}_j = \sum_{i,j=1}^n \left[ -b(2\lambda_i \lambda_j) \right] z_i \bar{z}_j
\]
\[
= 2b \left| \sum_{i,j=1}^n \lambda_i z_i \right|^2 \geq 0.
\]

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
\[
f_X(\lambda) = \mathbb{E} \left[ e^{i\lambda X} \right] = \sum_{n=0}^\infty \mathbb{E} \left[ e^{i\lambda X} : N = n \right]
\]
\[
= \sum_{n=0}^\infty \mathbb{E} \left[ e^{i\lambda(Z_1 + \cdots + Z_n)} : N = n \right]
\]
\[
= e^{-\lambda} \sum_{n=0}^\infty \frac{\lambda^n}{n!} [f_{Z_1}(\lambda)]^n = \exp(\lambda(f_{Z_1}(\lambda) - 1)).
\]

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 26.36.** 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 \( (A_{ij})_{i,j=1}^n \) 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
\[
A_{ij}^n = (A^@e_j, e_i^@n)
\]
and therefore,
\[\sum_{i,j=1}^{d} A_{ij}^n z_i z_j = \sum_{i,j=1}^{d} (A^\otimes_n e_j^\otimes_n, e_i^\otimes_n) z_i z_j = (A^\otimes_n \psi, \psi)\]
where \(\psi := \sum_{i=1}^{d} z_i e_j^\otimes_n \in (C^d)^\otimes_n\). So it suffices to show \(A^\otimes_n \geq 0\). To do this let \(u_i\) be an O.N. basis for \(C^d\) such that \(A u_i = \lambda_i u_i\) for all \(i\). Since \(A \geq 0\) we know that \(\lambda_i \geq 0\) and therefore
\[A^\otimes_n (u_i \otimes \cdots \otimes u_{ia}) = (\lambda_i \cdots \lambda_{ia}) (u_i \otimes \cdots \otimes u_{ia})\]
where \((\lambda_1 \cdots \lambda_{ia}) \geq 0\). This shows that \(A^\otimes_n\) is unitarily equivalent to a diagonal matrix with non-negative entries and hence is positive semi-definite.

**Proposition 26.37.** Suppose that \(\{A_{ij}\}_{i,j=1}^{d} \subset C\) is a matrix such that \(A^* = A\) and \(A\) is conditionally positive definite, for example \(A_{ij} := \psi (\lambda_i - \lambda_j)\) as above. Then the matrix with entries, \((e^{A_{ij}})_{i,j=1}^{d}\) is positive definite.

**Proof.** Let \(u := (1, \ldots, 1)^{tr} \in C^d\). Let \(\xi \in C^d\) and write \(\xi = z + \alpha u\) where \((z, u) = 0\) and \(\alpha := (\xi, u)/d\). Letting \(B := \sqrt{A}\) on \(u^+\) and on \(C \cdot u\), we have
\[(A \xi, \xi) = (A (z + \alpha u), z + \alpha u) = (A z, z) + 2 \Re [\alpha (A z, u)] + |\alpha|^2 (A u, u) = (A z, z) + 2 \Re [\alpha (B z, u)] + |\alpha|^2 (A u, u) \geq (A z, z) - 2 \|Bz\| \cdot |\alpha| \|B^* u\| + |\alpha|^2 (A u, u) \geq (A z, z) - (\|Bz\|^2 + |\alpha|^2 \|B^* u\|^2) + |\alpha|^2 (A u, u) = |\alpha|^2 \left[(A u, u) - \|B^* u\|^2\right].\]
Since
\[(u \ u^{tr} \xi, \xi) = |\alpha|^2 (u \ u^{tr} u, u) = |\alpha|^2 d^2,\]
it follows that
\[\left((A + \lambda u \ u^{tr}) \xi, \xi\right) \geq |\alpha|^2 \left[(A u, u) - \|B^* u\|^2 + \lambda d^2\right] \geq 0\]
provided \(\lambda d^2 \geq \|B^* u\|^2 - (A u, u)\).

We now fix such a \(\lambda \in R\) so that \((A + \lambda u \ u^{tr}) \geq 0\). It then follows from Lemma 26.36 that

\[e^{\lambda t} A_{ij} = e^{\lambda t} + \lambda = e^{(A + \lambda u \ u^{tr}) t} = \sum_{n=0}^{\infty} \frac{(A + \lambda u \ u^{tr}) t}{n!} i_{ij} \]
are the matrix entries of a positive definite matrix. Scaling this matrix by \(e^{-\lambda} > 0\) then gives the result that \((e^{A_{ij}})_{i,j=1}^{d} \geq 0\).

As a consequence it follows that \(e^{\psi(t)}\) is a positive definite function whenever \(\psi\) is conditionally positive definite.

**Proposition 26.38.** Suppose that \(\{Z_t\}_{i=1}^{\infty}\) are i.i.d. random vectors in \(R^d\) with Law \((Z_t) = \mu\) 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 \([E [e^{ik \cdot X_t}] = e^{\psi(k)}\]
where
\[\psi (k) = \lambda \int_{R^n} e^{ik \cdot x} d\mu (x) = E [e^{ik \cdot Z_t}].\]

**Proof.** It has already been shown in Theorem 20.30 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,
\[E [e^{ik \cdot X_t}] = \left[Q_t (x \rightarrow e^{ik \cdot x})\right] (0) \]
\[= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} E [e^{ik \cdot (Z_1 + \cdots + Z_n)}] \]
\[= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \hat{\mu}(k)^n \]
\[= e^{\lambda (\hat{\mu} (k) - 1)} \exp \left(i \lambda \int_{R^n} e^{ik \cdot x} d\mu (x)\right).\]

More generally, if we let \(B_t\) be Brownian motion in \(R^n\) with Cov \([B_i, B_j]\) = \(A_{ij} t\) 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 Lévy process whose Fourier transform is given by,
\[E [e^{ik \cdot X_t}] = \exp \left(i bt + Ak \cdot k + \lambda \int_{R^n} e^{ik \cdot x} d\mu (x)\right).\]

Thus
\[\psi(t) = ib + Ak \cdot k + \lambda \int_{R^n} (e^{ik \cdot x} - 1) d\mu (x)\]
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 \(\psi (k)\) will be a Lévy exponent iff \(\psi\) has the form given in Eq. 26.29 below.

**Theorem 26.39 (Lévy Kintchine formula).** If \(\psi\) is continuous at zero and conditionally positive definite, then
\[ \psi(\lambda) = i\lambda b - \frac{1}{2}a\lambda^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{i\lambda x} - 1 - i\lambda x \cdot 1_{|x|\leq 1}) \, d\nu(x) \]  

(26.29)

for some \( b \in \mathbb{R}, a \geq 0 \), and some measure \( \nu \) such that

\[ \int_{\mathbb{R}\setminus\{0\}} (x^2 \wedge 1) \, d\nu(x) < \infty. \]
Stochastic Processes II
We are now going to discuss continuous time stochastic processes in more detail. We will be using Poisson processes (Definition 14.8) and Brownian motion (Definition 20.24) 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 20.24 except for the continuity of the sample paths. We will then use Kolmogorov’s continuity criteria (Theorem 28.8) 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 – mention the free Euclidean field and its connections to SLE.)
Gaussian Random Fields

Recall from Section 12.9 (which the reader should review if necessary) that a random variable, \( Y : \Omega \to \mathbb{R} \) is said to be **Gaussian** if
\[
\mathbb{E}[e^{i\lambda Y}] = \exp \left( -\frac{1}{2} \text{Var}(Y) + i\lambda Y \right) \quad \forall \lambda \in \mathbb{R}.
\]

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
\[
\mathbb{E}[e^{i\lambda X}] = \exp \left( -\frac{1}{2} \text{Var}(\lambda \cdot X) + i\lambda \cdot \mathbb{E}[X] \right) \quad \forall \lambda \in \mathbb{R}^n. \tag{27.1}
\]

**Remark 27.1.** 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, \((X,X \cdot Y)^t\) has Gaussian components but is not Gaussian.

**Exercise 27.1 (Same as Exercise 12.10).** Prove the assertion made in Remark 27.1 by computing \( \mathbb{E}[e^{i(\lambda_1 X + \lambda_2 Y)}] \). (Another proof that \((X,X \cdot Y)^t\) is not Gaussian follows from the fact that \(X\) and \(XY\) are uncorrelated but not independent\(^1\), which would then contradict Lemma 13.28)

### 27.1 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 |\(X| \leq \frac{1}{2}\) iff |\(XY| \leq \frac{1}{2}\) and therefore,
\[
P\left(|X| \leq \frac{1}{2} \text{ and } |XY| \leq \frac{1}{2}\right) = P\left(|X| \leq \frac{1}{2}\right) =: \alpha
\]

while
\[
P\left(|X| \leq \frac{1}{2}\right) P\left(|XY| \leq \frac{1}{2}\right) = \alpha^2 \neq \alpha.
\]

**Theorem 27.2.** 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, \( q : \mathbb{R}^n \to \mathbb{R} \),
\[
\mathbb{E}[q(X)] = \left( e^{\frac{1}{2} L q} \right)(0) := \sum_{n=0}^{\infty} \frac{1}{n!} \left( \left( \frac{L}{2} \right)^n q \right)(0) \quad \text{(a finite sum).} \tag{27.2}
\]

**Proof. First Proof.** The first proof is conceptually clear but technically a bit more difficult. In this proof we will begin by proving Eq. (27.2) when \( q(x) = e^{i\lambda x} \) where \( \lambda \in \mathbb{R}^n \). The function \( q \) is not a polynomial, but never mind. In this case,
\[
\mathbb{E}[q(X)] = \mathbb{E}[e^{i\lambda X}] = e^{-\frac{1}{2} Q \lambda \lambda}.
\]

On the other hand,
\[
\left( e^{\frac{1}{2} L q} \right)(x) = \frac{1}{2} Q_{ij} \partial_i \partial_j e^{i\lambda x} = \frac{1}{2} (Q \lambda \cdot \lambda) e^{i\lambda x} = -\frac{1}{2} (Q \lambda \cdot \lambda) q(x).
\]

Therefore,
\[
e^{\frac{1}{2} L q} = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{1}{2} Q \lambda \lambda \right)^n q = e^{-\frac{1}{2} Q \lambda \lambda} q
\]

and hence
\[
\left( e^{\frac{1}{2} L q} \right)(0) = e^{-\frac{1}{2} Q \lambda \lambda}.
\]

Thus we have shown
\[
\mathbb{E}[e^{i\lambda X}] = e^{\frac{1}{2} L e^{i\lambda x}}|_{x=0}.
\]

The result now formally follows by differentiating this equation in \(\lambda\) and then setting \(\lambda = 0\). Indeed observe that
\[
\mathbb{E}[(iX)^n] = \partial^n \mathbb{E}[e^{i\lambda X}] |_{\lambda = 0} = \partial^n e^{\frac{1}{2} L e^{i\lambda x}} |_{x=0, \lambda = 0} = e^{\frac{1}{2} L} \partial^n e^{i\lambda x} \big|_{x=0}.
\]

To justify this last equation we must show,
\[
\partial^n e^{\frac{1}{2} L} e^{i\lambda x} = e^{\frac{1}{2} L} \partial^n e^{i\lambda x}
\]
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

\[ \partial_q^\infty L e^{i\lambda x} = \partial_q^\infty \sum_{n=0}^{\infty} \frac{1}{n!} \left( -\frac{1}{2} Q \lambda \cdot \lambda \right)^n e^{i\lambda x} \]

while on the other,

\[ e^{\frac{1}{2} L} \partial_q^\infty e^{i\lambda x} = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \left( \frac{L}{2} \right)^n e^{i\lambda x} \right) = \sum_{n=0}^{\infty} \frac{1}{n!} \partial_q^\infty \left( \left( \frac{L}{2} \right)^n e^{i\lambda x} \right) \]

\[ = \sum_{n=0}^{\infty} \partial_q^\infty \left( \frac{1}{n!} \left( -\frac{1}{2} Q \lambda \cdot \lambda \right)^n e^{i\lambda x} \right) . \]

Thus to complete the proof we must show,

\[ \partial_q^\infty \sum_{n=0}^{\infty} \frac{1}{n!} \left( -\frac{1}{2} Q \lambda \cdot \lambda \right)^n e^{i\lambda x} = \sum_{n=0}^{\infty} \partial_q^\infty \left( \frac{1}{n!} \left( -\frac{1}{2} Q \lambda \cdot \lambda \right)^n e^{i\lambda x} \right) . \]

Perhaps the easiest way to do this would be to use the Cauchy estimates,\(^2\) which allow one to show that if \( \left\{ f_n (\lambda) \right\}_{n=0}^{\infty} \) is a sequence of analytic functions such that \( \sum_{n=0}^{\infty} f_n (\lambda) \) is uniformly convergent on compact subsets, then \( \sum_{n=0}^{\infty} \partial_q^\infty f_n (\lambda) \) is also uniformly convergent on compact subsets and therefore,

\[ \partial_q^\infty \sum_{n=0}^{\infty} f_n (\lambda) = \sum_{n=0}^{\infty} \partial_q^\infty f_n (\lambda) . \]

Now apply this result with \( f_n (\lambda) := \frac{1}{n!} \left( -\frac{1}{2} Q \lambda \cdot \lambda \right)^n e^{i\lambda x} \) 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. (27.2) holds for \( q (x) = p (x) e^{i\lambda x} \) where \( \lambda \in \mathbb{R}^N \) and \( p \) is a polynomial.

**Second Proof.** Let

\[ u (t, y) := E \left[ q \left( y + \sqrt{t} X \right) \right] = Z^{-1} \int_{\mathbb{R}^N} q \left( y + \sqrt{t} x \right) e^{-Q^{-1} x \cdot x} dx \tag{27.3} \]

\[ = Z^{-1} \int_{\mathbb{R}^N} q (y + x) e^{-Q^{-1} x \cdot x} \frac{dx}{t^{n/2}} . \tag{27.4} \]

One now verifies that

\[ \partial_t e^{-Q^{-1} x \cdot x / 2t} = t^{-n/2} \frac{1}{2} e^{-Q^{-1} x \cdot x / 2t} . \]

Using this result and differentiating under the integral in Eq. (27.4) then shows,

\[ \partial_t u (t, y) = \frac{1}{2} L_y u (t, y) \quad \text{with} \quad u (0, y) = q (y) . \]

Moreover, from Eq. (27.3), 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,

\[ v (t, y) := \sum_{n=0}^{\infty} \frac{1}{n!} \left( \left( \frac{L}{2} \right)^n q \right) (y) = e^{L/2 q} (y) \]

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

\[ u (1, 0) = E \left[ q (0 + \sqrt{1} X) \right] = E \left[ q (X) \right] \quad \text{and} \]

\[ v (1, 0) = \left( e^{L/2 q} \right) (0) . \]

**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

\[ E \left[ (Q^{-1} X \cdot u) p (X) \right] = -\frac{1}{Z} \int_{\mathbb{R}^N} p (x) \partial_u \exp \left( -\frac{1}{2} Q^{-1} x \cdot x \right) dx \]

\[ = \frac{1}{Z} \int_{\mathbb{R}^N} \partial_u p (x) \exp \left( -\frac{1}{2} Q^{-1} x \cdot x \right) dx \]

\[ = E \left[ (\partial_u p) (X) \right] . \]

Replacing \( u \) by \( Qu \) in this equation leads to important identity,

\[ E \left[ (Q \cdot u) p (X) \right] = E \left[ (\partial Q u) p (X) \right] . \tag{27.5} \]

It is clear that using this identity and induction it would be possible to compute \( E \left[ p (X) \right] \) for any polynomial \( p \). So to finish the proof it suffices to show

\[ e^{L/2 ((x \cdot u) p (x))} \big|_{x=0} = e^{L/2 ((\partial Q u) p (x))} \big|_{x=0} = 0. \]
This is correct that
\[
e^{L/2} ((x \cdot u) p(x)) = e^{L/2} \left( (x \cdot u) e^{-L/2} e^{L/2} p(x) \right) = e^{L/2} M_{(x \cdot u)} e^{-L/2} e^{L/2} p(x)
\]
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
\[
e^{L/2} ((x \cdot u) p(x)) |_{x=0} = \left( (M_{(x \cdot u)} + \partial_{Qu}) e^{L/2} p(x) \right) |_{x=0} = \left[ \partial_{Qu} e^{L/2} p(x) \right] |_{x=0} = e^{L/2} \partial_{Qu} p(x) |_{x=0}
\]
which is the same identity as in Eq. \((27.5)\).

**Example 27.3.** Suppose \( X \overset{d}{=} N(1, 0) \in \mathbb{R} \), then
\[
\mathbb{E} \left[ X^{2n} \right] = \left[ e^{L/2} x^{2n} \right] |_{x=0} = \frac{1}{n!} \cdot 2^n \Delta^n \|x\|^{2n} = \frac{(2n)!}{2^n \cdot n!}.
\]

## 27.2 Existence of Gaussian Fields

**Definition 27.4.** 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 \( Q(s,t) := \text{Cov}(X_s, X_t) \). 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 27.5.** Suppose \( T \) is a set and \( \{X_t : t \in T\} \) is a random field. For any \( \Lambda \subset T \), let \( B_{\Lambda} := \sigma(X_t : t \in \Lambda) \).

**Proposition 27.6.** 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
\[
\mathbb{E} [X_s] = c(s) \text{ and } \text{Cov}(X_s, X_t) = Q(s,t) \quad (27.7)
\]
for all \( s, t \in T \).

**Proof.** Since we will construct \((\Omega, \mathcal{B}, P)\) by Kolmogorov’s extension Theorem \([20.68]\), 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 \((\mathbb{R}^\Lambda, B_\Lambda := \mathcal{B}_{\mathbb{R}^\Lambda})\) such that
\[
\int_{\mathbb{R}^\Lambda} e^{i \sum_{t \in \Lambda} \lambda(t) x(t)} d\mu_\Lambda(x) = \exp \left( -\frac{1}{2} \sum_{s,t \in \Lambda} Q(s,t) \lambda(s) \lambda(t) + i \sum_{s \in \Lambda} c(s) \lambda(s) \right).
\]
The main point now is to show \( \{ (\mathbb{R}^\Lambda, B_\Lambda, \mu_\Lambda) \}_{\Lambda \subset_f T} \) 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 \( \bar{\lambda} \in \mathbb{R}^\Gamma \) be defined so that \( \bar{\lambda} = \lambda \) on \( \Lambda \) and \( \bar{\lambda} = 0 \) on \( \Gamma \setminus \Lambda \). We then have,
\[
\int_{\mathbb{R}^\Lambda} e^{i \sum_{t \in \Lambda} \lambda(t) x(t)} d\mu_\Lambda (x)
\]
\[
= \int_{\mathbb{R}^\Gamma} e^{i \sum_{t \in \Gamma} \lambda(t) x(t)} d\mu_\Gamma (x)
\]
\[
= \int_{\mathbb{R}^\Gamma} e^{i \sum_{t \in \Gamma} \bar{\lambda}(t) x(t)} d\mu_\Gamma (x)
\]
\[
= \exp \left( -\frac{1}{2} \sum_{s,t \in \Gamma} Q(s,t) \bar{\lambda}(s) \bar{\lambda}(t) + i \sum_{s \in \Gamma} c(s) \bar{\lambda}(s) \right)
\]
\[
= \exp \left( -\frac{1}{2} \sum_{s,t \in \Lambda} Q(s,t) \lambda(s) \lambda(t) + i \sum_{s \in \Lambda} c(s) \lambda(s) \right)
\]
\[
= \int_{\mathbb{R}^\Lambda} e^{i \sum_{t \in \Lambda} \lambda(t) x(t)} d\mu_\Lambda (x)
\]
27 Gaussian Random Fields

Since this is valid for all \( \lambda \in \mathbb{R}^4 \), it follows that \( \mu_\Gamma \circ \pi^{-1} = \mu_A \) as desired. Hence by Kolmogorov’s theorem, there exists a unique probability measure, \( P \) on \((\Omega, B)\) such that

\[
\int_\Omega f(\omega) dP(\omega) = \int_{\mathbb{R}^4} f(x) d\mu_A(x)
\]

for all \( A \subset_f T \) and all bounded measurable functions, \( f: \mathbb{R}^4 \to \mathbb{R} \). In particular, it follows that

\[
E \left[ e^{i\sum_{t\in A} \lambda(t)X_t} \right] = \int_\Omega e^{i\sum_{t\in A} \lambda(t)\omega(t)} dP(\omega)
\]

\[
= \exp \left( -\frac{1}{2} \sum_{s,t\in A} Q(s,t) \lambda(s) \lambda(t) + i \sum_{s\in A} c(s) \lambda(s) \right)
\]

for all \( \lambda \in \mathbb{R}^4 \). From this it follows that \( \{X_t\}_{t\in T} \) is a Gaussian random field satisfying Eq. (27.7). \( \blacksquare \)

Exercise 27.2. Suppose \( T = [0, \infty) \) and \( \{X_t : t \in T\} \) is a mean zero Gaussian random field (process). Show that \( B_{[0,\sigma]} \perp B_{(\sigma,\infty)} \) for all \( 0 \leq \sigma < \infty \) iff

\[
Q(s,\sigma) Q(\sigma,t) = Q(s,\sigma) Q(s,t) \ \forall \ 0 \leq s \leq \sigma \leq t < \infty. \quad (27.8)
\]

Hint: see use Exercises 11.12 and 11.11

27.3 Gaussian Field Interpretation of Pre-Brownian Motion

Lemma 27.7. Suppose that \( \{B_t\}_{t \geq 0} \) is a pre-Brownian motion as described in Proposition 26.21, also see Corollary 20.25. Then \( \{B_t\}_{t \geq 0} \) is a mean zero Gaussian random process with \( E[B_tB_s] = s \wedge t \) for all \( s, t \geq 0 \).

Proof. Suppose we are given \( 0 = t_0 < t_1 < \cdots < t_n < \infty \) and recall from Proposition 26.21 that \( B_0 = 0 \) a.s. and \( \{B_{t_j} - B_{t_{j-1}}\}_{j=1}^n \) are independent mean zero Gaussian random variables. Hence it follows from Corollary 13.29 that \( \{B_{t_j} - B_{t_{j-1}}\}_{j=1}^n \) is a Gaussian random vector. Since the random vector \( \{B_{t_j}\}_{j=1}^n \) is a linear transformation \( \{B_{t_j} - B_{t_{j-1}}\}_{j=1}^n \), it follows from Lemma 12.42 that \( \{B_{t_j}\}_{j=1}^n \) is a Gaussian random vector. Since \( 0 = t_0 < t_1 < \cdots < t_n < \infty \) was arbitrary, it follows that \( \{B_t\}_{t \geq 0} \) is a Gaussian process. Since \( B_t = B_t - B_0 + N(0,t) \) we see that \( E B_t = 0 \) for all \( t \). Moreover we have for \( 0 \leq s < t < \infty \) that

\[
E[B_tB_s] = E[(B_t - B_s + B_s - B_0)(B_s - B_0)]
\]

\[
= E[(B_t - B_s)(B_s - B_0)] + E[(B_s - B_0)^2] = E[B_t - B_s] \cdot E[B_s - B_0] + s = 0 \cdot 0 + s
\]

which completes the proof. \( \blacksquare \)

Theorem 27.8. The function \( Q(s, t) := s \wedge t \) defined on \( s, t \geq 0 \) 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 \( EX_j = 0 \) and \( \text{Var}(X_j) = t_j - t_{j-1} \). Let \( Y_j := X_1 + \cdots + X_j \) for \( j = 1, 2, \ldots, n \). We then have, for \( j \leq k \) that

\[
\text{Cov}(Y_j, Y_k) = \sum_{m \leq j, \ n \leq k} \text{Cov}(X_m, X_n) = \sum_{m \leq j, \ n \leq k} \delta_{m,n} (t_m - t_{m-1}) = \sum_{m \leq j} (t_m - t_{m-1}) = t_j,
\]

i.e. \( t_j \wedge t_k = \text{Cov}(Y_j, Y_k) \). But such covariance matrices are always positive definite. Indeed,

\[
\sum_{j,k \leq n} t_j \wedge t_k \lambda_j \lambda_k = \sum_{j,k \leq n} \lambda_j \lambda_k \text{Cov}(Y_j, Y_k) = \text{Var}(\lambda_1 Y_1 + \cdots + \lambda_n Y_n) \geq 0
\]

with equality holding iff \( \lambda_1 Y_1 + \cdots + \lambda_n Y_n = 0 \) from which it follows that

\[
0 = E[Y_j (\lambda_1 Y_1 + \cdots + \lambda_n Y_n)] = \lambda_j (t_j - t_{j-1}),
\]

i.e. \( \lambda_j = 0 \).

2. According to Exercise 24.13 we can find stochastic processes \( \{B_n(t) = \sqrt{nS_{[nt]}}\}_{n=1}^\infty \) such that \( E[B_n(t)B_n(s)] \to s \wedge t \) as \( n \to \infty \) and therefore

\[
\sum_{s,t \in A} (s \wedge t) \lambda_s \lambda_t = \lim_{n\to\infty} \sum_{s,t \in A} E[B_n(t)B_n(s)] \lambda_s \lambda_t = \lim_{n\to\infty} E \left[ \left( \sum_{\ell \in A} B_n(t) \lambda_\ell \right)^2 \right] \geq 0.
\]
3. Appealing to Corollary \[\text{Corollary 20.25}\] there exists a time homogeneous Markov processes \(\{B_t\}_{t \geq 0}\) with Markov transition kernels given by

\[
Q_t(x, dy) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2} dy.
\] (27.9)

It is now easy to see that \(s \wedge t = \text{Cov} (B_s, B_t)\) which is automatically non-negative as we saw in the proof of item 2.

4. Let \(A = \{0 < t_1 < \cdots < t_n < \infty\}\) and \(\{\lambda_i\}_{i=1}^n \subset \mathbb{R}\) be given. Further let \(\alpha_j := \lambda_1 + \lambda_{i+1} + \cdots + \lambda_n\) with the convention that \(\alpha_{n+1} = 0\). We then have,

\[
\sum_{i=1}^n t_i \wedge t_j \lambda_i = \sum_{i=1}^n t_i \wedge t_j (\alpha_i - \alpha_{i+1}) = \sum_{1 \leq i \leq j} [t_i \wedge t_j - t_{i-1} \wedge t_j] \alpha_i = \sum_{1 \leq i \leq j} [t_i - t_{i-1}] \alpha_i
\]

where \(t_0 := 0\). Hence it follows that

\[
\sum_{i,j=1}^n t_i \wedge t_j \lambda_i \lambda_j = \sum_{i=1}^n \sum_{1 \leq i \leq j \leq n} [t_i - t_{i-1}] \alpha_i \lambda_j = \sum_{1 \leq i \leq j \leq n} [t_i - t_{i-1}] \alpha_i \lambda_j
\]

with equality iff \(\alpha_i = 0\) for all \(i\) which is equivalent to \(\lambda_i = 0\) for all \(i\).

5. Let \(h_\tau (\tau) := t \wedge \tau\) be as after Theorem \[\text{Theorem 27.8}\] below and using the results and notation proved there we find,

\[
\sum_{s,t \in \mathcal{A}} (s \wedge t) \lambda_s \lambda_t = \sum_{s,t \in \mathcal{A}} \langle h_s, h_t \rangle_T \lambda_s \lambda_t = \left\| \sum_{s \in \mathcal{A}} \lambda_s h_s \right\|^2_T \geq 0.
\]

This shows \(Q\) is positive semi-definite and equality holds iff \(\sum_{t \in \mathcal{A}} \lambda_t h_t = 0\). 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 \[\text{Corollary 20.25}\] Given Theorem \[\text{Theorem 27.8}\] we can also give a Gaussian process proof of the existence of pre-Brownian motion which we summarize in the next proposition.

**Proposition 27.9 (Pre-Brownian process).** Let \(\{B_t\}_{t \geq 0}\) be a mean zero Gaussian process such that \(\text{Cov} (B_s, B_t) = s \& t\) for all \(s, t \geq 0\) and let \(B_t := \sigma (B_s : s \leq t)\) and \(B_{t+} := \bigcap_{\sigma > t} B_\sigma\) Then;

1. \(B_0 = 0\) a.s.
2. \(\{B_t\}_{t \geq 0}\) has independent increments with \(B_t - B_s = d N (0, (t-s))\) for all \(0 \leq s < t < \infty\).
3. For all \(t \geq s \geq 0\), \(B_t - B_s\) is independent of \(B_{s+}\).
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. (27.9). [\[\text{Proof}\]. See Exercise 27.3 - 27.5\]

**Exercise 27.3 (Independent increments).** Let \(\mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\}\)

be a partition of \([0,T]\), \(\Delta_t B := B_t - B_{t-}\) and \(\Delta t := t_i - t_{i-1}\). Show \(\{\Delta_i B\}_{i=1}^n\) are independent mean zero normal random variables with \(\text{Var}(\Delta_i B) = \Delta_i t\).

**Exercise 27.4 (Increments independent of the past).** Let \(B_t := \sigma (B_s : s \leq t)\). For each \(s \in (0,\infty)\) and \(t > s\), show;

1. \(B_t - B_s\) is independent of \(B_s\) and
2. more generally show, \(B_t - B_s\) is independent of \(B_{s+} := \bigcap_{\sigma > s} B_\sigma\).

**Exercise 27.5 (The simple Markov property).** Show \(B_t - B_s\) is independent of \(B_s\) for all \(t \geq s\). Use this to show, for any bounded measurable function, \(f : \mathbb{R} \to \mathbb{R}\) that

\[
\mathbb{E} [f (B_t) | B_{s+}] = \mathbb{E} [f (B_t) | B_s] = \mathbb{E} [f (B_t) | B_s] = (p_t - s * f) (B_s) = (t - s) \Delta / 2 f (B_s)\text{ a.s.,}
\]

where

\[
p_t (x) := \frac{1}{\sqrt{2 \pi t}} e^{-\frac{x^2}{2t}}
\]

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 \(\frac{\Delta}{2} = \frac{1}{2} \frac{d^2}{dx^2}\) as there “infinitesimal generator.”

**Exercise 27.6.** 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
\[ \mathbb{E} [ f ( B_{t_1}, \ldots, B_{t_n} ) ] = \int_{\mathbb{R}^n} f ( x_1, \ldots, x_n ) q_p ( x ) \, dx \]

where

\[ q_p ( x ) := p_{t_1} ( x_1 ) p_{t_2 - t_1} ( x_2 - x_1 ) \cdots p_{t_n - t_{n-1}} ( x_n - x_{n-1} ) . \]

**Hint:** Either use Exercise 27.3 by writing

\[ f ( x_1, \ldots, x_n ) = g ( x_1, x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1} ) \]

for some function, \( g \) or use Exercise 27.5 first for functions, \( f \) of the form,

\[ f ( x_1, \ldots, x_n ) = \prod_{j=1}^{n} \varphi_j ( x_j ) . \]

**Better yet,** do it by both methods!
Version 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 28.1 (Versions).** Suppose, \(X_t : \Omega \to S\) and \(\tilde{X}_t : \Omega \to S\) are two processes defined on \(T\). We say that \(X\) 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 28.2.** We say two processes are indistinguishable iff \(P^* (Y \neq X) = 0\), i.e., iff there is a measurable set, \(E \subset \Omega\), such that \(P (E) = 0\) and \(\{Y \neq X\} \subset E\) where

\[
\{Y \neq X\} = \{\omega \in \Omega : Y_t (\omega) \neq X_t (\omega) \text{ for some } t \in [0, \infty)\}
\]

\[= \bigcup_{t \in (0, \infty)} \{\omega \in \Omega : Y_t (\omega) \neq X_t (\omega)\}. \tag{28.1}
\]

So \(Y\) is a modification of \(X\) iff

\[
0 = \sup_{t \in T} P \left( X_t \neq \tilde{X}_t \right) = \sup_{t \in \mathbb{T}} P \left( \left\{ \rho \left( X_t, \tilde{X}_t \right) > 0 \right\} \right)
\]

while \(Y\) is indistinguishable for \(X\) iff

\[
0 = P^* \left( X_t \neq \tilde{X}_t \quad \forall t \right) = P^* \left( \left\{ \sup_{t \in T} \rho \left( X_t, \tilde{X}_t \right) > 0 \right\} \right).
\]

Thus the formal difference between the two notions is simply whether the supremum is taken outside or inside the probabilities. See Exercise 28.1 for an example of two processes which are modifications of each other but are not indistinguishable.

**Exercise 28.1.** 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 \(E := \{Y \neq X\}\) is a measurable set such that \(P (E) = 0\) and hence \(X\) and \(Y\) are indistinguishable. **Hint:** replace the union in Eq. (28.1) by an appropriate countable union.

**Exercise 28.2.** 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, \(\{\tilde{X}_t(N)\}_{0 \leq t < N}\) of \(\{X_t\}_{0 \leq t < N}\). Show that \(X\) admits a right continuous modifications, \(\tilde{X}\), defined for all \(t \geq 0\).

### 28.1 Kolmogorov’s Continuity Criteria

Let \(\mathbb{D}_n := \{\frac{i}{2^n} : i \in \mathbb{Z}\}\) and \(\mathbb{D} := \bigcup_{n=0}^\infty \mathbb{D}_n\) be the dyadic rational numbers.

**Lemma 28.3.** Let \(\mathbb{D}_+ = \mathbb{D} \cap [0, \infty)\) and \(s \in \mathbb{D}_+ \cap [0, \infty)\) 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

\[
s = \frac{i}{2^n} + \frac{\sum_{k=1}^\infty a_k}{2^{n+k}},
\]

where \(a_k = a_k (n, s) \in \{0, 1\}\) with \(a_k = 0\) for all sufficiently large \(k\).

**Example 28.4.** Suppose that \(s = 85/32 = 85/2^5 \in \mathbb{D}\) and \(2 \in \mathbb{N}_0\) are given, then \(2^2s = 85/8 = 10 + 5/8\), 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 \(\frac{1}{2^n} + \frac{a_2}{2^{n+1}} \leq s\) and \(0\) otherwise, then choose \(a_2 = 1\) if \(\frac{1}{2^n} + \frac{a_2}{2^{n+1}} + \frac{a_3}{2^{n+2}} \leq s\) and \(0\) otherwise, etc. It is easy to check that \(s_m := \frac{1}{2^n} + \sum_{k=1}^m \frac{a_k}{2^{n+k}}\) 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 \(s, s_m \in \mathbb{D}_{m+n}\).
Moreover there exists a unique continuous function for some and this extension is still α and observe that if ∑ γ

If n ≥ 0 then

Suppose now that (S, ρ) is a metric space and x : Q := [0, 1] → S. For n ∈ N let

\[ \Delta_n(x) = \max \{ \rho(x(2^{-n}), x((i - 1)2^{-n}))) : 1 \leq i \leq 2^n \} \]

= \max \{ \rho(x(t), x(s)) : s, t \in Q \cap \mathbb{D}_n \text{ with } |s - t| \leq \frac{1}{2^n} \}.

If γ ∈ (0, 1) and x : \mathbb{D}_1 := \mathbb{D} \cap [0, 1] → S is a γ - Hölder continuous function, i.e.

\[ \rho(x(t), x(s)) \leq K |t - s|^\gamma \]

for some and K < ∞, then \( \Delta_n(x) \leq K2^{-n\gamma} \) for all n and in particular for all α ∈ (0, γ) we have

\[ \sum_{n=0}^{\infty} 2^{n\alpha} \Delta_n(x) \leq K \sum_{n=0}^{\infty} 2^{n\alpha} 2^{-n\gamma} = K \left( 1 - \frac{1}{2^{\gamma-\alpha}} \right)^{-1} < \infty. \]

Our next goal is to produce the following “converse” to this statement.

**Lemma 28.5.** Suppose \( \alpha > 0 \) and x : Q := [0, 1] → S is a map such that \( \sum_{n=0}^{\infty} 2^{n\alpha} \Delta_n(x) < \infty \), then

\[ \rho(x(t), x(s)) \leq 2^{1+\alpha} \cdot \left[ \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(x) \right] |t - s|^\alpha \quad \text{for all } s, t \in Q. \tag{28.2} \]

Moreover there exists a unique continuous function \( \hat{x} : [0, 1] \to S \) extending x and this extension is still α - Hölder continuous. (When \( \alpha > 1 \) it follows by Exercise 28.3 that x(t) is constant.)

**Proof.** Let s, t ∈ Q with s < t and choose n so that

\[ \frac{1}{2^{n+1}} < t - s \leq \frac{1}{2^n} \tag{28.3} \]

and observe that if \( s \in [i2^{-n}, (i + 1)2^{-n}] \), then \( t \in [i2^{-n}, (i + 2)2^{-n}] \) and therefore

\[ s = \frac{i}{2^n} + \sum_{k=1}^{\infty} \frac{a_k}{2^{n+k}} \quad \text{and} \quad t = \frac{j}{2^n} + \sum_{k=1}^{\infty} \frac{b_k}{2^{n+k}} \]

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 \( |s_m - s_{m+1}| \leq 2^{-(n+m+1)} \), it follows from the definition of \( \Delta_{n+m+1} \) that \( \rho(x(s_{m+1}), x(s_m)) \leq \Delta_{n+m+1}(x) \) which combined with the triangle inequality shows

\[ \rho(x(s), x(s_0)) \leq \sum_{m=1}^{N} \rho(x(s_m), x(s_{m-1})) \leq \sum_{m=1}^{\infty} \Delta_{n+m}(x). \]

Similarly \( \rho(x(t), x(t_0)) \leq \sum_{m=1}^{\infty} \Delta_{n+m}(x) \) while \( \rho(x(s_0), x(t_0)) \leq \Delta_n(x) \).

One more application of the triangle inequality now shows

\[ \rho(x(t), x(s)) \leq \Delta_n(x) + 2 \cdot \sum_{m=1}^{\infty} \Delta_{n+m}(x) \]

\[ \leq 2 \cdot \sum_{k=n}^{\infty} \Delta_k(x) = 2 \cdot \sum_{k=n}^{\infty} 2^{-\alpha k} \cdot \alpha k \Delta_k(x) \]

\[ \leq 2 \cdot (2^{-n})^\alpha \sum_{k=n}^{\infty} 2^{\alpha k} \Delta_k(x). \]

Combining this with the lower bound in Eq. (28.3) in the form

\[ (2^{-n})^\alpha = 2^\alpha \left( 2^{-(n+1)} \right)^\alpha < 2^\alpha (t - s)^\alpha, \]

gives the estimate in Eq. (28.2).

For the last assertion we define \( \hat{x}(t) := \lim_{Q \ni x \to t} x(s) \). 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. (28.2) 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

\[ \rho(\hat{x}(t), \hat{x}(s)) \leq 2^{1+\alpha} \cdot \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(x) \cdot |t - s|^\alpha \quad \text{for all } s, t \in [0, 1], \]

which shows that \( \hat{x} \) is Hölder continuous. As we had no choice but to define \( \hat{x} \) the way we did if \( \hat{x} \) is to be continuous, the extension is unique.

**Exercise 28.3.** Show if x : Q → S is α - Hölder continuous for some \( \alpha > 1 \), then x is constant.

**Notation 28.6.** For \( 0 < \alpha < 1 \) and x : Q → S, let

\[ K_\alpha(x) := 2^{(1+\alpha)} \sum_{k=0}^{\infty} 2^{\alpha k} \Delta_k(x). \tag{28.4} \]
Theorem 28.7 (Kolmogorov’s Continuity Criteria). Let \( \{X_t\}_{t \in \Omega} \) be an \( S \)-valued stochastic process and suppose there exists \( \gamma, \varepsilon > 0 \) such that
\[
\mathbb{E}[\rho(X_t, X_s)] \leq C |t-s|^{1+\varepsilon} \quad \text{for all} \ s,t \in \Omega. \tag{28.5}
\]
Then for all \( \alpha \in (0, \varepsilon/\gamma) \),
\[
\rho(X_t, X_s) \leq K_\alpha(X)|t-s|^{\alpha} \quad \text{for all} \ s,t \in \Omega \tag{28.6}
\]
where \( K_\alpha(X) \) is as in Eq. (28.4) satisfies \( K_\alpha(X) \in L^\gamma(P) \) and in fact,
\[
\|K_\alpha(X)\|_\gamma \leq \frac{C^{1/\gamma} \cdot 2^{(1+\alpha)}}{1-2^{\alpha-\varepsilon/\gamma}} < \infty. \tag{28.7}
\]

Proof. According to Exercise 28.4 below when \( \gamma < 1 \) or more generally when \( \gamma < 1 + \varepsilon \) it actually follows that \( X_t = X_0 \) a.s. and therefore Eq. (28.6) 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 \( K_\cdot(\cdot) \) is defined as in Eq. (28.4). The estimate in Eq. (28.6) is now a consequence of Lemma 28.5. So it only remains to verify Eq. (28.7). From the following simple estimate,
\[
\Delta_k(X)^\gamma = \max_{1 \leq i \leq 2^k} \rho(X_{i2^{-k}}, X_{(i-1)2^{-k}})^\gamma \leq \sum_{i=1}^{2^k} \rho(X_{i2^{-k}}, X_{(i-1)2^{-k}})^\gamma,
\]
we find
\[
\mathbb{E}[\Delta_k(X)^\gamma] \leq \sum_{i=1}^{2^k} \mathbb{E}[\rho(X_{i2^{-k}}, X_{(i-1)2^{-k}})^\gamma] \leq 2^k \cdot C (2^{-k})^{1+\varepsilon} = C2^{-k\varepsilon} \tag{28.8}
\]
and therefore \( \|\Delta_k(X)\|_\gamma \leq C^{1/\gamma}2^{-k\varepsilon/\gamma} \). Combining this inequality with Minkowski’s inequality shows
\[
\|K_\alpha(X)\|_\gamma = 2^{(1+\alpha)} \left[ \sum_{k=0}^{\infty} \left( \sum_{\alpha} \|\Delta_k(X)\|_\gamma \right) \right] \leq 2^{(1+\alpha)} \sum_{k=0}^{\infty} 2^{\alpha k} \|\Delta_k(X)\|_\gamma \leq 2^{(1+\alpha)} C^{1/\gamma} \sum_{k=0}^{\infty} \left( 2^{\alpha-\varepsilon/\gamma} \right)^k = \frac{C^{1/\gamma} 2^{(1+\alpha)}}{1-2^{\alpha-\varepsilon/\gamma}} < \infty
\]
provided \( \alpha < \varepsilon/\gamma \).

Theorem 28.8 (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
\[
\mathbb{E}[\rho(X_t, X_s)] \leq C |t-s|^{1+\varepsilon} \quad \text{for all} \ s,t \in D. \tag{28.9}
\]
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
\[
\rho(\tilde{X}_t, \tilde{X}_s) \leq K_\alpha(X)|t-s|^{\alpha} \quad \text{for all} \ s,t \in D. \tag{28.10}
\]
(Again according to Exercise 28.4 we will have \( \tilde{X}_t = X_0 \) a.s. for all \( t \in D \) unless \( \gamma \geq 1 + \varepsilon \).

Proof. From Theorem 28.7 we know for all \( \alpha \in (0, \varepsilon/\gamma) \) there is a random variable \( K_\alpha(X) \in L^\gamma(P) \) such that
\[
\rho(X_t, X_s) \leq K_\alpha(X)|t-s|^{\alpha} \quad \text{for all} \ s,t \in D \cap \mathbb{D}.
\]
On the set \( \{K_\alpha(X) < \infty\}, \{X_t\}_{t \in D \cap \mathbb{D}} \) has a unique continuous extension to \( \mathbb{D} \) which we denote by \( \{\tilde{X}_t\}_{t \in D} \). Moreover this extension is easily seen to satisfy Eq. (28.10). Lastly we have for \( s \in D \cap \mathbb{D} \) and \( t \in D \) that
\[
\rho(\tilde{X}_t, \tilde{X}_s) \leq \liminf_{D \cap \mathbb{D} \ni \delta \to t} \mathbb{E}[\rho(X_t, X_s)] \leq \liminf_{D \cap \mathbb{D} \ni \delta \to t} C |t-s|^{1+\varepsilon} = 0.
\]
This certainly implies that \( \rho(\tilde{X}_t, \tilde{X}_s) = 0 \) a.s. for every \( t \in D \) and therefore that \( \tilde{X} \) is a modification of \( X \).

Our construction of Brownian motion in Theorem 29.3 below will give us an opportunity to apply Theorem 28.8. At this time let us observe that it is important that \( \varepsilon \) is greater than 0 in the previous two theorems.

Example 28.9. 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 25.11), one can show that for any \( k \in \mathbb{N}, \)
\[
\mathbb{E}[N_t - N_s]^k \sim \lambda^k |t-s| \quad \text{for} \quad |t-s| \quad \text{small}.
\]
Notice that we cannot use Eq. (28.11) for any \( k \in \mathbb{N} \) to satisfy the hypothesis of Theorem 28.8 which is good since \( \{N_t\}_{t \geq 0} \) is integer value and does not have a continuous modification. However, see Example 31.27 below where it is shown that \( \{N_t\}_{t \geq 0} \) has a right continuous modification.
Exercise 28.4. Let $T \in \mathbb{N}, D = [0, T] \subset \mathbb{R}, (S, \rho)$ be a complete separable metric space and suppose that $X_t : \Omega \rightarrow S$ is a process for $t \in D$. Assume there exists $\gamma > 0, C > 0,$ and $\varepsilon > 0$ such that $(1 + \varepsilon) / \gamma > 1$ and

$$E \left[ \rho \left( X_t, X_s \right) \right] \leq C |t - s|^{1 + \varepsilon} \text{ for all } s,t \in D.$$ 

Show $X_t = X_0$ a.s. for each $t \in D$. Hint: for $\gamma \in (0, 1)$ use the inequality $(a + b)^{\gamma} \leq a^{\gamma} + b^{\gamma}$ for all $a, b \geq 0$ while for $\gamma \geq 1$ use Minkowski’s inequality. (This Exercise was inspired by questions posed by Dennis Leung.)

28.2 Kolmogorov’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 (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 defining

$$\rho_\infty (x, y) := \max_{0 \leq t \leq 1} \rho (x (t), y (t)) \forall x, y \in C ([0, 1], S). \quad (28.12)$$

It is standard and not hard to verify that $(C ([0, 1], S), \rho_\infty)$ is also a complete separable metric space. Moreover the compact subset, $K$, of $(C ([0, 1], S), \rho_\infty)$ are precisely those sets which are closed, uniformly bounded, and equicontinuous, 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

$$K (\alpha, C) := \{ x \in C ([0, 1], S) : \rho (x (t), x (s)) \leq K |t - s|^{\alpha} \forall s,t \in [0, 1] \}$$

is a closed, uniformly bounded, and equicontinuous, subset of $C ([0, 1], S)$ and hence is compact.

Exercise 28.5. Show $(C ([0, 1], S), \rho_\infty)$ is separable. Hints:

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 $g (x) := f (x + a) - f (a)$. By looking at a picture or just noting that $g' (x) \leq 0$ since $f'$ is decreasing, it follows that $g$ is a decreasing function of $x$. In particular it follows that

$$f (b + a) - f (b) = g (b) \leq g (0) = f (a) - f (0) = f (a).$$

2. The Heine–Borel property means that closed and bounded sets are compact.

1. Choose a countable dense subset, $A$, of $S$ and then choose finite subset $A_n \subset A$ such that $A_n \uparrow A$.
2. Let $D_n := \{ \frac{2^k}{2^n} : 0 \leq k \leq 2^n \}$ and $D = \bigcup_{n=1}^{\infty} D_n$. Further let $F_n := \{ x : [0, 1] \rightarrow A_n \}$ such that $x |_{[\frac{k}{2^n}, \frac{k+1}{2^n}]}$ is constant for all $1 \leq k \leq 2^n$ and further suppose that $x |_{[\frac{2^n-1}{2^n}, \frac{2^n}{2^n}]}$ is constant.
3. Given $y \in C ([0, 1], S)$ and $\varepsilon > 0$, show there exists $n \in \mathbb{N}$ and an $x \in F_n$ such that $\rho_\infty (y, x) \leq \varepsilon$.
4. For $k, n \in \mathbb{N}$ let

$$F_n^k := \left\{ y \in C ([0, 1], S) : \min_{x \in F_n} \rho_\infty (y, x) \leq \frac{1}{k} \right\}$$

and let $\Gamma := \{ (k, n) \in \mathbb{N}^2 : F_n^k \neq \emptyset \}$. For each $(k, n) \in \Gamma$, choose a function, $y_{k,n} \in F_n^k$.
5. Now show that $\{ y_{k,n} : (k, n) \in \Gamma \}$ is a countable dense subset of $(C ([0, 1], S), \rho_\infty)$.

Theorem 28.10 (Tightness Criteria). Let $S$ be a complete metric space satisfying the Heine–Borel property. Suppose that $\{ B_n (t) : 0 \leq t \leq 1 \}^{\infty}_{n=1}$ is a sequence of $S$ – valued continuous stochastic processes and suppose there exists $\gamma, \varepsilon > 0$ and $C < \infty$ such that

$$\sup_n E \left[ \rho (B_n (t), B_n (s)) \right] \leq C |t - s|^{1 + \varepsilon} \text{ for all } 0 \leq s, t \leq 1$$

and for some point $s_0 \in S$ we have

$$\lim_{N \uparrow \infty} \sup_n P \left[ \rho (B_n (0), s_0) > N \right] = 0.$$ 

Then the collection of measures, $\{ \mu_n := \text{Law}_{P} (B_n) \}^{\infty}_{n=1}$ 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

$$\Delta_k (\omega) := \max \left\{ \rho \left( \omega \left( \frac{j-1}{2^k} \right), \omega \left( \frac{j}{2^k} \right) \right) : 1 \leq j \leq 2^k \right\}.$$

The assumptions of this theorem allows us to apply Theorem 28.7 in order to learn;

$$\sup_n E \left[ K_\alpha (B_n) \right] \leq M (C, \gamma, \varepsilon, \alpha) < \infty.$$
Now let \( \Omega_N \) denote those \( \omega \in C([0,1] , S) \) such that \( \rho(\omega(0) , s_0) \leq N \) and \( K_n(\omega) \leq N \). We then have that

\[
\mu_n(\Omega_N) = P(B_n \notin \Omega_N) = P(\rho(B_n(0) , s_0) > N \text{ or } K_n(B_n) > N) \\
\leq P(\rho(B_n(0) , s_0) > N) + P(K_n(B_n) > N) \\
\leq P[\rho(B_n(0) , s_0) > N] + \frac{1}{N^\gamma} EK_\alpha(B_n) \\
\leq P[\rho(B_n(0) , s_0) > N] + \frac{1}{N^\gamma} M(C, \gamma, \varepsilon, \alpha).
\]

From this inequality and the hypothesis of the theorem it follows that \( \lim_{n \to \infty} \sup_n \mu_n(\Omega_N) = 0 \). To complete the proof it suffices to observe that for \( \omega \in \Omega_N \) we have \( \rho(\omega(0) , s_0) \leq N \) and

\[
\rho(\omega(t) , \omega(s)) \leq K_n(\omega)|t-s|^\alpha \leq N|t-s|^\alpha \quad \forall \ 0 \leq s, t \leq 1.
\]

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 28.11 (Weak Convergence Theorem).** Keeping the notation above and further assume \( \{B_n(t) : 0 \leq t \leq 1\} \) is another \( S \)-valued continuous process. Then \( B_n \Rightarrow B_0 \) iff \( \{B_n\}_{n=1}^\infty \) is tight and \( B_n \overset{\text{f.d.}}{\Rightarrow} B_0 \), i.e. \( B_n \) converges to \( B_0 \) in the sense of finite dimensional distributions.

**Proof.** If \( B_n \Rightarrow B_0 \) then \( \{B_n\}_{n=1}^\infty \) is tight by Prokhorov’s Theorem. Moreover if \( f : \mathbb{R}^k \to \mathbb{R} \) is a bounded continuous function and \( 0 \leq t_1 < t_2 < \cdots < t_k \leq 1 \), then

\[
F(x) := f(x(t_1), \ldots , x(t_k)) \quad \text{for} \quad x \in C([0,1] , S)
\]

(28.15)
defines a bounded continuous (cylinder) function on \( C([0,1] , S) \). Therefore by the definition of weak convergence it follows that

\[
\lim_{n \to \infty} E[f(B_n(t_1), \ldots , B_n(t_k))] = \lim_{n \to \infty} E[F(B_n(\cdot))] = E[F(B_0(\cdot))]
\]

\[
= E[f(B_0(t_1), \ldots , B_0(t_k))]
\]

and we have shown \( B_n \overset{\text{f.d.}}{\Rightarrow} B_0 \).

For the converse we now suppose that \( \{B_n\}_{n=1}^\infty \) is tight and \( B_n \overset{\text{f.d.}}{\Rightarrow} B_0 \) 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 \( |E[F(B_n)] - E[F(B_0)]| \geq \varepsilon \) for infinitely many \( n \). Therefore there is a subsequence \( B'_{n_k} = B_{n_k} \) such that

\[
|E[F(B'_{n_k})] - E[F(B_0)]| \geq \varepsilon > 0 \quad \text{for all} \quad k \in \mathbb{N}.
\]

(28.16)

Moreover by the assumed tightness, Prokhorov’s Theorem allows us to pass to a further subsequence (still denoted by \( B'_{n_k} \)) if necessary so that \( B'_{n_k} \Rightarrow X \) 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. (28.16) then implies that

\[
|E[F(X)] - E[F(B_0)]| \geq \varepsilon > 0.
\]

(28.17)

On the other hand \( B_n \overset{\text{f.d.}}{\Rightarrow} B_0 \) 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. However, this then implies \( |E[F(X)] - E[F(B_0)]| = 0 \) which contradicts Eq. (28.17).
Brownian Motion I

Our next goal is to prove existence of Brownian motion and then describe some of its basic path properties.

**Definition 29.1 (Brownian Motion).** A Brownian motion \( \{B_t\}_{t \geq 0} \) is an adapted mean zero Gaussian random process on some filtered probability space, \( (\Omega, \mathcal{B}, \{B_t\}_{t \geq 0}, P) \), satisfying: 1) for each \( \omega \in \Omega \), \( t \to B_t(\omega) \) is continuous, and 2) \( E[B_t B_s] = t \wedge s \) for all \( s, t \geq 0 \).

(29.1)

So a Brownian motion is a pre-Brownian motion with continuous sample paths.

**Remark 29.2.** If no filtration is given, we can use the process to construct one. Namely, let \( B^0_t := \sigma(B_s : s \leq t) \) and replace \( \mathcal{B} \) by \( \mathcal{B}^0_t \) if necessary. We call \( \{B^0_t\} \) the raw filtration associated to \( \{B_t\} \).

**Theorem 29.3 (Wiener 1923).** Brownian motions exist. 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

\[
E \left( (\tilde{B}_t - \tilde{B}_s)^2 \right) = E \left( \tilde{B}_t^2 + \tilde{B}_s^2 - 2B_t \tilde{B}_s \right) = t + s - 2s = t - s.
\]

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 \( p \in [1, \infty) \),

\[
E \left| \tilde{B}_t - \tilde{B}_s \right|^p = (t-s)^{p/2} E |N|^p.
\] (29.2)

Hence an application of Theorem 28.8 shows, with \( \varepsilon = p > 2 \), \( \beta = p/2 - 1 \), \( \alpha \in \left(0, \frac{p^2 - 1}{p}\right) = (0, \frac{1}{2} - 1/p) \), there exists a modification, \( \tilde{B} \) of \( \tilde{B} \) such that

\[
|B_t - B_s| \leq C_{\alpha, T} |t-s|^\alpha \quad \text{for} \ s, t \in [0, T).
\]

By applying this result with \( T = N \in \mathbb{N} \), we find there exists a continuous version, \( \tilde{B} \), of \( B \) 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, \( (\Omega, \mathcal{B}, \{B_t\}_{t \geq 0}, P) \) and \( \mathcal{B}_t := \sigma(B_s : s \leq t) \).

29.1 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,

\[
B_n(t) := \frac{1}{\sqrt{n}} \left( S_{[nt]} + (nt - [nt]) X_{[nt]+1} \right)
\] (29.3)

where for \( \tau \in \mathbb{R}_+ \), \( \lfloor \tau \rfloor \) 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 29.4.** Let \( B \) be a standard Brownian motion, then \( B_n \overset{f.d.}{\to} B \).

**Proof.** In Exercise 24.13 you showed that \( \frac{1}{\sqrt{n}} S_{[nt]} \overset{f.d.}{\to} B \) as \( n \to \infty \). Suppose that \( 0 < t_1 < t_2 < \cdots < t_k < \infty \) are given and we let

\[
W_n := \frac{1}{\sqrt{n}} \left( S_{[nt_1]}, \ldots, S_{[nt_k]} \right),
\]

\[
Y_n := (B_n(t_1), \ldots, B_n(t_k))
\]

and \( \varepsilon_n := Y_n - W_n \in \mathbb{R}^k \). From Eq. (29.3) and Chebyshev’s inequality (for all \( \delta > 0 \)),

\[
P\left( |\varepsilon_n| > \delta \right) \leq \frac{1}{\delta} E \left[ \frac{1}{\sqrt{n}} \left( nt_i - [nt_i] \right) |X_{[nt_i]+1}| \right] \leq \frac{1}{\sqrt{n}} \frac{E|X_1|}{\delta} \to 0
\]
as \( n \to \infty \). This then easily implies that \( \varepsilon_n \overset{p}{\to} 0 \) as \( n \to \infty \) and therefore by Slutsky’s Theorem 24.43 it follows that \( Y_n = X_n + \varepsilon_n \overset{d}{\to} (B(t_1), \ldots, B(t_k)) \).

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 29.5 (Donsker’s invariance principle). Let Ω, B_n, and B be as above. Then for B_n → B, i.e.,
\[ \lim_{n \to \infty} E[F(B_n)] = E[F(B)] \]
for all bounded continuous functions F : Ω → 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 29.4. 2) Show the distributions of B_n are tight. A proof of the tightness may be based on Ottaviani’s maximal inequality in Corollary 23.52 [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 possess fourth moments.

Proposition 29.6 (Random walk approximation bounds). Suppose that \( \{X_n\}_{n=1}^{\infty} \subset L^4(P) \) are i.i.d. random variables with \( E[X_n] = 0, \ E[X_n^2] = 1 \) and \( \gamma := E[X_n^4] < \infty \). Then there exists \( C < \infty \) such that
\[ E|B_n(t) - B_n(s)|^4 \leq C |t-s|^2 \]
for all \( s, t \in \mathbb{R}_+ \). (29.4)

Exercise 29.1. Provide a proof of Proposition 29.6. Hints: Use the results of Exercise 13.6, namely that
\[ E|S_t|^4 = t\gamma + 3t(l-1), \]
(29.5)
to verify that Eq. (29.4) holds for \( s, t \in D_n := \frac{1}{n} \mathbb{N}_0 \). Take care of the case where \( s, t \geq 0 \) with \( |t-s| < 1/n \) by hand and finish up using these results along with Minkowski’s inequality.

Theorem 29.7 (Baby Donsker Theorem). Continuing the notation used in Proposition 29.6 and let \( T < \infty \) and B be a Brownian motion. Then
\[ B_n[0,T] \Rightarrow B[0,T], \]
(29.6)
i.e. \( \text{Law}(B_n[0,T]) \Rightarrow \text{Law}(B[0,T]) \) 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. (29.6) fails to hold there would exists \( g \in BC(\Omega_T) \) and a subsequence \( B'_{n_k} = B_{n_k} \) such that
\[ \varepsilon := \inf_k E[g(B'_{n_k}[0,T])] - Eg(B[0,T]) > 0. \]
(29.7)
Since \( B_n(0) = 0 \) for all \( n \) and the estimate in Eq. (29.4) of Proposition 29.6 holds, it follows from Theorem 28.10 that \( \{B_n[0,T]\}_{n=1}^\infty \) is tight. So by Prokhorov’s Theorem 24.89 there is a further subsequence \( B''_{n_k} = B'_{n_k} \), which is weakly convergent to some \( \Omega_T \)-valued process X. Replacing \( B'_{n_k} \) bit \( B''_{n_k} \) in Eq. (29.7) and then letting \( l \to \infty \) in the resulting equation shows
\[ |E[g(X[0,T])] - Eg(B[0,T])| \geq \varepsilon > 0. \]
(29.8)
On the other hand by Proposition 29.4 we know that \( B_n \overset{f.d.}{\to} B \) 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 28.1. However this is in contradiction to Eq. (29.8).

29.2 Path Regularity Properties of BM

Definition 29.8. Let \( (V, \|\cdot\|) \) be a normed space and \( Z \in C([0,T], V) \). For \( 1 \leq p < \infty \), the \( p \)-variation of \( Z \) is,
\[ v_p(Z) := \sup_{\Pi} \left( \sum_{j=1}^{n} \|Z_{t_j} - Z_{t_{j-1}}\|^p \right)^{1/p}. \]
where the supremum is taken over all partitions, \( \Pi := \{0 = t_0 < t_1 < \cdots < t_n = T \} \), of \([0,T]\).

Lemma 29.9. 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
\[ \|a\|_q = \sum_{j=1}^{n} a_j^{q+r} \leq \left( \max_j a_j \right)^r \left( \sum_{j=1}^{n} a_j^p \right) \leq \|a\|_p \cdot \|a\|_p = \|a\|_p^q, \]
wherein we have used,
\[ \max_j a_j = \left( \max_j a_j^p \right)^{1/p} \leq \left( \sum_{j=1}^{n} a_j^p \right)^{1/p} = \|a\|_p. \]
Proposition 29.12. Suppose that \( \{P_m\}_{m=1}^{\infty} \) is a sequence of partitions of \([0, T]\) such that \( P_m \subset P_{m+1} \) for all \( m \) and \( \text{mesh}(P_m) \to 0 \) as \( m \to \infty \). Then \( Q_m \to T \) a.s. where \( Q_m \) is defined as in Eq. (29.10).

**Proof.** It is always possible to find another sequence of partitions, \( \{P'_n\}_{n=1}^{\infty} \), of \([0, T]\) such that \( P'_n \subset P'_{n+1} \), mesh \( (P'_n) \to 0 \) as \( n \to \infty \), \# \( (P'_n) = \# (P'_n) + 1 \), and \( P_m = P'_m \) where \( \{n_m\}_{m=1}^{\infty} \) is a subsequence of \( \mathbb{N} \). If we let \( Q'_n \) denote the quadratic variations associated to \( P'_n \) and we can show \( Q'_n \to T \) a.s. then we will also have \( Q_m = Q'_m \to T \) a.s. as well. So with these comments we may now assume that \( \# (P_{n+1}) = \# (P_n) + 1 \).

We already know from Exercise 29.2 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,

\[
\mathcal{F}_m := \sigma (Q_m, Q_{m+1}, \ldots).
\]

To do this, suppose that \( P_{m+1} = P_m \cup \{v\} \) and \( u = t_{i-1}, v = t_{i+1} \in P_m \) such that \( u < v < w \). Let \( X := B_u - B_w \) and \( Y := B_v - B_u \). Then

\[
Q_m = Q_{m+1} - (B_v - B_u)^2 - (B_w - B_u)^2 + (B_v - B_u)^2
= Q_{m+1} - X^2 - Y^2 + (X + Y)^2
= Q_{m+1} + 2XY
\]

therefore,

\[
E[Q_m|\mathcal{F}_{m+1}] = Q_{m+1} + 2E[XY|\mathcal{F}_{m+1}].
\]

So to finish the proof it suffices to show \( E[XY|\mathcal{F}_{m+1}] = 0 \) a.s.

To do this let

\[
b_t := \begin{cases} 
B_t & \text{if } t \leq v \\
B_u - (B_t - B_u) & \text{if } t > v,
\end{cases}
\]

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 \( E [b_t b_s] = s \wedge t \). Thus \( \{b_t\}_{t \geq 0} \) is still a Brownian motion. Moreover, if \( Q_{m+n} (b) \) is the quadratic variation of \( b \) relative to \( P_{m+n} \), 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 \( (X, Y, Q_{m+1}, Q_{m+2}, \ldots) \) and \( (-X, Y, Q_{m+1}, Q_{m+2}, \ldots) \) have the same distribution, if we write

\[
E[XY|\mathcal{F}_{m+1}] = f (Q_m, Q_{m+2}, \ldots) \text{ a.s.,}
\]

then it follows from Exercise 17.7 that

\[
E[-XY|\mathcal{F}_{m+1}] = f (Q_m, Q_{m+2}, \ldots) \text{ a.s.}
\]

Hence we may conclude,

\[
E[XY|\mathcal{F}_{m+1}] = E[-XY|\mathcal{F}_{m+1}] = -E[XY|\mathcal{F}_{m+1}],
\]

and thus \( E[XY|\mathcal{F}_{m+1}] = 0 \) a.s.
Corollary 29.13. If $p < 2$, then $v_p (B_{[0,T]}) = \infty$ a.s.

**Proof.** Choose partitions, $\{P_m\}$, of $[0,T]$ such that $\lim_{m \to \infty} Q_m = T$ a.s. where $Q_m$ is as in Eq. (29.10) and let $\Omega_0 := \{\lim_{m \to \infty} Q_m = T\}$ so that $P (\Omega_0) = 1$. If $v_p (B_{[0,T]} (\omega)) < \infty$ for some $\omega \in \Omega_0$, then, with $b = B (\omega)$, we would have

$$Q_m (\omega) = \sum_{i=1}^{n_m} (\Delta^m b)^2 = \sum_{i=1}^{n_m} |\Delta^m b|^p \cdot |\Delta^m b|^{2-p}$$

$$\leq \max_i |\Delta^m b|^{2-p} \cdot \sum_{i=1}^{n_m} |\Delta^m b|^p \leq [v_p (b)]^p \cdot \max_i |\Delta^m b|^{2-p}$$

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 $v_p (B_{[0,T]}) = \infty$ on $\Omega_0$. \hfill \blacksquare

Remark 29.14. The reader may find a proof that Corollary 29.13 also holds for $p = 2$ in [10] Theorem 13.69 on p. 382. You should consider why this result is not in contradiction with Exercise 29.2 and Theorem 29.12

**Hint:** unlike the case of $p = 1$, when $p > 1$ the quantity:

$$v_p^H (Z) := \left( \sum_{j=1}^{n} \|Z_{t_j} - Z_{t_{j-1}}\|^p \right)^{1/p}$$

does not increase under refinement of partitions so that

$$v_p (Z) = \sup_H v_p^H (Z) \neq \lim_{|H| \to 0} v_p^H (Z)$$

when $p > 1$.

Corollary 29.15 (Roughness of Brownian Paths). A Brownian motion, $\{B_t\}_{t \geq 0}$, is not almost surely $\alpha$–Hölder continuous for any $\alpha > 1/2$.

**Proof.** According to Exercise 29.2, we may choose partition, $P_m$, such that mesh $(P_m) \to 0$ and $Q_m \to T$ a.s. If $B$ were $\alpha$–Hölder continuous for some $\alpha > 1/2$, then

$$Q_m = \sum_{i=1}^{n_m} (\Delta^m B)^2 \leq C \sum_{i=1}^{n_m} (\Delta^m t)^{2\alpha} \leq C \max (|\Delta t|^{2\alpha-1}) \sum_{i=1}^{n_m} \Delta^m t$$

$$\leq C \text{mesh} (P_m)^{2\alpha-1} T \to 0 \text{ as } m \to \infty$$

which contradicts the fact that $Q_m \to T$ as $m \to \infty$. \hfill \blacksquare

Lemma 29.16. For any $\alpha > 1/2$, $\limsup_{t \to 0} |B_t|/t^\alpha = \infty$ a.s. (See Exercise ?? below to see that $\alpha = 1/2$ would work as well.)

**Proof.** If $\limsup_{t \to 0} |\omega_t|/t^\alpha < \infty$ then there would exists $C < \infty$ such that $|\omega_t| \leq Ct^\alpha$ for all $t \leq 1$ and in particular, $|\omega_{1/n}| \leq Cn^{-\alpha}$ for all $n \in \mathbb{N}$. Hence we have shown

$$\left\{ \limsup_{t \to 0} |B_t|/t^\alpha < \infty \right\} \subset \bigcup_{C \in \mathbb{N}} \cap_{n \in \mathbb{N}} \{ |B_{1/n}| \leq Cn^{-\alpha} \}.$$ 

This completes the proof because,

$$P (\cap_{n \in \mathbb{N}} \{ |B_{1/n}| \leq Cn^{-\alpha} \}) \leq \liminf_{n \to \infty} P ( |B_{1/n}| \leq Cn^{-\alpha} )$$

$$= \liminf_{n \to \infty} P \left( \frac{1}{\sqrt{n}} |B_1| \leq Cn^{-\alpha} \right)$$

$$= \liminf_{n \to \infty} P ( |B_1| \leq Cn^{1/2-\alpha} ) = P ( |B_1| = 0 ) = 0$$

if $\alpha > 1/2$.

**Theorem 29.17 (Nowhere $1/2 + \varepsilon$–Hölder Continuous).** Let

$$W := \{ \omega \in C ([0, \infty) \to \mathbb{R}) : \omega (0) = 0 \},$$

$B$ denote the standard Brownian motion on $(0, \infty)$, and $\mu$ be Wiener measure on $(W, B)$, 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 $\mu^x (E_\alpha) = 0$, i.e. there exists a set $\tilde{E}_\alpha \in B$ such that

$$E_\alpha = \left\{ \inf_{0 \leq t \leq 1} \limsup_{h \to 0} \frac{|\omega (t+h) - \omega (t)|}{|h|^\alpha} < \infty \right\} \subset \tilde{E}_\alpha$$

and $\mu (\tilde{E}_\alpha) = 0$. 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 $\nu \in \mathbb{N}$ be chosen more specifically later. If $\omega \in E_\alpha$, then there exists, $t \in [0, 1]$, $C < \infty$, such that

$$|\omega (t) - \omega (s)| \leq C |t-s|^\alpha$$

for all $|s| \leq \nu + 1$.

For all $n \in \mathbb{N}$ we may choose $i \geq 0$ so that $|t - \frac{i}{n}| < \frac{1}{n}$. By the triangle inequality, for all $j = 1, 2, \ldots, \nu$, we have

$$\left| \sum_{j=1}^{\nu} \omega (t + \frac{j}{n}) - \omega (t + \frac{i}{n}) \right| \leq \sum_{j=1}^{\nu} C |\frac{j}{n} - \frac{i}{n}|^\alpha \leq \sum_{j=1}^{\nu} C (|j - i|/n)^\alpha \leq C \nu^{\alpha}.$$
Letting &Auml; on µ. The last limit in Eq. (29.13) is zero provided we choose
P we have shown that
Therefore, ω ∈ Eα implies there exists D ∈ N such that for all n ∈ N there exists i ∈ n such that
\[ \left| \omega \left( \frac{i + j}{n} \right) - \omega \left( \frac{i + j - 1}{n} \right) \right| \leq Dn^{-\alpha} \quad \forall \ j = 1, 2, \ldots, \nu. \]
Letting
\[ A_D := \bigcap_{n=1}^{\infty} \bigcup_{i \leq n} E_{\alpha} \bigcap_{j=1}^{\nu} \left\{ \omega : \left| \omega \left( \frac{i + j}{n} \right) - \omega \left( \frac{i + j - 1}{n} \right) \right| \leq Dn^{-\alpha} \right\}, \]
we have shown that Eα ⊆ ∪D∈N AD. We now complete the proof by showing P (AD) = 0. To do this, we compute,
\[ P (AD) \leq \lim_{n \to \infty} \inf \left( \bigcup_{i \leq n} \bigcap_{j=1}^{\nu} \left\{ \omega : \left| \omega \left( \frac{i + j}{n} \right) - \omega \left( \frac{i + j - 1}{n} \right) \right| \leq Dn^{-\alpha} \right\} \right), \]
\[ \leq \lim_{n \to \infty} \inf \sum_{i \leq n, j=1}^{\nu} P \left( \left| \omega \left( \frac{i + j}{n} \right) - \omega \left( \frac{i + j - 1}{n} \right) \right| \leq Dn^{-\alpha} \right) \]
\[ = \lim_{n \to \infty} \inf \left[ \sum_{i \leq n} \prod_{j=1}^{\nu} P \left( \left| \omega \left( \frac{i + j}{n} \right) - \omega \left( \frac{i + j - 1}{n} \right) \right| \leq Dn^{-\alpha} \right) \right] \]
\[ = \lim_{n \to \infty} \inf \left[ Cn^{\frac{1}{2} - \alpha} \nu \right] \nu \]
\[ = \nu \lim_{n \to \infty} \inf \left[ Cn^{\frac{1}{2} - \alpha} \right] \nu \]
\[ \leq \lim_{n \to \infty} \inf \left[ Cn^{\frac{1}{2} - \alpha} \right] \nu \] (29.13)
wherein we have used
\[ \mu (|N| \leq \delta) = \frac{1}{\sqrt{2\pi}} \int_{|x| \leq \delta} e^{-\frac{1}{2} x^2} dx \leq \frac{1}{\sqrt{2\pi}} 2\delta. \]
The last limit in Eq. (29.13) is zero provided we choose \( \alpha > \frac{1}{2} \) and \( \nu (\alpha - \frac{1}{2}) > 1. \)

We end this section with a often useful, albeit heuristic, interpretation of Wiener measure, \( \mu := \text{Law} (B(\cdot)) \). This interpretation emphasizes the Gaussian nature of \( \mu \).

**Theorem 29.18 (Fake Theorem!).** Let \( \mu := \text{Law} (B(\cdot)) \) thought of as a measure on \( C([0,1], \mathbb{R}) \) and let
\[ H := \left\{ x \in C ([0,1], \mathbb{R}) : x (0) = 0 \quad \text{and} \quad \int_0^1 |\dot{x} (t)|^2 \, dt < \infty \right\}. \]
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 \( \mathcal{P} := \{ 0 = t_0 < t_1 < \cdots < t_n = 1 \} \) be a partition of \([0,1]\), let
\[ H_\mathcal{P} := \{ x \in H : x (0) = 0 \quad \text{and} \quad x'' (t) = 0 \quad \forall t \notin \mathcal{P} \}, \]
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 \( 1 \leq i \leq n \). Now suppose that \( F : C ([0,1], \mathbb{R}) \to \mathbb{R} \) is a bounded continuous function. The key points are;
1. \( \max_{0 \leq t \leq 1} |B^P_t (\omega) - B_t (\omega)| \to 0 \) as mesh (\( \mathcal{P} \)) → 0, and therefore
\[ F \left( B^P_t \right) \to F (B_t) \] as mesh (\( \mathcal{P} \)) → 0.
2. By the dominated convergence theorem,
\[ \int_{C([0,1],\mathbb{R})} F (x) \, d\mu (x) := \mathbb{E} [F (B_t)] = \lim_{\text{mesh}(\mathcal{P}) \to 0} \mathbb{E} [F \left( B^\mathcal{P}_t \right)]. \]
3. We have
\[ \mathbb{E} \left[ F \left( B^\mathcal{P}_t \right) \right] = \int_{H_{\mathcal{P}}} f (x) \frac{1}{Z_{\mathcal{P}}} \exp \left( -\frac{1}{2} \int_0^1 |\dot{x} (t)|^2 \, dt \right) \, dm_{\mathcal{P}} (x) \]
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 mesh (\( \mathcal{P} \)) → 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 \( f (x (t_1), \ldots, x (t_n)) = F (x^\mathcal{P}) \), then
\[ \mathbb{E} \left[ F \left( B^\mathcal{P}_t \right) \right] = \mathbb{E} [f (B_{t_1}, \ldots, B_{t_n})] \]
\[ = \int_{\mathbb{R}^n} f (x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi \Delta_j}} \exp \left( -\frac{1}{2\Delta_j} |x_j - x_{j-1}|^2 \right) dx_j \]
where \( x_0 = 0 \) by convention. The key point is that the above expression may be written as
\[
E[f(B_{t_1}, \ldots, B_{t_n})] = \int_{H_P} f(x(t_1), x(t_2), \ldots, x(t_n)) \frac{1}{Z_P} \exp \left( -\frac{1}{2} \int_0^1 |\dot{x}(t)|^2 \, dt \right) dm_P(x)
\]
where
\[
Z_P := \prod_{j=1}^n \sqrt{2\pi \Delta_j} \text{ and } dm_P(x) := d[x(t_1)] \ldots d[x(t_n)].
\]
The reason this is true is that for \( x \in H_P, \)
\[
\dot{x}(t) = \frac{x(t_j) - x(t_{j-1})}{\Delta_j} \text{ for } t_{j-1} < t < t_j
\]
and therefore
\[
\int_0^1 |\dot{x}(t)|^2 \, dt = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |\dot{x}(t)|^2 \, dt = \sum_{j=1}^n \frac{|x(t_j) - x(t_{j-1})|^2}{\Delta_j}
\]
(29.14)
which is routine in all cases. Let us work out item 3. in detail to illustrate the method. For \( 0 < s < t, \)
\[
E[b_s b_t] = s t E[B_{t-s} B_{t-t}] = s t (s^{-1} \wedge t^{-1}) = s t \cdot t^{-1} = s.
\]
Notice that \( t \to b_t \) is continuous for \( t > 0, \) so to finish the proof we must show that \( \lim_{t \to 0} b_t = 0 \text{ a.s.} \) 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 29.1 we know that
\[
E := \{ \omega \in \Omega : b_t(\omega) \neq \tilde{b}_t(\omega) \text{ for some } t > 0 \}
\]
is a null set. Hence \( \omega \notin E \) it follows that
\[
\lim_{t \downarrow 0} b_t(\omega) = \lim_{t \downarrow 0} \tilde{b}_t(\omega) = 0.
\]

29.3 Scaling Properties of B. M.

**Theorem 29.19 (Transformations preserving B. M.).** Let \( \{B_t\}_{t \geq 0} \) be a Brownian motion and \( B_t := \sigma(B_s) : s \leq t \). 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. \( b_t := t B_{1/t} \) for \( t > 0 \) and \( b_0 = 0 \) is a Brownian motion. In particular, \( \lim_{t \downarrow 0} t B_{1/t} = 0 \text{ a.s.} \)
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 \( B_T. \)
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, \( E[b_s b_t] = s \wedge t \)
which is routine in all cases. Let us work out item 3. in detail to illustrate the method. For \( 0 < s < t, \)
\[
E[b_s b_t] = s t E[B_{t-s} B_{t-t}] = s t (s^{-1} \wedge t^{-1}) = s t \cdot t^{-1} = s.
\]

**Corollary 29.20 (B. M. Law of Large Numbers).** Suppose \( \{B_t\}_{t \geq 0} \) is a Brownian motion, then almost surely, for each \( \beta > 1/2, \)
\[
\limsup_{t \to \infty} \frac{|B_t|}{t^\beta} = \begin{cases} 0 & \text{if } \beta > 1/2, \\ \infty & \text{if } \beta \in (0, 1/2). \end{cases}
\]

**Proof.** Since \( b_t := t B_{1/t} \) for \( t > 0 \) and \( b_0 = 0 \) is a Brownian motion, we know that for all \( \alpha < 1/2 \) there exists, \( C_\alpha(\omega) < \infty \) such that, almost surely,
\[
t |B_{1/t}| = |t B_{1/t}| = |b_t| \leq C_\alpha |t|^\alpha \text{ for all } t \leq 1.
\]
Replacing \( t \) by \( 1/t \) in this inequality implies, almost surely, that
\[
\frac{1}{t} |B_t| \leq C_\alpha |t|^{1-\alpha} \text{ for all } t \geq 1.
\]
or equivalently that
\[
|B_t| \leq C_\alpha t^{1-\alpha} \text{ for all } t \geq 1.
\]
(29.16)
Hence if \( \beta > 1/2, \) let \( \alpha < 1/2 \) such that \( \beta < 1 - \alpha. \) Then Eq. (29.15) follows from Eq. (29.16).

On the other hand, taking \( \alpha > 1/2, \) we know by Lemma 29.16 (or Theorem 29.17) that
\[
\limsup_{t \downarrow 0} \frac{|B_{1/t}|}{t^\alpha} = \limsup_{t \downarrow 0} \frac{|b_t|}{t^\alpha} = \infty \text{ a.s.}
\]
This may be expressed as saying
\[
\infty = \limsup_{t \to \infty} \frac{t^{-\alpha} |B_t|}{t^{-\alpha}} = \limsup_{t \to \infty} \frac{|B_t|}{t^{1-\alpha}} \text{ a.s.}
\]
Since \( \beta := 1 - \alpha \) is any number less than 1/2, the proof is complete.
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 \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+})\) 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 \( \sigma \) – algebra of \( \mathcal{B} \) and \( \mathcal{B}_s \subset \mathcal{B}_t \) for all \( s \leq t \).

### 30.1 Measurability Structures

**Notation 30.1** \((\mathcal{B}_t^\pm)\) Let \( \mathcal{B}_t^\pm = \mathcal{B}_t = \sigma(\bigcup_{t \in \mathbb{R}_+} \mathcal{B}_t) \subset \mathcal{B} \).

Also let \( \mathcal{B}_{t-} := \bigvee_{s > t} \mathcal{B}_s \).

Also let \( \mathcal{B}_{0-} := \mathcal{B}_0 \) and for \( t \in (0, \infty) \) let \( \mathcal{B}_t := \sigma(\bigvee_{s < t} \mathcal{B}_s) \).

(Observe that \( \mathcal{B}_{t-} = \mathcal{B}_t \).

The filtration, \( \{\mathcal{B}_t^\pm\}_{t \in \mathbb{R}_+} \) “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 30.2** (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 30.3** (Right continuous extension). Suppose \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0})\) is a filtered space and \( \mathcal{B}_t^+ := \mathcal{B}_{t+} := \bigcap_{s \geq t} \mathcal{B}_s \). Then \( \{\mathcal{B}_t^+\}_{t \geq 0} \) is right continuous. (We refer to \( \{\mathcal{B}_t^+\}_{t \in \mathbb{R}_+} \) as the right continuous filtration associated to \( \{\mathcal{B}_t\}_{t \in \mathbb{R}_+} \).

**Exercise 30.1.** Suppose \((\Omega, \mathcal{F})\) is a measurable space, \((S, \rho)\) is a separable metric space\(^1\) and \( S \) is the Borel \( \sigma \) – algebra on \( S \) – i.e. the \( \sigma \) – algebra generated by all open (or closed) balls of the form

\[
B(a, \varepsilon) := \{s \in S : \rho(s, a) < \varepsilon\} \quad (30.1)
\]

\[
(B(a, \varepsilon) := \{s \in S : \rho(s, a) \leq \varepsilon\} \quad (30.2)
\]

with \( a \in D \) and \( \varepsilon \in \mathbb{Q}_+ \).

1. Let \( D \subset S \) be a countable dense set and \( \mathbb{Q}_+ := \mathbb{Q} \cap \mathbb{R}_+ \). Show \( S \) may be described as the \( \sigma \) – algebra generated by all open (or closed) balls of the form

\[
B(a, \varepsilon) := \{s \in S : \rho(s, a) < \varepsilon\} \quad (30.1)
\]

\[
(B(a, \varepsilon) := \{s \in S : \rho(s, a) \leq \varepsilon\} \quad (30.2)
\]

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 \( \rho(x, \cdot) : \mathcal{S} \to \mathbb{R}_+ \) is a measurable map.

3. If \( X_n : \Omega \to S \) is a sequence of \( \mathcal{F}/\mathcal{S} \) – measurable maps such that \( X_n(\omega) \) exists in \( S \) for all \( \omega \in \Omega \), then the limiting function, \( X_\omega \), is \( \mathcal{F}/\mathcal{S} \) – measurable as well. (Hint: use item 2.)

**Definition 30.4.** Suppose \( S \) is a metric space, \( S \) is the Borel \( \sigma \) – algebra on \( S \), and \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+})\) is a filtered measurable space. A process, \( X_t : \Omega \to S \) for \( t \in \mathbb{R}_+ \) is:

1. adapted if \( X_t \) is \( \mathcal{B}_t/S \)– 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 \( \mathcal{B}_{[0,T]} \otimes \mathcal{B}_T/S \) – measurable.

**Lemma 30.5.** Let \( \varphi(t, \omega) := X_t(\omega) \) where we are continuing the notation in Definition 30.1. If \( X_t : \Omega \to S \) is a progressively measurable process then \( X_t \) is adapted and \( \varphi : \mathbb{R}_+ \times \Omega \to S \) is \( \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B} \) measurable and the \( \mathcal{X} \) is adapted. \( \square \)

**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 uncomfortable 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 $\eta_T^{-1}([0, a] \times A) = A \in B_T$ if $a = T$. This shows $\eta_T$ is $B_T \otimes B_{[0, T]} \otimes B - \text{measurable}$. Therefore, the composition, $\varphi^T \circ \eta_T = X_T$ is $B_T / S$ - measurable for all $T \in \mathbb{R}_+$ which is the statement that $X$ is adapted.

For $V \in S$ and $T < \infty$, we have

$$\varphi^{-1}(V) \cap ([0, T] \times \Omega) = (\varphi^T)^{-1}(V) \in B_{[0, T]} \otimes B_T \subset B_{[0, T]} \otimes B. \quad (30.3)$$

Since $B_{[0, T]} \otimes B$ and $(B_{\mathbb{R}_+} \otimes B)_{[0, T] \times \Omega}$ are $\sigma$ - algebras which are generated by sets of the form $[0, a] \times A$ with $a \in [0, T]$ and $A \in B$, they are equal - measurable. This observation along with Eq. (30.3) then implies,

$$\varphi^{-1}(V) \cap ([0, T] \times \Omega) \in \left(\bigcup_{n \in \mathbb{N}} \varphi^{-1}(V) \cap \left([0, T] \times \Omega\right)\right) \in B_{\mathbb{R}_+} \otimes B.$$

and therefore,

$$\varphi^{-1}(V) = \left(\bigcup_{n \in \mathbb{N}} \varphi^{-1}(V) \cap \left([0, T] \times \Omega\right)\right) \in B_{\mathbb{R}_+} \otimes B.$$

This shows $\varphi$ is $B_{\mathbb{R}_+} \otimes B / S$ - measurable as claimed. \hfill \blacksquare

**Lemma 30.6.** Suppose $S$ is a separable metric space, $S$ is the Borel $\sigma$ - algebra on $S$, $\left(\Omega, \mathcal{B}, \{B_t\}_{t \in \mathbb{R}_+}\right)$ is a filtered measurable space, and $X_t : \Omega \rightarrow S$ for $t \in \mathbb{R}_+$ is a adapted right continuous process. Then $X$ is progressively measurable and the map, $\varphi : \mathbb{R}_+ \times \Omega \rightarrow S$ defined by $\varphi(t, \omega) = X_t(\omega)$ is $[B_{\mathbb{R}_+} \otimes B] / S$ - measurable.

**Proof.** Let $T \in \mathbb{R}_+$. To each $n \in \mathbb{N}$ let $\varphi_n(0, \omega) = X_0(\omega)$ and

$$\varphi_n(t, \omega) := X_{\frac{k-1}{2^n}}(\omega) \quad \text{if} \quad \frac{(k-1)T}{2^n} < t \leq \frac{kT}{2^n} \quad \text{for} \quad k \in \{1, 2, \ldots, 2^n\}.$$

Then

$$\varphi_n^{-1}(A) = \left[0 \times X_n^{-1}(A)\right] \cup \left[\left(\frac{(k-1)T}{2^n}, \frac{kT}{2^n}\right) \times X^{-1}_{\frac{T-2n-1}{2^n}}(A)\right] \in B_{[0, T]} \otimes B_T,$$

showing that $\varphi_n$ is $[B_{[0, T]} \otimes B_T] / S$ - measurable. Therefore, by Exercise 30.1, $\varphi^T = \lim_{n \rightarrow \infty} \varphi_n$ is also $[B_{[0, T]} \otimes B_T] / S$ - measurable. The fact that $\varphi$ is $[B_{\mathbb{R}_+} \otimes B] / S$ - measurable now follows from Lemma 30.5. \hfill \blacksquare

**Lemma 30.7.** Suppose that $T \in (0, \infty)$, $\Omega_T := C([0, T], \mathbb{R})$, and $F_T = \sigma(B^T_t : t \leq T)$, where $B^T_t(\omega) = \omega(t)$ for all $t \in [0, T]$ and $\omega \in \Omega_T$. Then:

1. The map, $\pi : \Omega \rightarrow \Omega_T$ defined by $\pi(\omega) := \omega|_{[0, T]}$ is $B_T / F_T$ - measurable.
2. A function, $F : \Omega \rightarrow \mathbb{R}$ is $B_T$ - measurable iff there exists a function, $f : \Omega_T \rightarrow \mathbb{R}$ which is $F_T$ - measurable such that $F = f \circ \pi$.
3. Let $\|\cdot\|_T := \max_{t \in [0, T]} |\omega(t)|$ so that $(\Omega_T, \|\cdot\|_T)$ is a Banach space. The Borel $\sigma$ - algebra, $B_{\Omega_T}$ on $\Omega_T$ is the same as $F_T$.
4. If $F = f \circ \pi$ where $f : \Omega_T \rightarrow \mathbb{R}$ is a $\|\cdot\|_T$ - continuous function, then $F$ is $B_T$ - measurable.

**Proof.** 1. Since $B^T_t \circ \pi = B_t$ is $B_T$ - measurable for all $t \in [0, T]$, it follows that $\pi$ is measurable.

2. Clearly if $f : \Omega_T \rightarrow \mathbb{R}$ is $F_T$ - measurable, then $F = f \circ \pi : \Omega \rightarrow \mathbb{R}$ is $B_T$ - measurable. For the converse assertion, let $\mathbb{H}$ denote the bounded $B_T$ - measurable functions of the form $F = f \circ \pi$ with $f : \Omega_T \rightarrow \mathbb{R}$ being $F_T$ - 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, $G(B_{t_1}, \ldots, B_{t_n}) = G(B_{t_1}^{T}, \ldots, B_{t_n}^{T}) \circ \pi$ 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 $B_T$ - measurable functions. For a general $B_T$ - measurable function, $F : \Omega \rightarrow \mathbb{R}$, the truncation by $N \in \mathbb{N}$, $F_N = \min(F \wedge N)$, is of the form $F_N = f_N \circ \pi$ for some $F_T$ - measurable function, $f_N : \Omega_T \rightarrow \mathbb{R}$. Since every $\omega \in \Omega_T$ extends to an element of $\omega \in \Omega$, it follows that $\lim_{N \rightarrow \infty} F_N(\omega) = \lim_{N \rightarrow \infty} F_N(\omega) = F(\omega)$ exists. Hence if we let $F := \lim_{N \rightarrow \infty} F_N$, we will have $F = F \circ \pi$ with $F$ being a $F_T$ - measurable function.

3. Recall that $B_{\Omega_T} = \sigma(\text{open sets})$. Since $B_t : \Omega = \mathbb{R}$ is continuous for all $t$, it follows that $\sigma(B_t^T) \subset B_{\Omega_T}$ for all $s$ and hence $F_T \subset B_{\Omega_T}$. Conversely, since

$$\|\cdot\| := \sup_{t \in [0, T]} |\omega(t)| = \sup_{t \in [0, T]} \left|B_t^T(\omega)\right|,$$

it follows that $\|\cdot - \omega_0\| = \sup_{t \in [0, T]} \left|B_t^T(\cdot) - \omega_0(t)\right|$ is $F_T$ - 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 $F_T$. By the classical Weierstrass approximation theorem we know that $\Omega$ is separable and hence we may now conclude that $F_T$ contains all open subsets of $\Omega$. This shows that $B_{\Omega_T} = \sigma(\text{open sets}) \subset F_T$.

4. Any continuous function, $f : \Omega \rightarrow \mathbb{R}$ is $B_{\Omega_T} = F_T$ - measurable and therefore, $F = f \circ \pi$ is $B_T$ - measurable since it is the composition of two measurable functions. \hfill \blacksquare

### 30.2 Stopping and optional times

**Definition 30.8.** A random time $T : \Omega \rightarrow [0, \infty)$ is a stopping time iff $\{T \leq t\} \in B_t$ for all $t \geq 0$ and is an optional time iff $\{T < t\} \in B_t$ for all $t \geq 0$. 
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 30.9.** 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 $\{B_t\}$ - optional time iff $T$ is a $\{B_t^+\}$ - stopping time. In particular, if $B_t$ is right continuous, i.e. $B_t^+ = B_t$, for all $t$, then the notion of optional time and stopping time are the same.

**Proof.** 1.

$$\{T \leq t\} = \{ \emptyset \text{ if } t < s \}$$

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

$$\{T < t\} = \cup_{n \in N} \{T \leq t_n\} \in B_{t -} \subset B_t.$$

This shows $T$ is an optional time.

3. If $T$ is $\{B_t\}$ - 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 $\{T \leq t\} \in B_{t +} = B_t^+$. Conversely if $T$ is an $\{B_t^+\}$ - stopping time, $t > 0$, and $t_n \in (0, t)$ with $t_n \uparrow t$, then $\{T \leq t_n\} \in B_{t_n} \subset B_t$ for all $n$ and therefore,

$$\{T < t\} = \cup_{n = 1}^{\infty} \{T \leq t_n\} \in B_t.$$

**Exercise 30.2.** Suppose, for all $t \in \mathbb{R}_+$, that $X_t : \Omega \to \mathbb{R}$ is a function. Let $B_t := B_t^X := \sigma(X_s : s \leq t)$ and $\mathcal{B} = \mathcal{B}_\infty := \vee_{0 \leq t < \infty} B_t$. (Recall that the general element, $\mathcal{A} \in \mathcal{B}_t$ is of the form, $\mathcal{A} = X^{-1}_A(A)$ where $A$ is a countable subset of $[0, \infty)$, $\bar{A} \subset \mathbb{R}$ is a measurable set relative to the product $\sigma$ - algebra on $\mathbb{R}^4$, and $X_t : \Omega \to \mathbb{R}^4$ is defined by, $X_t(\omega)(t) = X_t(\omega)$ for all $t \in \mathcal{A}$.) 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 30.10.** Given a process, $X_t : \Omega \rightarrow S$ and $A \subset S$, let

$$T_A(\omega) := \inf \{t > 0 : X_t(\omega) \in A\}$$

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, $D_A \subseteq T_A$ 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 $X_0(\omega) \notin A$.

In the sequel will typically assume that $(S, \rho)$ is a metric space and $S$ 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 $t \rightarrow X_t(\omega)$ is right continuous, then $T_A = D_A$. Indeed, if $X_0(\omega) \in A$, then by the right continuity of $X(\omega)$, we know that $\lim_{t \uparrow 0} X_t(\omega) = X_0(\omega) \in A$ 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 $X_0(\omega) \in \text{bd} (A)$, there is no need for $T_A(\omega) = 0$ and hence in this case, typically $D_A = T_A$.

**Proposition 30.11.** Suppose $\left(\Omega, \{B_t\}_{t \geq 0}, \mathcal{B} \right)$ is a filtered measurable space, $(S, \rho)$ is a metric space, and $X_t : \Omega \rightarrow S$ is a right continuous $\{B_t\}_{t \geq 0}$ - 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 $\{T_A < \infty\}$ $(\{D_A < \infty\})$, $X_{T_A} \in A \cap (X_{D_A} \in A)$.
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, $\{T_A \leq t\} \subset B_t$ for all $t > 0$ while $\{T_A = 0\} \in B_0$.

**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,

$$\{D_A < t\} = \bigcup_{\exists s < t} X_s^{-1}(A) \subset B_{t -} \subset B_t.$$
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 assume that $X$ is a continuous process and $A$ is a closed subset of $S$.

3. Observe that $D_A(t) > t$ iff $X_{[0,t]}(\omega) \cap A = \emptyset$. Since $X$ is continuous, $X_{[0,t]}(\omega)$ is a compact subset of $S$ and therefore

$$
\varepsilon := \rho (X_{[0,t]}(\omega), A) > 0
$$

where

$$
\rho (A, B) := \inf \{ \rho (a, b) : a \in A \text{ and } b \in B \}.
$$

Hence we have shown,

$$
\{ D_A > t \} = \bigcup_{n=1}^{\infty} \{ \omega : \rho (X_{[0,n]}(\omega), A) \geq 1/n \}
$$

which 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 $X_{[\delta,\epsilon]}(\omega) \cap A = \emptyset$ or equivalently iff for all $\delta \in (0,t)$, $\varepsilon := \rho (X_{[\delta,\epsilon]}(\omega), A) > 0$. Using these observations we find,

$$
\{ T_A > t \} = \bigcap_{n>1/t} \bigcup_{m=1}^{\infty} \{ \omega : \rho (X_{[1/n,1]}(\omega), A) \geq 1/m \}
$$

This shows $\{ T_A \leq t \} = \{ T_A > t \}^c \in \mathcal{B}_t$ for all $t > 0$. Since, for $t > 0$, $\{ T_A < t \} = \bigcup_{\varepsilon \in \mathbb{Q} \cap (0,t]} \{ T_A < s \} \in \mathcal{B}_t$ 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 $\{ T_A = 0 \} \in \mathcal{B}_{\mathbb{R}_+}$ rather than $\{ T_A = 0 \} \in \mathcal{B}_0$. It should be clear that, in general, $\{ T_A = 0 \} \notin \mathcal{B}_0$ for $\{ T_A = 0 \} \in \mathcal{B}_0$ if $f(\cdot, X)$ 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 $X_A$.

**Notation 30.12** If $\tau : \Omega \to [0, \infty]$ is a random $\mathcal{B}_\infty$ - measurable time, let

$$
\mathcal{B}_\tau := \{ A \in \mathcal{B}_\infty : \{ \tau \leq t \} \cap A \in \mathcal{B}_t \text{ for all } t \in [0, \infty] \}.
$$

and

$$
\mathcal{B}_{\tau^+} := \{ A \in \mathcal{B}_\infty : A \cap \{ \tau < t \} \in \mathcal{B}_t \text{ for all } t \in \infty \}.
$$

**Exercise 30.3.** 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 30.4.** Suppose $\tau : \Omega \to [0, \infty]$ is the constant function, $\tau = s$, show $\mathcal{B}_s = \mathcal{B}_0$ and $\mathcal{B}_{s^+} = \bigcap_{t>s} \mathcal{B}_t = \mathcal{B}_+$ so that the notation introduced in Notation [30.12] is consistent with the previous meanings of $\mathcal{B}_s$ and $\mathcal{B}_{s^+}$.

**Exercise 30.5.** Suppose that $\tau$ is an optional time and let

$$
\mathcal{B}_{\tau^+} := \{ A \in \mathcal{B}_\infty : A \cap \{ \tau \leq t \} \in \mathcal{B}_t = \mathcal{B}_t^+ \text{ for all } t \leq \infty \}.
$$

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 $\{ \mathcal{B}_t^+ \}$.

**Lemma 30.13.** 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 30.5 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 30.14.** If $\tau$ is a $\{ \mathcal{B}_t \}$ - stopping time and $X_\tau : \Omega \to S$ is a $\{ \mathcal{B}_t \}$ - progressively measurable process, then $X_\tau$ defined on $\{ \tau < \infty \}$ is $(\mathcal{B}_T)^{[\tau<\infty]}$ - measurable. Similarly, if $\tau$ is a $\{ \mathcal{B}_t \}$ - optional time and $X_\tau : \Omega \to S$ is a $\{ \mathcal{B}_t^+ \}$ - progressively measurable process, then $X_\tau$ defined on $\{ \tau < \infty \}$ is $(\mathcal{B}_{\tau^+})^{(\tau<\infty)}$ - measurable.

**Proof.** In view of Proposition 30.9 and Exercise 30.5 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 $\mathcal{B}_{[0,T]} \otimes \mathcal{B}_T$ - 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 $(\mathcal{B}_T)^{(\tau\leq T)} / \mathcal{B}_T$ - measurable.

For $V \in \mathcal{S}$ and $T \in \mathbb{R}_+$,

$$
X_\tau^{-1}(A) \cap \{ \tau \leq T \} = \{ \omega \in \Omega : \tau(\omega) \leq T \text{ and } X_\tau(\omega) \in A \}
$$

$$
= \{ \omega \in \Omega : \tau(\omega) \leq T \text{ and } \varphi_T(\omega) \in A \}
$$

$$
= \{ \tau \leq T \} \cap \{ \varphi_T \circ \psi_T \in A \} \in (\mathcal{B}_T)^{(\tau\leq T)} \subset \mathcal{B}_T.
$$

This is true for arbitrary $T \in \mathbb{R}_+$ we conclude that $X_\tau^{-1}(A) \in \mathcal{B}_T$ and since, by definition, $X_\tau^{-1}(A) \subset \{ \tau < \infty \}$, it follows that $X_\tau^{-1}(A) \in (\mathcal{B}_T)^{(\tau<\infty)}$. This completes the proof since $A \in \mathcal{S}$ was arbitrary.
Lemma 30.15 (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 $T \land S$, $T \lor S$, and $T + S$ are stopping times.
6. If $\{T_n\}_{n=1}^{\infty}$ are optional times, then

$$\sup_{n \geq 1} T_n, \inf_{n \geq 1} T_n, \liminf_{n \to \infty} T_n, \text{ and } \limsup_{n \to \infty} T_n$$

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

$$\{T + S \leq \tau\} = \{S = 0, T \leq \tau\} \cup \{0 < S, T + S \leq \tau\}$$

and $\{S = 0, T \leq \tau\} \in \mathcal{B}_\tau$. Hence it suffices to show $\{0 < S, T + S \leq \tau\} \in \mathcal{B}_\tau$. 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

$$\{S + T \leq \tau\} = \cup_{n \in \mathbb{N}} \cap_{n \geq m} \{r < S < r + 1/n, T \leq \tau - r : 0 < r < r + 1/n < \tau\}$$

which is in $\mathcal{B}_{\tau-} \subset \mathcal{B}_\tau$. In showing $\{0 < S, T + S \leq \tau\} \in \mathcal{B}_\tau$ we only need for $S$ and $T$ to be optional times and so if $S > 0$ and $T > 0$, then

$$\{T + S \leq \tau\} = \{0 < S, T + S \leq \tau\} \in \mathcal{B}_{\tau-} \subset \mathcal{B}_\tau.$$ 

5. If $T$ and $S$ are stopping times and $\tau \geq 0$, then

$$\{T \land S \leq \tau\} = \{T \leq \tau\} \cup \{S \leq \tau\} \in \mathcal{B}_\tau,$$

$$\{T \lor S \leq \tau\} = \{T \leq \tau\} \cap \{S \leq \tau\} \in \mathcal{B}_\tau,$$

and

$$\{S + T > \tau\} = \{T = 0, S > \tau\} \cup \{0 < T, T + S > \tau\}$$

$$= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \{T \geq \tau, T + S > \tau\}$$

$$= \{T = 0, S > \tau\} \cup \{0 < T < \tau, T + S > \tau\} \cup \{S = 0, T > \tau\} \cup \{S > 0, T \geq \tau\}.$$ 

The first, third, and fourth events are easily seen to be in $\mathcal{B}_\tau$. As for the second event,

$$\{0 < T \leq \tau, T + S > \tau\} = \cup \{r < T < \tau, S > \tau - r : r \in \mathbb{Q} \text{ with } 0 < r < \tau\}.$$

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 $\sup_{n} T_n$ is a $\mathcal{B}_{\tau}$ stopping time and hence $\sup_{n} T_n$ is an $\mathcal{B}_{\tau}$ stopping time and hence $\sup_{n} T_n$ is an $\mathcal{B}_{\tau}$ optional time, wherein we have used Proposition 30.9 twice.

Lemma 30.16 (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 $\tau \land t$ is $\mathcal{B}_t$ – measurable.
3. If $\sigma \leq \tau$, then $\mathcal{B}_\sigma \subset \mathcal{B}_\tau$.
4. $\{\mathcal{B}_\sigma\}_{\sigma \leq \tau} \subset \mathcal{B}_{\sigma \land \tau}$ and in particular $\{\sigma \leq \tau\}, \{\sigma < \tau\}, \{\tau \leq \sigma\},$ and $\{\tau < \sigma\}$ are all in $\mathcal{B}_{\sigma \land \tau}$.
5. $\{\mathcal{B}_\sigma\}_{\sigma \leq \tau} \subset \mathcal{B}_{\sigma \land \tau}$.
6. $\mathcal{B}_\sigma \cap \mathcal{B}_\tau = \mathcal{B}_{\sigma \land \tau}$.
7. If $\mathbb{D}$ is a countable set and $\tau : \Omega \rightarrow \mathbb{D} \subset [0, \infty]$ 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 $f : \Omega \rightarrow \mathbb{R}$ 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 $B$ is replaced by $B^+$.\[\]

**Proof.** Recall from Definition [17.27] that if $G$ is a $\sigma$-algebra on $\Omega$ and $A \subset \Omega$, then $G_A := \{B \cap A : B \in G\}$ is a sub-$\sigma$-algebra of $2^A$. Moreover if $G$ and $F$ are two $\sigma$-algebras on $\Omega$ and $A \in G \cap F$, then (by definition) $G = F$ on $A$ if $G_A = F_A$.

1. If $A \in B_\tau$, then $A \cap \{ \tau = t \} = A \cap \{ \tau \leq t \} \cap \{ \tau < t \}^c \in B_t$.

Conversely if $A \in B_t$ and $s \in \mathbb{R}_+$,

$$A \cap \{ \tau = t \} \cap \{ \tau \leq s \} = \begin{cases} \emptyset & \text{if } s < t \\ A \cap \{ \tau = t \} & \text{if } s \geq t \end{cases}$$

from which it follows that $A \cap \{ \tau = t \} \in B_\tau$.

2. To see $\tau \wedge t$ is $B_t$-measurable simply observe that

$$\{ \tau \wedge t \leq s \} = \begin{cases} \Omega \in B_t & \text{if } s \leq t \\ \{ \tau \leq s \} \in B_s \subset B_t & \text{if } t > s \end{cases}$$

and hence $\{ \tau \wedge t \leq s \} \in B_t$ for all $s \in [0, \infty]$.

3. If $A \in B_\sigma$ and $\sigma \leq \tau$, then

$$A \cap \{ \tau \leq t \} = [A \cap \{ \sigma \leq t \}] \cap \{ \tau \leq t \} \in B_t$$

for all $t \leq \infty$ and therefore $A \in B_\tau$.

4. If $A \in B_\sigma$ then $A \cap \{ \sigma \leq \tau \}$ is the generic element of $(B_\sigma)_{(\sigma \leq \tau)}$. We now have

$$(A \cap \{ \sigma \leq \tau \}) \cap \{ \tau \wedge s \leq t \} = (A \cap \{ \sigma \leq \tau \}) \cap \{ \sigma \leq t \}$$

$$= (A \cap \{ \sigma \wedge t \leq \tau \wedge t \}) \cap \{ \sigma \leq t \}$$

$$= (A \cap \{ \sigma \leq t \}) \cap \{ \sigma \wedge t \leq \tau \wedge t \} \in B_t$$

since $(A \cap \{ \sigma \leq t \}) \in B_t$ and $\sigma \wedge t$ and $\tau \wedge t$ are $B_t$-measurable and hence $\{ \sigma \wedge t \leq \tau \wedge t \} \in B_t$. Since $\Omega \in B_\sigma$, it follows from what we have just proved that $\{ \sigma \leq \tau \} = \Omega \cap \{ \sigma \leq \tau \} \in B_{\sigma \wedge \tau}$ and hence also $\{ \tau < \sigma \} = \{ \sigma \leq \tau \}^c \in B_{\sigma \wedge \tau}$. By symmetry we may also conclude that $\{ \tau \leq \sigma \}$ and $\{ \sigma < \tau \}$ are in $B_{\sigma \wedge \tau}$.

5. By item 4., if $A \in B_\sigma$, then

$$A \cap \{ \sigma < \tau \} = A \cap \{ \sigma \leq \tau \} \cap \{ \sigma < \tau \} \in B_{\sigma \wedge \tau}.$$
Lemma 30.18 (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
\[
\tau_n := \frac{1}{2^n} [2^n \tau] = \infty \mathbf{1}_{\tau=\infty} + \sum_{k=1}^{\infty} \frac{k}{2^n} \mathbf{1}_{\tau < \frac{k}{2^n}}.
\] (30.5)

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
\[
A \cap \{ \tau_n = k2^{-n} \} = A \cap \{ (k-1)2^{-n} \leq \tau < k2^{-n} \} = \left[ A \cap \{ \tau < k2^{-n} \} \right] \setminus \{ \tau < (k-1)2^{-n} \} \in \mathcal{B}_{k2^{-n}}.
\]

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 30.16. 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.

### 30.3 Filtration considerations

For this action suppose that \( \left( \Omega, \mathcal{B}, \{ \mathcal{B}_t \}_{t \geq 0}, P \right) \) is a given filtered probability space.

**Notation 30.19 (Null sets)** Let \( \mathcal{N}^P := \{ N \in \mathcal{B} : P(N) = 0 \} \) be the collection of null sets of \( P \).

**Definition 30.20.** If \( A \subset \mathcal{B} \) is a sub-sigma-algebra of \( \mathcal{B} \), then the *augmentation* of \( A \) is the \( \sigma \)-algebra, \( \tilde{\mathcal{A}} := \mathcal{A} \vee \mathcal{N} := \sigma(A \cup \mathcal{N}) \).

**Definition 30.21 (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, \( \left( \Omega, \mathcal{B}, P \right) \) is complete (i.e. if \( N \in \mathcal{N}^P \), \( A \in \mathcal{N} \), then \( A \in \mathcal{N}^P \)), then we say that \( \left( \Omega, \mathcal{B}, \{ \mathcal{B}_t \}_{t \geq 0}, P \right) \) 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 30.22 (Augmentation lemma).** Continuing the notation in Definition 30.20, we have
\[
\tilde{\mathcal{A}} := \{ B \in \mathcal{B} : \exists A \in \mathcal{A} \supseteq A \Delta B \in \mathcal{N} \}.
\] (30.6)

**Proof.** Let \( \mathcal{G} \) denote the right side of Eq. (30.6). If \( B \in \mathcal{G} \) and \( A \in \mathcal{A} \) such that \( N := A \Delta B \in \mathcal{N} \), then
\[
B = [A \cap B] \cup [A \setminus B] = [A \setminus (A \setminus B)] \cup [B \setminus A].
\] (30.7)

Since \( A \setminus B \subset \mathcal{N} \) and \( B \setminus A \subset \mathcal{N} \) implies \( A \setminus B \) and \( B \setminus A \) are in \( \mathcal{N} \), it follows that \( B \in \mathcal{A} \vee \mathcal{N} = \tilde{\mathcal{A}} \). Thus we have shown, \( \mathcal{G} \subset \tilde{\mathcal{A}} \). Since it is clear that \( \tilde{\mathcal{A}} \subset \mathcal{G} \) 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 \( \tilde{\mathcal{A}} = \mathcal{A} \vee \mathcal{N} \subset \mathcal{G} \).

Since \( A \cap B \subset \mathcal{N} \) and \( A \cup B \subset \mathcal{N} \), 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 \( A_j \Delta B_j \in \mathcal{N} \) for all \( j \). So letting \( A = \bigcup_j A_j \in \mathcal{A} \) and \( B = \bigcup_j B_j \in \mathcal{B} \), we have
\[
B \supset A \Delta \bigcup_j \{ A_j \Delta B_j \} \in \mathcal{N}
\]
from which we conclude that \( A \Delta B \in \mathcal{N} \) 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 30.23 (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 \( s > t \), \( \mathcal{B}_t \subseteq \mathcal{B}_s \) it follows that \( \overline{\mathcal{B}_{t+}} \subset \overline{\mathcal{B}_{s+}} \) and therefore that
\[
\overline{\mathcal{B}_{t+}} \subset \cap_{s>t} \overline{\mathcal{B}_s} = \overline{\mathcal{B}_{t+}}.
\]
Conversely if \( B \in \overline{\mathcal{B}_{t+}} = \cap_{s>t} \overline{\mathcal{B}_s} \) and \( t_n \uparrow 0 \) then for each \( n \in \mathbb{N} \) there exists \( A_n \in \mathcal{B}_{t_n} \) such that \( A_n \Delta B \in \mathcal{N} \). We will now show that \( B \in \overline{\mathcal{B}_{t+}} \), by showing \( B \Delta A \in \mathcal{N} \) where
\[
A := \{ A_n \mathbf{i.o.} \} = \bigcup_{n \in \mathbb{N}} \cap_{n \geq m} A_n \in \mathcal{B}_{t+}.
\]
To prove this let \( A'_m := \bigcup_{n \geq m} A_n \) so that \( A'_m \downarrow A \) as \( n \uparrow \infty \). Then
30.3.1 ***More Augmentation Results (This subsection need 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 30.24.** Suppose \((\Omega, B)\) is a measurable space. A collection of subsets, \(\mathcal{N} \subset B\) is a null like collection in \(B\) if: 1) \(\mathcal{N}\) is closed under countable union, and 2) if \(A \in B\) and there exists \(N \in \mathcal{N}\) such that \(A \subset N\), then \(A \in \mathcal{N}\).

**Example 30.25.** Let \(\{P_i\}_{i \in I}\) be any collection of probability measures on a measurable space, \((\Omega, B)\). Then

\[
\mathcal{N} := \{N \in B : P_i(N) = 0 \text{ for all } i \in I\}
\]

is a null like collections of subsets in \(B\).

**Example 30.26.** If \(\mathcal{N}\) is a null like collection in \(B\), then

\[
\mathcal{N}' := \{A \subset 2^\Omega : A \subset N \text{ for some } N \in \mathcal{N}\}
\]

is a null like collection in \(2^\Omega\).

**Example 30.27.** Let \(\{P_i\}_{i \in I}\) be any collection of probability measures on a measurable space, \((\Omega, B)\). Then

\[
\mathcal{N} := \{N \subset 2^\Omega : \exists B \in B \ni P_i(B) = 0 \text{ and } N \subset B \text{ for all } i \in I\}
\]

is a null like collections of subsets of \(2^\Omega\). Similarly,

\[
\mathcal{N} := \{N \subset 2^\Omega : \exists B_i \in B \ni P_i(B_i) = 0 \text{ and } N \subset B_i \text{ for all } i \in I\}
\]

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 30.28.** If \(\mathcal{N}_i \subset B\) are null like collections in \(B\) for all \(i \in I\), then \(\mathcal{N} := \cap_{i \in I} \mathcal{N}_i\) is another null like collection in \(B\). Indeed, if \(B \in B\) 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 30.29.** If \((\Omega, B)\) is a measurable space, \(A \subset B\) is a sub-sigma-algebra of \(B\) and \(\mathcal{N} \subset B\) is a null like collection in \(B\), we say \(A^\mathcal{N} := A \vee \mathcal{N}\) is the augmentation of \(A\) by \(\mathcal{N}\).

**Lemma 30.30 (Augmentation lemma).** If \(A\) is a sub-sigma-algebra of \(B\) and \(\mathcal{N}\) is a null like collection in \(B\), then the augmentation of \(A\) by \(\mathcal{N}\) is

\[
A^\mathcal{N} := \{B \in B : \exists A \in A \ni A \triangle B \in \mathcal{N}\}. 
\]

**Proof.** Let \(G\) denote the right side of Eq. (30.3). If \(B \in G\) and \(A \in A\) such that \(N := A \triangle B \in \mathcal{N}\), then

\[
B = [A \cap B] \cup [A \setminus B] = [A \setminus (A \setminus B)] \cup [B \setminus A].
\]

Since \(A \setminus B \subset N\) and \(B \setminus A \subset \mathcal{N}\) implies \(A \setminus B\) and \(B \setminus A\) are in \(\mathcal{N}\), it follows that \(B \in A \vee \mathcal{N} = A^\mathcal{N}\). Thus we have shown, \(G \subset A^\mathcal{N}\). Since it is clear that \(A \subset G\) and \(N \subset G\), to finish the proof it suffices to show \(G\) is a \(\sigma\) – algebra. For if we do this, then \(A^\mathcal{N} = A \vee N \subset G\).

Since \(A^\mathcal{N} \cap B^\mathcal{N} = A \triangle B\), we see that \(G\) is closed under complementation. Moreover, if \(B_j \in G\), there exists \(A_j \in A\) such that \(A_j \triangle B_j \in \mathcal{N}\) for all \(j\). So letting \(A = \bigcup_j A_j \in A\) and \(B = \bigcup_j B_j \in B\), we have

\[
B \ni A \triangle B \ni \bigcup_j [A_j \triangle B_j] \in \mathcal{N}
\]

from which we conclude that \(A \triangle B \in \mathcal{N}\) and hence \(B \in G\). This shows that \(G\) is closed under countable unions, complementation, and contains \(A\) and hence the empty set and \(\Omega\), thus \(G\) is a \(\sigma\) – algebra.

**Lemma 30.31 (Commutation lemma).** Let \((\Omega, B, \{B_t\}_{t \geq 0})\) be a filtered space and \(\mathcal{N} \subset B\) be null like collection and for \(G \subset B\), let \(G := G \vee \mathcal{N}\). Then \(\bar{B}_t = \bar{B}_t^{G}\).

**Proof.** Since for any \(s > t\), \(B_t \subset B_s\) it follows that \(\bar{B}_t^{G} \subset \bar{B}_s\) and therefore that

\[
\bar{B}_t^{G} \subset \cap_{s > t} \bar{B}_s = \bar{B}_t^{G}.
\]

Conversely if \(B \in \bar{B}_t^{G} = \cap_{s > t} \bar{B}_s\) and \(t_n > t\) such that \(t_n \downarrow 0\), then for each \(n \in \mathbb{N}\) there exists \(A_n \in B_{t_n}\) such that \(A_n \triangle B \in \mathcal{N}\). We will now show that \(B \in \bar{B}_t^{G}\), by shown \(B \triangle A \in \mathcal{N}\) where

\[
A := \{A_n \text{ i.o.} \} = \cap_{m \in \mathbb{N}} \cup_{n \geq m} A_n \in B_{t_t}.
\]
To prove this let $A'_n := \bigcup_{n \geq m} A_n$ so that $A'_n \downarrow A$ as $n \uparrow \infty$. Then

$$B \triangle A = B \triangle [\bigcap_n A'_m] = (B \setminus [\bigcap_n A'_m]) \cup ([\bigcap_n A'_m] \setminus B) \subset \bigcup_n (B \setminus A'_m) \cup (A'_m \setminus B) \in \mathcal{N},$$

because measurable subsets of elements in $\mathcal{N}$ are still in $\mathcal{N}$, $\mathcal{N}$ is closed under countable unions,

$$B \setminus A'_m \subset B \setminus A_m \subset B \triangle A_m \subset \mathcal{N}, \quad \text{and} \quad A'_1 \setminus B = \bigcap_{n=1}^{\infty} [A_n \setminus B] \subset \bigcap_{n=1}^{\infty} [A_n \triangle B] \in \mathcal{N}.$$

\[ \square \]

**Corollary 30.32.** 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 $\mathcal{N} \in \mathcal{N}$ there exists $\mathcal{N}' \in \mathcal{A} \cap \mathcal{N}$ such that $\mathcal{N}' \subset \mathcal{N}$. Then

$$\mathcal{A}^\mathcal{N} := \{ \mathcal{A} \cap \mathcal{N} : \mathcal{A} \in \mathcal{A} \text{ and } \mathcal{N} \in \mathcal{N}' \}. \tag{30.10}$$

**Proof.** Let $\mathcal{G}$ denote the right side of Eq. (30.10). It is clear that $\mathcal{G} \subset \mathcal{A}^\mathcal{N}$. Conversely if $\mathcal{B} \in \mathcal{A}^\mathcal{N}$, we know by Lemma 30.22 that there exists, $\mathcal{A} \in \mathcal{A}$ such that $\mathcal{N} := \mathcal{A} \triangle \mathcal{B} \in \mathcal{N}$. Since $C := \mathcal{A} \setminus \mathcal{B} \subset \mathcal{N}$, $C \in \mathcal{N}$ and so by assumption there exists $\mathcal{N}' \in \mathcal{A} \cap \mathcal{N}$ such that $\mathcal{C} \subset \mathcal{N}'$. Therefore, according to Eq. (30.7), we have

$$\mathcal{B} = [\mathcal{A} \setminus \mathcal{C}] \cup [\mathcal{B} \setminus \mathcal{A}] = [\mathcal{A} \setminus \mathcal{N}' \setminus \mathcal{C}] \cup [\mathcal{B} \setminus \mathcal{A}],$$

Since $\mathcal{A} \setminus \mathcal{N}' \in \mathcal{A}$ and $[\mathcal{A} \setminus \mathcal{N}' \setminus \mathcal{C}] \cup [\mathcal{B} \setminus \mathcal{A}] \in \mathcal{N}'$, it follows that $\mathcal{B} \in \mathcal{G}$. \[ \square \]

**Example 30.33.** Let $\nu$ be a probability measure on $\mathbb{R}$. As in Notation ??, let $P_{\nu} := \int_{\mathbb{R}} d\nu(x) \ P_{\nu}$ be the Wiener measure on $\Omega := C([0,\infty), \mathbb{R})$, $B_t : \Omega \rightarrow \mathbb{R}$ be the projection map, $B_t(\omega) = \omega(t)$, $B_t(\omega) = \sigma(B_i : s \leq t)$, and $\mathcal{N}'(\nu) := \{ N \in \mathcal{B}_t : P_{\nu}(N) = 0 \}$. Then by Corollary ??, $B_t = B_t \cup \mathcal{N}'(\nu)$. Hence if we let

$$\mathcal{N}(\nu) := \{ N \in \mathcal{B} : P_{\nu}(N) = 0 \}$$

and $\mathcal{N}(\nu) := \{ B \subset \Omega : B \subset N \text{ for some } N \in \mathcal{N}(\nu) \}$ then $B_t \cup \mathcal{N}(\nu) = B_t \cup \mathcal{N}(\nu) = B_t$, and $B_t \cup \mathcal{N}(\nu) = B_t \cup \mathcal{N}(\nu)$ for all $t \in \mathbb{R}_+$. This shows that the augmented Brownian filtration, $\{ \mathcal{B}_t \}_{t \geq 0}$, is already right continuous.

**Definition 30.34.** Recall from Proposition 6.66 if $(\Omega, \mathcal{B}, P)$ is a probability space and $\mathcal{N}$ is a null-like collection in $\mathcal{B}$ then the completion, $\tilde{P} := P^\mathcal{B} \setminus \mathcal{N}$, is a probability measure on $\mathcal{B} \setminus \mathcal{N}$ which extends $P$ so that $\tilde{P}(\Omega) = 0$ for all $\mathcal{A} \in \mathcal{N}^\mathcal{P}$.

Suppose that $(\mathcal{O}, \mathcal{B})$ 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 $A \subset \mathcal{B}$ with $A \subset \mathcal{N}$ 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, $B \in \mathcal{N}_i$ for all $i$ and hence $B \subset \mathcal{N}$. Moreover, it is clear that $\mathcal{N}$ is still closed under countable unions.

**Lemma 30.35 (Augmentation).** Let us now suppose that $\mathcal{A}$ is a sub-sigma-algebra of $\mathcal{B}$. Then the augmentation of $\mathcal{A}$ by $\mathcal{N}$,

$$\mathcal{A}^\mathcal{N} := \{ B \in \mathcal{B} : \exists A \in A \ni A \triangle \mathcal{B} \in \mathcal{N} \},$$

is a sub-sigma-algebra of $\mathcal{B}$. Moreover if $\mathcal{N} = \cap_i \mathcal{N}_i$ and $A_i = \mathcal{A}_i^{\mathcal{N}_i}$ is the augmentation of $\mathcal{A}$ by $\mathcal{N}_i$, then

$$\mathcal{A}^\mathcal{N} = \cap_i A_i.$$

**Proof.** To prove this, first observe that

$$A \triangle \mathcal{B} = (\mathcal{A} \triangle \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A}) = (\mathcal{A} \cap \mathcal{B}^c) \cup (\mathcal{B} \cap \mathcal{A}^c) = (\mathcal{B}^c \setminus \mathcal{A}^c) \cup (\mathcal{A}^c \setminus \mathcal{B}^c) = \mathcal{A}^c \triangle \mathcal{B}^c$$

from which it follows that $\mathcal{A}^\mathcal{N}$ is closed under countable union. 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 $j$. Letting $A = \bigcup_j A_j \in \mathcal{A}$ and $B = \bigcup_j B_j \in \mathcal{B}$, we have

$$B \ni A \triangle \mathcal{B} \in \bigcup_j [A_j \triangle B_j] \in \mathcal{N}$$

from which we conclude that $A \triangle \mathcal{B} \in \mathcal{N}$ and hence $B \in \mathcal{A}^\mathcal{N}$. This shows that $\mathcal{A}^\mathcal{N}$ is closed under unions and hence we have show $\mathcal{A} \subset \mathcal{A} \subset \mathcal{A}^\mathcal{N}$ 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

$$\mathcal{A}^\mathcal{N} \subset \cap_i A_i \ni \cap_i \mathcal{A}_i^{\mathcal{N}_i}.$$ 

For the converse inclusion, suppose that $B \subset \cap_i \mathcal{A}_i^{\mathcal{N}_i}$ in which case there exists $A_i \in \mathcal{A}$ such that $B \triangle A_i \in \mathcal{N}_i$ for all $i \in I$. \[ \square \]
Notation 30.36 Let $\bar{B}^P$ denote the completion of $B$. Let $B^P_t$ denote the augmentation of $B_t$ by the $P$–null subsets of $B$. We also let $\bar{B}^P_t$ denote the augmentation of $B_t$ by the $P$–null subsets of $\bar{B}^P_t$. 
Continuous time (sub)martingales

For this chapter, let \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \in \mathbb{R}_+}, P) \) be a filtered probability space as described in Chapter 30.

**Definition 31.1.** Given a filtered probability space, \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P) \), 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 31.2.** If \( \sigma \) and \( \tau \) are two \( \{\mathcal{B}_t\} \) – optional times, then \( \sigma \wedge \tau \) is as well. Indeed, \( \tau \in \mathbb{R}_+ \cup \{\infty\} \), 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 [21.40].

**Proposition 31.3 (Discrete optional sampling).** Suppose \( \{X_t\}_{t \in \mathbb{R}_+} \) is a submartingale on a filtered probability space, \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P) \), and \( \tau \) and \( \sigma \) are two \( \{\mathcal{B}_t\}_{t \geq 0} \) – stopping times with values in \( \mathbb{N} \), \( \{Y_k \}_{k \in \mathbb{N}} := \{\frac{X_{\tau(k)}}{X_{\tau(k)+1}} : k \in \mathbb{N}\} \) for some \( n \in \mathbb{N} \). If \( M := \sup_{\omega \in \Omega} \tau(\omega) < \infty \), then \( X_\tau \in L^1(\Omega, \mathcal{B}_\sigma, P) \), \( X_{\sigma \wedge \tau} \in L^1(\Omega, \mathcal{B}_{\sigma \wedge \tau}, P) \), and
\[
X_{\sigma \wedge \tau} \leq \mathbb{E}[X_\tau|\mathcal{B}_\sigma].
\]

**Proof.** For \( k \in \mathbb{N} \), let \( F_k := B_{k_2-n} \) and \( Y_k := X_{k_2-n} \). Then \( \{Y_k\}_{k=0}^\infty \) is a \( \{F_k\} \) – submartingale and \( 2^n \sigma, 2^n \tau \) are two \( \mathbb{N} \) – valued stopping times with \( 2^n \tau < 2^n M < \infty \). Therefore, we may apply the optional sampling Theorem 31.40 to find
\[
X_{\sigma \wedge \tau} = Y_{(2^n \sigma) \wedge (2^n \tau)} \leq \mathbb{E}[Y_{2^n \tau}|F_{2^n \sigma}] = \mathbb{E}[X_\tau|\mathcal{B}_\sigma].
\]

We have used \( F_{2^n \sigma} = B_{\sigma} \) (you prove) in the last equality. \( \blacksquare \)

**Lemma 31.4 (L^1 convergence I).** Suppose \( \{X_t\}_{t \in \mathbb{R}_+} \) is a submartingale on a filtered probability space, \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P) \). 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 \( F_n := \mathcal{B}_{t_n} \) for \( n \in \mathbb{N} \). Then \( \{(Y_n, F_n)\}_{n \in \mathbb{N}} \) is a backwards submartingale such that \( \inf_n \mathbb{E}Y_n \geq \mathbb{E}X_t \) and hence the result follows by Theorem 21.79. \( \blacksquare \)

**Lemma 31.5 (L^1 convergence II).** Suppose \( \{X_t\}_{t \in \mathbb{R}_+} \) is a submartingale on a filtered probability space, \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P) \), \( \tau \) is a bounded \( \{\mathcal{B}_t\} \) – optional time, and \( \{\tau_n\}_{n=1}^\infty \) is the sequence of approximate stopping times defined in Lemma 30.18. Then \( \lim_{n \to \infty} X_{\tau_n} \) 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 31.3 \( X_{\tau_n} \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_{\tau_m}] \) and \( X_0 \leq \mathbb{E}[X_{\tau_n}|\mathcal{B}_0] \). Hence, if we let \( Y_n := X_{\tau_n} \) and \( F_n := \mathcal{B}_{\tau_n} \) for \( n \in \mathbb{N} \). Then \( \{(Y_n, F_n)\}_{n \in \mathbb{N}} \) 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 21.79. \( \blacksquare \)

**Lemma 31.6 (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 \( Z \in L^1(P) \),
\[
\lim_{n \to \infty} \mathbb{E}[Z|\mathcal{B}_n] = \mathbb{E}[Z|\cap_{n=1}^\infty \mathcal{B}_n]
\]
where the above convergence is almost surely and in \( L^1(P) \).

**Proof.** This is a special case of Corollary 21.81 applied to the reverse martingale, \( M_m = \mathbb{E}[|Z|\mathcal{F}_m] \) where, \( m \in \mathbb{N} \), \( \mathcal{F}_m := \mathcal{B}_{-m} \). 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 31.7.** 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. \( \mathbb{E}[Z|\mathcal{B}_\tau] = \mathbb{E}[Z|\mathcal{B}_{\sigma \wedge \tau}] \).
Moreover, both results hold if \( \sigma \) and \( \tau \) are optional times provided every occurrence of the letter \( B \) is replaced by \( B^+ \).

**Proof.** 1. From Corollary \( 30.17 \) \( 1_{\sigma \leq t} E[Z|B_{\sigma}] \) is \( B_{\sigma \wedge \tau} \) measurable and therefore,

\[
1_{\sigma \leq t} E[Z|B_{\sigma}] = E[1_{\sigma \leq \tau} E[Z|B_{\sigma}]|B_{\sigma \wedge \tau}] = 1_{\sigma \leq \tau} E[E[Z|B_{\sigma}]|B_{\sigma \wedge \tau}] = 1_{\sigma \leq \tau} E[Z|B_{\sigma \wedge \tau}]
\]
as desired.

2. Writing

\[
Z = 1_{\sigma \leq \tau} Z + 1_{\tau < \sigma} Z
\]
we find, using item 1. that,

\[
E[Z|B_{\sigma}] = 1_{\sigma \leq \tau} E[Z|B_{\sigma}] + 1_{\tau < \sigma} E[Z|B_{\sigma}] = 1_{\sigma \leq \tau} E[Z|B_{\sigma \wedge \tau}] + 1_{\tau < \sigma} E[Z|B_{\sigma}].
\]

(31.2)

Another application of item 1. shows,

\[
E[1_{\tau < \sigma} E[Z|B_{\sigma}]|B_{\tau}] = 1_{\tau < \sigma} E[E[Z|B_{\sigma}]|B_{\tau}] = 1_{\tau < \sigma} E[Z|B_{\tau \wedge \sigma}].
\]

Using this equation and the fact that \( 1_{\sigma \leq \tau} E[Z|B_{\sigma \wedge \tau}] \) is \( B_{\sigma \wedge \tau} \) measurable, we may condition Eq. (31.2) on \( B_{\tau \wedge \sigma} \) to find

\[
E[E[Z|B_{\sigma}]|B_{\tau}] = 1_{\sigma \leq \tau} E[Z|B_{\sigma \wedge \tau}] + 1_{\tau < \sigma} E[Z|B_{\tau \wedge \sigma}] = E[Z|B_{\tau \wedge \sigma}].
\]

**Lemma 31.8.** Suppose \( \sigma \) is an optional time and \( \{\sigma_m\}_{m=1}^{\infty} \) are stopping times such that \( \sigma_m \downarrow \sigma \) as \( m \to \infty \) and \( \sigma < \sigma_m \) on \( \{\sigma < \infty\} \) for all \( m \in \mathbb{N} \). Then \( B_{\sigma_m} \downarrow B_{\sigma}^+ \) as \( m \to \infty \), i.e. \( B_{\sigma_m} \) is decreasing in \( m \) and

\[
B_{\sigma}^+ = \bigcap_{m=1}^{\infty} B_{\sigma_m}.
\]

(31.3)

**Proof.** If \( A \in B_{\sigma}^+ \), then \( A \in B_{\infty} \) and for all \( t \in \mathbb{R}_+ \) and \( m \in \mathbb{N} \) we have

\[
A \cap \{\sigma_m \leq t\} = A \cap \{\sigma < t\} \cap \{\sigma_m \leq t\} \in B_t.
\]

This shows \( A \in \bigcap_{m=1}^{\infty} B_{\sigma_m} \). For the converse, observe that

\[
\{\sigma < t\} = \bigcup_{m=1}^{\infty} \{\sigma_m \leq t\} \quad \forall \ t \in \mathbb{R}_+.
\]

Therefore if \( A \in \bigcap_{m=1}^{\infty} B_{\sigma_m} \) then \( A \in B_{\infty} \) and

\[
A \cap \{\sigma < t\} = \bigcup_{m=1}^{\infty} \{A \cap \{\sigma_m \leq t\}\} \in B_t \quad \forall \ t \in \mathbb{R}_+.
\]

**Theorem 31.9 (Continuous time optional sampling theorem).** Let \( \{X_t\}_{t \geq 0} \) be a right continuous \( \{B_t\} \) (or \( \{B_t^+\} \)) - submartingale and \( \sigma \) and \( \tau \) be two \( \{B_t\} \) (or \( \{B_t^+\} \)) - optional (or stopping) times such that \( M := \sup_{\omega \in \Omega} \tau(\omega) < \infty \)

Then \( X_t \in L^1(\Omega, B_t^+, P) \), \( X_{\sigma \wedge \tau} \in L^1(\Omega, B_{\sigma \wedge \tau}^+, P) \)

\[
X_{\sigma \wedge \tau} \leq E[X_t|B_{\sigma}^+].
\]

(31.4)

**Proof.** Let \( \{\tau_m\}_{m=1}^{\infty} \) and \( \{\tau_n\}_{n=1}^{\infty} \) be the sequences of approximate times for \( \sigma \) and \( \tau \) respectively defined Lemma 31.18 i.e.

\[
\tau_n := 1_{\tau = \infty} + \frac{k}{2^n} 1_{\frac{k-1}{2^n} \leq \tau < \frac{k}{2^n}}.
\]

By the discrete optional sampling Proposition 31.3 we know that

\[
X_{\sigma \wedge \tau_n} \leq E[X_{\tau_n}|B_{\sigma_m}] \text{ a.s.}
\]

(31.5)

Since \( X_t \) is right continuous, \( X_{\tau_n}(\omega) \to X_\tau(\omega) \) for all \( \omega \in \Omega \) which combined with Lemma 31.5 implies \( X_{\tau_n} \to X_\tau \) in \( L^1(P) \) and in particular \( X_\tau \in L^1(P) \)

Similarly, \( X_{\sigma_m \wedge \tau_n} \to X_{\sigma \wedge \tau} \) 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. (31.5) to find

\[
X_{\sigma \wedge \tau} \leq \lim_{m \to \infty} E[X_{\tau_n}|B_{\sigma_m}] = E[X_{\tau}|B_{\sigma}^+] \quad \text{a.s.}
\]

(31.6)

From the right continuity of \( \{X_t\} \) and making use of Lemma 31.6 (or Corollary 21.81) and Lemma 31.8 we may let \( m \to \infty \) in Eq. (31.6) to find

\[
X_{\sigma \wedge \tau} \leq \lim_{m \to \infty} E[X_{\tau_n}|B_{\sigma_m}] = E[X_{\tau}|B_{\sigma}^+] \quad \text{a.s.}
\]

which is Eq. (31.4).

**Corollary 31.10 (Optional stopping).** Let \( \{X_t\}_{t \geq 0} \) be a right continuous \( \{B_t\} \) (or \( \{B_t^+\} \)) - submartingale and \( \tau \) be any \( \{B_t\} \) - optional (or stopping) time. Then the stopped process, \( X_t^\tau := X_{\sigma \wedge t} \) is a right continuous \( \{B_t^\tau\} \) - submartingale.

**Proof.** Let \( 0 \leq s \leq t < \infty \) and apply Theorem 31.9 with to the two stopping times, \( \sigma \wedge s \) and \( \sigma \wedge t \) to find

\[
X_s^\tau = X_{\sigma \wedge s} \leq E[X_{\sigma \wedge t}|B_{\sigma \wedge \tau}^+] = E[X_t^\tau|B_{\sigma \wedge \tau}^+] \quad \text{a.s.}
\]

From Proposition 31.7

\[
E[X_t^\tau|B_{\sigma \wedge \tau}^+] = E[X_{\sigma \wedge t}|B_{\sigma \wedge \tau}^+] = E[E[X_{\sigma \wedge t}|B_{\sigma \wedge \tau}^+]|B_t^+] = E[X_{\sigma \wedge t}|B_t^+]
\]

and therefore, we have shown \( X_s^\tau \leq E[X_t^\tau|B_t^+] \). Since \( X_s^\tau = X_{\sigma \wedge s} \) is \( B_{\sigma \wedge s}^+ \) measurable and \( B_{\sigma \wedge \tau}^+ \subset B_t^+ \), it follows that \( X_s^\tau \) is \( B_t^+ \) measurable.

\[\text{This will be shown below in Theorem 31.28 that the boundedness restriction on } \tau \text{ may be replaced by the assumption that } \{X_t^\tau\}_{t \geq 0} \text{ is uniformly integrable.}\]
31.1 Submartingale Inequalities

Let \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, P)\) 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 \(\sup_{D \ni t \leq T} X_t = \inf_{D \ni t \leq T} X_t\), and \(\sup_{s \in [0,T]} |X_s|\) by \(\sup_{t \leq T} X_t\) and \(\inf_{t \leq T} X_t\), 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
\[
\sup_{t \leq T} X_t = \sup_{D \ni t \leq T} X_t, \quad \inf_{t \leq T} X_t = \inf_{D \ni t \leq T} X_t \quad \text{and} \quad \sup_{s \in [0,T]} |X_s| = \sup_{s \in D \cap [0,T]} |X_s|.
\]

Our immediate goal is to generalize the submartingale inequalities of Section 21.5 to this context.

Proposition 31.11 (Maximal Inequalities of Bernstein and Lévy). With \(\mathbb{T} = \mathbb{D}\) for \(\mathbb{T} = \mathbb{R}_+\), for any \(a \geq 0\) and \(T \in \mathbb{T}\), we have,
\[
\begin{align*}
ap \left( \sup_{t \leq T} X_t \geq a \right) & \leq E \left[ X_T : \sup_{t \leq T} X_t \geq a \right] \leq E \left[ X_T^+ \right], \quad (31.8) \\
ap \left( \inf_{t \leq T} X_t \leq -a \right) & \leq E \left[ X_T : \inf_{t \leq T} X_t \leq -a \right] - E [X_0] \leq E \left[ X_T^+ \right] - E [X_0], \quad (31.9) \\
\text{and} \quad aP (X_T^+ \geq a) & \leq 2E \left[ X_T^+ \right] - E [X_0]. \quad (31.10)
\end{align*}
\]

and
\[
P (M_T^+ \geq a) \leq \frac{1}{a} E [M_T^+ : M_T^+ \geq a] \leq \frac{1}{a} E [M_T] \quad (31.12)
\]

Proof. First assume \(\mathbb{T} = \mathbb{D}\). For each \(k \in \mathbb{N}\) let
\[
A_k = \{0 = t_0 < t_1 < \cdots < t_m = T \} \subset \mathbb{D} \cap [0, T]
\]
be a finite subset of \(\mathbb{D} \cap [0, T]\) containing \(\{0, T\}\) such that \(A_k \uparrow \mathbb{D} \cap [0, T]\). Noting that \(\{X_{t_n}\}_{n=0}^m\) is a discrete \((\Omega, \mathcal{F}, \{\mathcal{F}_{t_n}\}_{n=0}^m, P)\) submartingale, Proposition 21.4.3 implies all of the inequalities in Eqs. (31.8) - (31.11) holds provided we replace \(\sup_{t \leq T} X_t\) by \(\max_{t_n \in A_k} X_{t_n}\), \(\inf_{t \leq T} X_t\) by \(\min_{t_n \in A_k} X_{t_n}\), and \(X_T^+\) by \(\max_{t_n \in A_k} |X_{t_n}|\). Since \(\max_{t_n \in A_k} X_{t_n} \uparrow \sup_{t \leq T} X_t\), \(\min_{t_n \in A_k} X_{t_n} \downarrow \inf_{t \leq T} X_t\), and \(\max_{t_n \in A_k} |X_{t_n}| \uparrow X_T^+\), we may use the MCT and the DCT to pass to the limit (\(k \to \infty\)) in order to conclude Eqs. (31.8) - (31.11) are valid as stated. Equation (31.12) follows from Eq. (31.8) 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. (31.7), we see that Eqs. (31.8) - (31.12) remain valid for \(\mathbb{T} = \mathbb{R}_+\) by what we have just proved in the case \(\mathbb{T} = \mathbb{D} \cup \{T\}\).

Proposition 31.12 (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 \(T \in \mathbb{T}\),
\[
E X_T^p \leq \left( \frac{p}{p-1} \right)^p E X_T^p.
\]

Proof. Using the notation in the proof of Proposition 31.11 it follows from Corollary 21.47 that
\[
E \left[ \max_{t \in A_k} |X_t|^p \right] \leq \left( \frac{p}{p-1} \right)^p E X_T^p.
\]

Using the MCT, we may let \(k \uparrow \infty\) in this equation to arrive at Eq. (31.13) 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 31.11.

Lemma 31.13. 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
\[
|F(t) - F(s)| \leq |F(t) - F_n(t)| + |F_n(t) - F_n(s)| + |F_n(s) - F(s)| \\
\leq 2\varepsilon_n + |F_n(t) - F_n(s)|.
\]

Therefore we have
\[
\limsup_{s, t \uparrow a} |F(t) - F(s)| \leq 2\varepsilon_n \to 0 \text{ as } n \to \infty
\]

which shows that \(\xi := \lim_{t \downarrow a} F(t)\) exists. Similarly, for any \(t > a\),
\[
|\xi - \xi_n| \leq |\xi - F(t)| + |F(t) - F_n(t)| + |F_n(t) - \xi_n| \\
\leq |\xi - F(t)| + |F_n(t) - \xi_n| + \varepsilon_n
\]

and hence by passing to the limit as \(t \downarrow a\) in the previous inequality we have
\[
|\xi - \xi_n| \leq \varepsilon_n \to 0 \text{ as } n \to \infty.
\]
Corollary 31.14. Suppose that \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P)\) is a filtered probability space such that \(\mathcal{B}_t\) contains all \(P\)–null subsets of \(\mathcal{B}\) for all \(t \in \mathbb{R}_+\). For any \(T \in \mathbb{R}_+\), let \(\mathcal{M}_T\) denote the collection of (right) continuous \(L^2\)–martingales, 
\[ M := \{M_t\}_{t \leq T} \text{ equipped with the inner product,} \]
\[ (M, N)_T := \mathbb{E}[M_T N_T]. \]
(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, \((\mathcal{M}_T, (\cdot, \cdot)_T)\), is a Hilbert space and the map, \(U: \mathcal{M}_T \rightarrow \mathbb{L}^2(\Omega, \mathcal{B}_T, P)\) defined by \(U M := M_T\), 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 \(M = 0 \in \mathcal{M}_T\). This shows that \(U\) is injective and by definition \(U\) is an isometry and \((\cdot, \cdot)_T\) is an inner product on \(\mathcal{M}_T\). To finish the proof, we need only show \(H := \operatorname{Ran}(U)\) is a closed subspace of \(\mathbb{L}^2(\Omega, \mathcal{B}_T, P)\) or equivalently that \(\mathcal{M}_T\) is complete.

Suppose that \(\{M^n\}_{n=1}^\infty\) is a Cauchy sequence in \(\mathcal{M}_T\), then by Doob’s inequality (Proposition 31.12) and Hölder’s inequality, we have
\[
\mathbb{E} [(M^n - M^m)^2_T] \leq \mathbb{E} [2[M^n - M^m]_T^2] \leq 4\mathbb{E} |M^n_T - M^m_T|^2 = 2\|M^n - M^m\|_T \rightarrow 0 \text{ as } m,n \rightarrow \infty.
\]
By passing to a subsequence if necessary, we may assume
\[
\sum_{n=1}^\infty \mathbb{E} [(M^{n+1} - M^n)^*_T] \leq 2 \sum_{n=1}^\infty \|M^{n+1} - M^n\|_T < \infty
\]
from which it follows that
\[
\mathbb{E} \left[ \sum_{n=1}^\infty (M^{n+1} - M^n)^*_T \right] = \sum_{n=1}^\infty \mathbb{E} [(M^{n+1} - M^n)^*_T] < \infty.
\]
So if we let
\[
\Omega_0 := \left\{ \sum_{n=1}^\infty (M^{n+1} - M^n)_T < \infty \right\},
\]
then \(P(\Omega_0) = 1\). Hence if \(m \leq l\), the triangle inequality implies
\[
(M^l - M^m)^*_T \leq \sum_{n=m}^{l-1} (M^{n+1} - M^n)^*_T \rightarrow 0 \text{ on } \Omega_0 \text{ as } m,l \rightarrow \infty,
\]
which shows that \(\{M^n(\omega)\}_{n=1}^\infty\) is a uniformly Cauchy sequence and hence uniformly convergent for all \(\omega \in \Omega_0\). Therefore by Lemma 31.13 \(t \rightarrow M_t(\omega)\) is (right) continuous for all \(\omega \in \Omega_0\). We complete the definition of \(M\) by setting \(M(\omega) = 0\) for \(\omega \notin \Omega_0\). Since \(\mathcal{B}_t\) contains all of the null subset in \(\mathcal{B}\), it is easy to see that \(M\) is a \(\mathcal{B}_t\)–adapted process. Moreover, by Fatou’s lemma, we have
\[
\mathbb{E} \left[ (M^n - M^m)^*_T \right] \leq \liminf_{n \rightarrow \infty} \mathbb{E} \left[ (M^n - M^m)^*_T \right] \leq \liminf_{n \rightarrow \infty} \mathbb{E} \left[ (M^n - M^m)^*_T \right] \rightarrow 0 \text{ as } m \rightarrow \infty.
\]
In particular \(M^m_t \rightarrow M_t\) in \(\mathbb{L}^2(P)\) for all \(t \leq T\) from which follows that \(M\) is still an \(L^2\)–martingale. As \(M\) is (right) continuous, \(M \in \mathcal{M}_T\) and
\[
\|M - M^n\|_T = \|M_T - M^n_T\|_{\mathbb{L}^2(P)} \rightarrow 0 \text{ as } n \rightarrow \infty.
\]

31.2 Regularizing a submartingale

Lemma 31.15. Suppose that \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P)\) 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 \(\{\bar{\mathcal{B}}_t\}_{t \geq 0}\)–

submartingale. Moreover, we may first replace \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P)\) by its completion, \((\Omega, \mathcal{B}, \bar{P})\) (see Proposition 6.66), then \(\{X_t\}_{t \geq 0}\) is still a submartingale relative to the filtration \(\bar{\mathcal{B}} := \mathcal{B} \vee N\) where
\[
N := \{ B \in \bar{\mathcal{B}} : \bar{P}(B) = 0 \}.
\]

Proof. It suffices to prove the second assertion. By the augmentation Lemma 30.22 we know that \(B \in \mathcal{B}_s := \mathcal{B}_s \vee N\) iff there exists \(A \in \mathcal{B}_s\) such that \(B \triangle A \in N\). Then for any \(t > s\) we have
\[
\mathbb{E}_{\bar{P}} [X_t - X_s : B] = \mathbb{E}_{\bar{P}} [X_t - X_s : A] = \mathbb{E}_{\bar{P}} [X_t - X_s : A] \geq 0.
\]

Proposition 31.16. Suppose that \(\{X_t\}_{t \in \mathbb{R}_+}\) is an \(\{\mathcal{B}_t\}\)–submartingale such that \(t \rightarrow X_t\) 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 \(\{\mathcal{B}_t\}\)–submartingale.
Lemma 30.23. We use Lemma 30.23 to guarantee that the rest of this section.

Corollary 31.17. Suppose \((\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, \mathbb{P})\) is a filtered probability space and \(\{X_t\}_{t \in \mathbb{R}_+}\) is an \(\{\mathcal{B}_t\}\) - submartingale such that \(X_t \xrightarrow{\mathbb{P}} X_s\) as \(t \downarrow s\) for all \(s \in \mathbb{R}_+\). Let \((\Omega, \mathcal{B}, \mathbb{P})\) denote the completion of \((\Omega, \mathcal{B}_t, \mathbb{P})\), \(\mathcal{N}\) and \(\mathcal{N}_t\) the \(P\) and \(\mathbb{P}\) null sets respectively, then \((\Omega, \mathcal{B}, \{\mathcal{B}_t \vee \mathcal{N}_t\}_{t \geq 0}, \mathbb{P})\) satisfies the weak usual hypothesis (see Definition 30.21), \((\Omega, \mathcal{B}, \{\mathcal{B}_t \vee \mathcal{N}_t\}_{t \geq 0}, \mathbb{P})\) satisfies the usual hypothesis, and \(\{X_t\}_{t \geq 0}\) is a submartingale relative to each of these filtrations.

Proof. This follows directly from Proposition 31.16 Lemma 31.15 and Lemma 30.23. We use Lemma 30.23 to guarantee that \(\{\mathcal{B}_t \vee \mathcal{N}\}_{t \geq 0}\) and \(\{\mathcal{B}_t \vee \mathcal{N}_t\}_{t \geq 0}\) are right continuous.

In all of the examples of submartingales appearing in this book, the hypothesis and hence the conclusions of Proposition 31.16 will apply. For this reason there is typically no harm in assuming that our filtration is right continuous. By Corollary 31.17 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 31.1 (Continuous version of Example 21.7). 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 \(t \geq 0\) and \(x \in \Omega\),

\[
M_t^\varepsilon (x) := e^t \left( \varepsilon (t) 1_{0 < x < e^{-t}} + (1 - \varepsilon (t)) 1_{0 < x < e^{-t}} \right) = e^t \left( 1_{0 < x < e^{-t}} + \varepsilon (t) 1_{x = e^{-t}} \right).
\]

Further let \(\mathcal{B}_t^\varepsilon := \sigma (M_s^\varepsilon : s \leq t)\) for all \(t \geq 0\) and for \(a \in (0, 1]\) let

\[
\mathcal{F}_{[0,a]} := \{ [0, a] \cup A : A \in \mathcal{B}_{[a,1]} \} \cup \mathcal{B}_{[a,1]}
\]

and

\[
\mathcal{F}_{(0,a)} := \{ (0, a) \cup A : A \in \mathcal{B}_{[a,1]} \} \cup \mathcal{B}_{[a,1]}.
\]

Show:

1. \(\mathcal{F}_{[0,a]} \subset \mathcal{F}_{[0,1]}\) and \(\mathcal{F}_{(0,a)} \subset \mathcal{F}_{(0,1)}\) are sub – sigma – algebra of \(\mathcal{B}\) such that \(\mathcal{F}_{[0,a]} \subset \mathcal{F}_{(0,a)}\) 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]e^t \left( \varepsilon (t) 1_{0 < x < 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)
\]

3. \(M_t^\varepsilon = M_t^0\) for all \(t \geq 0\) 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 \(x \in \Omega\).

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, \(\mathcal{B}_t^\varepsilon = \mathcal{B}_{[0,e^{-t}]}^0 = \mathcal{F}_{[0,e^{-t}]}\) for all \(t \geq 0\).

6. \(M_t^\varepsilon\) is a \(\{\mathcal{B}_t^\varepsilon\}_{t \geq 0}\) – martingale and in fact it is a \(\{\mathcal{B}_t^\varepsilon\}_{t \geq 0}\) – 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

\[ N := \{ x : M^t_x (x) \neq M^t_{x^+} (x) \text{ for some } t \geq 0 \} . \]

Show \( N \) is measurable if \( \varepsilon \) is measurable. Also observe that if \( \varepsilon = 1 \), then \( P (N) = 1 \) and hence \( M^t_{x^+} \) and \( M^t_x \) are certainly not indistinguishable, see Definition 28.2.

9. Show \( \{ M^t_{x^+} \}_{t \geq 0} \) is not uniformly integrable.

10. Let \( Z \in L^1 (\Omega, B, P) \) find a version, \( N^*_t \), of \( \mathbb{E} [Z | \mathcal{B}^*_t] \). Verify that for any sequence, \( \{ t_n \}_{n=1}^\infty \subset [1, \infty) \), that \( N^*_n \rightarrow Z \) almost surely and in \( L^1 (P) \) as \( n \rightarrow \infty \).

**Definition 31.18 (Upcrossings).** Let \( \{ x_t \}_{t \in T} \) be a real valued function which is right continuous if \( T = \mathbb{R}_+ \). Given \( -\infty < a < b < \infty \), \( T \in \mathbb{T} \), and a finite subset, \( F \), of \( [0, T] \cap T \), let \( U^F_x (a, b) \) denote the number of upcrossings of \( \{ x_t \}_{t \in F} \) across \( [a, b] \), see Section 21.6. Also let

\[ U^F_x (a, b) := \sup \{ U^F_x (a, b) : F \subset \mathbb{T} \cap [0, T] \} \] (31.15)

be the number of upcrossings of \( \{ x_t \}_{t \in \mathbb{T} \cap [0, T]} \) across \( [a, b] \).

**Lemma 31.19.** 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 \( F_n \uparrow \mathbb{T} \cap [0, T] \), then

\[ U^F_x (a, b) := \lim_{n \rightarrow \infty} U^F_{F_n} (a, b) . \] (31.16)

In particular, \( U^F_x (a, b) \) is a \( \mathcal{B}_T \) measurable random variable when \( T = \mathbb{D} \).

**Proof.** It is clear that \( U^F_{F_n} (a, b) \leq U^F_x (a, b) \) for all \( n \) and \( U^F_{F_n} (a, b) \) is increasing with \( n \) and therefore the limit in Eq. (31.16) exists and satisfies, \( \lim_{n \rightarrow \infty} U^F_{F_n} (a, b) \leq U^F_x (a, b) \). Moreover, for any \( F \subset \mathbb{D} \cap [0, T] \) we may find an \( n \in \mathbb{N} \) sufficiently large so that \( F \subset F_n \). For this \( n \) we will have

\[ U^F_{F_n} (a, b) \leq U^F_x (a, b) \leq \lim_{n \rightarrow \infty} U^F_{F_n} (a, b) . \]

Taking supremum over all \( F \subset \mathbb{D} \cap [0, T] \) in this estimate then shows \( U^F_x (a, b) \leq \lim_{n \rightarrow \infty} U^F_{F_n} (a, b) \).

**Remark 31.20.** It is easy to see that if \( T = \mathbb{R}_+ \), \( x_t \) is right continuous, and \( a < \alpha < \beta < b \), then

\[ U^F_x (a, b) \leq \sup \{ U^F_x (a, \beta) : F \subset \mathbb{D} \cap [0, T] \} . \]

**Lemma 31.21.** Let \( T \in \mathbb{R}_+ \) and \( \{ x_t \}_{t \in \mathbb{D}} \) be a real valued function such that \( U^F_x (a, b) < \infty \) for all \( -\infty < a < b < \infty \) with \( a, b \in \mathbb{Q} \). Then

\[ x_{t-} := \lim_{\mathbb{D} \ni \delta \uparrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in (0, T] \] (31.17)
\[ x_{t+} := \lim_{\mathbb{D} \ni \delta \uparrow t} x_s \text{ exists in } \bar{\mathbb{R}} \text{ for } t \in [0, T) . \] (31.18)

Moreover, if we let \( U^\infty_x (a, b) = \lim_{t \uparrow \infty} U^F_x (a, b) \) 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. (31.17) 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

\[ \liminf_{\mathbb{D} \ni \delta \uparrow t} x_s < a < b < \limsup_{\mathbb{D} \ni \delta \uparrow t} x_s . \]

From the definition of the liminf and the limsup, 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

\[ \infty = U^\infty_x (a, b) \leq U^F_x (a, b) . \]

**Lemma 31.22.** Suppose that \( \mathbb{T} = \mathbb{D} \), \( S \) is a metric space, and \( \{ x_t \in S \}_{t \in \mathbb{D}} \).

1. If for all \( t \in \mathbb{R}_+ \),

\[ x_t^+ := x_{t+} = \lim_{\mathbb{D} \ni \delta \uparrow t} x_s \text{ exists in } S , \]

then \( \mathbb{R}_+ \ni t \rightarrow x_t^+ \in S \) is right continuous.

2. If we further assume that

\[ x_{t-} := \lim_{\mathbb{D} \ni \delta \uparrow t} x_s \text{ exists in } S \]

for all \( t > 0 \), then \( \lim_{t \uparrow \infty} x_t^+ = x_{t-} \) for all \( t > 0 \).

3. Moreover, if \( \lim_{\mathbb{D} \ni \delta \uparrow \infty} x_t \in S \) then again \( \lim_{t \uparrow \infty} x_t^+ = \lim_{\mathbb{D} \ni \delta \uparrow \infty} x_t \).

**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

\[ \rho (x_t^+, x_{t+}) = \lim_{\mathbb{D} \ni \delta \uparrow t} \rho (x_t^+, x_s) \leq \varepsilon \]

from which it follows that \( x_{t+} \rightarrow x_t^+ \) as \( \tau \downarrow t \).

2. Now suppose \( t > 0 \) such that \( x_{t-} \in S \). 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,

\[ \rho (x_{t-}, x_{t+}) = \lim_{s \downarrow t} \rho (x_{t-}, x_s) \leq \varepsilon \]
from which it follows that $x_{\tau^+} \leq x_{\tau-}$ as $\tau \uparrow t$.

3. Now suppose $x_{\infty} := \lim_{\tau \uparrow t} x_{\tau}$ exists in $S$. Then for every $\varepsilon > 0$, there exists $M = M(\varepsilon) < \infty$ such that $\rho(x_{\infty}, x_s) \leq \varepsilon$ if $s \in D \cap (M, \infty)$. Hence if $t \in (M, \infty)$ we have

$$\rho(x_{\infty}, x_t) = \lim_{\tau \uparrow t} \rho(x_{\infty}, x_s) \leq \varepsilon$$

from which we conclude that, $\lim_{t \uparrow t} x_{t+}$ exists in $S$ and is equal to $x_{\infty}$. □

Theorem 31.23 (Doob’s upcrossing inequality). Let $\{X_t\}_{t \in D}$ be a submartingale and $-\infty < a < b < \infty$. Then for all $T \in D$,

$$\mathbb{E}[U_T^{-}(a, b)] \leq \frac{1}{b - a} \left[ \mathbb{E}(X_T - a)^+ - \mathbb{E}(X_0 - a)_+ \right]. \quad (31.19)$$

Proof. Let $\{F_n\}_{n=1}^\infty$ be a sequence as in Lemma 31.19 and assume without loss of generality that $0, T \in F_n$ for all $n$. It then follows from Theorem 21.53 that

$$\mathbb{E}[U_{F_n}^{-}(a, b)] \leq \frac{1}{b - a} \left[ \mathbb{E}(X_T - a)^+ - \mathbb{E}(X_0 - a)_+ \right] \quad \forall n \in \mathbb{N}.$$

By letting $n \uparrow \infty$, Eq. (31.19) follows from this inequality, Lemma 31.19 and the MCT. □

Theorem 31.24. Let $\{X_t\}_{t \in D}$ be a submartingale,

$$\Omega_0 := \bigcap_{T \in D} \left( \sup_{D \ni t \leq T} |X_t| < \infty \right) \cap \left[ \bigcap_{n} \{U_n^X(a, b) < \infty : a < b \text{ with } a, b \in \mathbb{Q}\} \right),$$

for all $t \in \mathbb{R}_+$,

$$Y_t := \lim_{D \ni s \uparrow t} X_s \quad \text{and} \quad \tilde{X}_t := Y_t \cdot 1_{|Y_t| < \infty}. \quad (31.20)$$

Then;

1. $P(\Omega_0) = 1$.
2. on $\Omega_0$, $\sup_{D \ni t \leq T} |X_t| < \infty$ 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 $\{s_n\}_{n=1}^\infty \subset D \cap (t, \infty)$ such that $s_n \downarrow t$, then $X_{s_n} \rightarrow X_t^+$ in $L^1(P)$ as $n \rightarrow \infty$.
5. The process $\{\tilde{X}_t\}_{t \in \mathbb{R}_+}$ is a $\{\mathcal{B}^+_t\}_{t \geq 0}$ submartingale such that $t \rightarrow \tilde{X}_t$ is right continuous and has left limits on $\Omega_0$.
6. $X_t \leq \mathbb{E}[X_t | \mathcal{B}_t] \quad \text{a.s. for all } t \in D \text{ with equality at some } t \in D \text{ iff} \lim_{D \ni s \uparrow t} \mathbb{E}X_s = \mathbb{E}X_t$.

7. If $X_s^L \rightarrow X_t$ as $D \ni s \downarrow t$ at some $t \in D$ then $\tilde{X}_t = X_t$ a.s.
8. If $C := \sup_{D \ni t \leq T} \mathbb{E}|X_t| < \infty$ (or equivalently $\sup_{D \ni t \leq T} \mathbb{E}X_t^+ < \infty$), then $X_{\infty} := \lim_{D \ni t \uparrow T} X_t = \lim_{D \ni t \uparrow T} X_t$ exists in $\mathbb{R}$ a.s. and $\mathbb{E}|X_{\infty}| < C < \infty$.

Note: if $\{X_t^+\}_{t \in D}$ is uniformly integrable then $\sup_{D \ni t \leq T} \mathbb{E}|X_t^+| < \infty$.
9. If $\{X_t^+\}_{t \in D}$ is uniformly integrable iff there exists $X_{\infty} \in L^1(\Omega, \mathcal{B}, P)$ such that $\{X_t\}_{t \in D \cup \{\infty\}}$ is a submartingale. In other words, $\{X_t^+\}_{t \in D}$ is uniformly integrable iff here exists $X_{\infty} \in L^1(\Omega, \mathcal{B}, P)$ such that $X_t \leq \mathbb{E}[X_{\infty} | \mathcal{B}_t]$ a.s. for all $t \in D$.

Proof. 1. – 3. The fact that $P(\Omega_0) = 1$ follows from Doob’s upcrossing inequality and the maximal inequality in Eq. (31.11). The assertions in items 2. and 3. are now a consequence of the definition of $\Omega_0$ and Lemmas 31.21 and 31.22.

4. Let $Y_n := X_{s_n}$ and $F_n := B_{s_n}$ for $n \in \mathbb{N}$. Then $\{(Y_n, F_n)\}_{n \in \mathbb{N}}$ is a backwards submartingale such that $\mathbb{E}Y_n \geq \mathbb{E}X_t$ and hence by Theorem 21.79 $Y_n = X_{s_n} \rightarrow X_{t+}$ in $L^1(P)$ as $n \rightarrow \infty$.

5. Since $\tilde{X}_t = X_{t+}$ on $\Omega_0$ and $X_{t+}$ is right continuous with left hand limits, $\tilde{X}_t$ has these properties on $\Omega_0$ as well. Now let $0 \leq s < t < \infty$, $\{s_n\}, \{t_n\} \subset D$, 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 $X$,

$$\mathbb{E}[\tilde{X}_t - \tilde{X}_s : A] = \mathbb{E}[X_{t+} - X_{s+} : A] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n} - X_{s_n} : A] \geq 0$$

for all and $A \in B_{s+}$.

6. Let $A \in B_t$ and $\{t_n\} \subset D$ with $t_n \downarrow t \in D$, then

$$\mathbb{E}[\tilde{X}_t : A] = \mathbb{E}[X_{t+} : A] \geq \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n} : A].$$

Since $A \in B_t$ is arbitrary it follows that $X_t \leq \mathbb{E}[\tilde{X}_t | \mathcal{B}_t]$ a.s. If equality holds, then, taking $A = \Omega$ above, we find

$$\mathbb{E}X_t = \mathbb{E}\tilde{X}_t = \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n}].$$

Since $\{t_n\} \subset D$ with $t_n \downarrow t$ was arbitrary, we may conclude that $\lim_{D \ni s \uparrow t} \mathbb{E}X_s = \mathbb{E}X_t$. Conversely if $\lim_{D \ni s \uparrow t} \mathbb{E}X_s = \mathbb{E}X_t$, then along any sequence, $\{s_n\} \subset D$ with $s_n \downarrow s$, we have

$$\mathbb{E}X_t = \lim_{n \rightarrow \infty} \mathbb{E}X_{s_n} = \mathbb{E} \lim_{n \rightarrow \infty} X_{s_n} = \mathbb{E}X_t = \mathbb{E}[\tilde{X}_t | \mathcal{B}_t].$$

As $X_t \leq \mathbb{E}[\tilde{X}_t | \mathcal{B}_t]$ a.s. this identity implies $X_t = \mathbb{E}[\tilde{X}_t | \mathcal{B}_t]$ a.s. □

3 For example, this will hold if $\lim_{D \ni s \uparrow t} \mathbb{E}|X_t - X_s| = 0$. 

Page: 491  
job: prob  
macro: svmobonob.cls  
date/time: 2-Nov-2018/7:31
7. Let \( t_n \in \mathbb{D} \) such that \( t_n \downarrow t \), then as we have already seen \( X_{t_n} \to \tilde{X}_t \) in \( L^1(P) \). However by assumption, \( X_{t_n} \overset{p}{\rightarrow} X_t \), and therefore we must have \( \tilde{X}_t = X_t \) 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 21.56.

8. The proof here mimics closely the discrete version given in Corollary 21.56.

For any \(-\infty < a < b < \infty\), Doob's upcrossing inequality (Theorem 31.23) and the MCT implies,

\[
\mathbb{E} [U_\infty^{X} (a, b)] = \lim_{\mathbb{D} \ni t \to \infty} \mathbb{E} [U_T^{X} (a, b)] 
\leq \frac{1}{b-a} \left[ \sup_{T \in \mathbb{D}} \mathbb{E} (X_T - a) + \mathbb{E} (X_0 - a) \right] < \infty
\]

where

\[
U_T^{X} (a, b) = \lim_{\mathbb{D} \ni t \to \infty} U_T^{X} (a, b)
\]

is the total number of upcrossings of \( X \) across \([a, b]\). In particular it follows that

\[
\tilde{\Omega}_0 := \cap \{ U_T^{X} (a, b) < \infty : a, b \in \mathbb{Q} \text{ with } a < b \}
\]

has probability one. Hence by Lemma 31.21 for \( \omega \in \Omega_0 \) we have \( X_\infty (\omega) := \lim_{\mathbb{D} \ni t \to \infty} X_t (\omega) \) exists in \( \mathbb{R} \). By Fatou's lemma with \( \mathbb{D} \ni t_n \uparrow t \), it follows that

\[
\mathbb{E} [\|X_\infty\|] = \mathbb{E} \left[ \liminf_{n \to \infty} \|X_n\| \right] \leq \liminf_{n \to \infty} \mathbb{E} [\|X_n\|] \leq C < \infty
\]

and therefore that \( X_\infty \in \mathbb{R} \) a.s.

9. If \( \{X_t^+\}_{t \geq 0} \) is uniformly integrable, then, by Vitalli's convergence Theorem 15.51 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

\[
\mathbb{E} [X_t 1_A] \leq \limsup_{\mathbb{D} \ni s \to \infty} \mathbb{E} [X_s 1_A] = \limsup_{\mathbb{D} \ni s \to \infty} (\mathbb{E} [X_s^+ 1_A] - \mathbb{E} [X_s^- 1_A])
= \mathbb{E} [X_\infty^+ 1_A] - \liminf_{\mathbb{D} \ni s \to \infty} \mathbb{E} [X_s^- 1_A] \leq \mathbb{E} [X_\infty^+ 1_A] - \mathbb{E} \left[ \liminf_{\mathbb{D} \ni s \to \infty} X_s^- 1_A \right]
= \mathbb{E} [X_\infty^+ 1_A] - \mathbb{E} [X_\infty^- 1_A] = \mathbb{E} [X_\infty 1_A].
\]

Since \( A \in \mathcal{B}_t \) was arbitrary we may conclude that \( X_t \leq \mathbb{E} [X_\infty] \mathbb{1}_A \) a.s. for all \( t \in \mathbb{R}_+ \).

Conversely if we suppose that \( X_t \leq \mathbb{E} [X_\infty] \mathbb{1}_A \) a.s. for all \( t \in \mathbb{R}_+ \), then by Jensen's inequality, \( X_t^+ \leq \mathbb{E} [X_\infty^+] \mathbb{1}_B \) and therefore \( \{X_t^+\}_{t \geq 0} \) is uniformly integrable by Proposition 21.8 and Exercise 15.5.

**Example 31.25.** In this example we show that there exists a right continuous submartingale, \( \{X_t\}_{t \geq 0} \), such that \( \{X_n\}_{n=1}^\infty \) is not uniformly integrable for some bounded increasing sequence \( \{s_n\}_{n=1}^\infty \). Indeed, let

\[
X_t := -M_{\tan(\frac{\pi}{t+1})}^0,
\]

where \( \{M_t\}_{t \geq 0} \) is the martingale constructed in Exercise 31.1. Then it is easily checked that \( \{X_t\}_{t \geq 0} \) is a \( \{ \mathcal{B}_{\tan(\frac{\pi}{t+1})} \}_{t \geq 0} \) - 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 31.24 we cannot suppose \( \{s_n\}_{n=1}^\infty \subset \mathbb{D} \cap [0, t) \) with \( s_n \uparrow t \).

**Exercise 31.2.** If \( \{X_t\}_{t \geq 0} \) is a right continuous submartingale on a filtered probability space, \( \left( \Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P \right) \), then \( s \to \mathbb{E} X_s \) is right continuous at \( t \) and \( X \) is a \( \{\mathcal{B}_t^+\} \) - submartingale.

**Exercise 31.3.** Let \( \{X_t\}_{t \geq 0} \) be a submartingale on a filtered probability space, \( \left( \Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P \right) \), \( t \in \mathbb{R}_+ \), and \( X_s \overset{P}{\rightarrow} X_t \) as \( s \downarrow t \), then \( s \to \mathbb{E} X_s \) is right continuous at \( t \).

**Theorem 31.26 (Regularizing Submartingales).** Let \( \{X_t\}_{t \geq 0} \) be a submartingale on a filtered probability space, \( \left( \Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, P \right) \), and let \( \Omega_0 \) and \( \tilde{X}_t \) be as in Theorem 31.24 applied to \( \{X_t\}_{t \in \mathbb{D}} \). Further let

\[
\tilde{X}_t (\omega) := \begin{cases} X_t (\omega) & \text{if } \omega \in \Omega_0 \\ 0 & \text{if } \omega \notin \Omega_0, \end{cases}
\]

and \( \{\tilde{B}_t := \tilde{B} + \mathcal{N}\}_{t \geq 0} \) where \( \mathcal{N} \) is the collection of \( P - \text{null } \) subsets of \( \Omega \) in \( \mathcal{B} \). Then:

1. \( \{\tilde{X}_t\}_{t \geq 0} \) is a \( \{\tilde{B}_t^+\} \) - submartingale which is right continuous with left hand limits.
2. \( \mathbb{E} [X_t | \mathcal{B}_t] \geq X_t \) 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 \( \{\mathcal{B}_t\}_{t \geq 0} \) is right continuous, then \( \tilde{X}_t \geq X_t \) 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 \( \{X_t\}_{t \geq 0} \) is right continuous in probability, then \( \{X_t\}_{t \geq 0} \) has a right continuous modification possessing left hand limits, \( \{\hat{X}_t\}_{t \geq 0} \), such that \( \{\hat{X}_t\}_{t \geq 0} \) is a \( \{\mathcal{B}_t\}_{t \geq 0} \) – submartingale.

**Proof.** 1. Since \( \Omega \setminus \Omega_0 \in \mathcal{N} \) and \( \hat{X}_t \) is a \( \mathcal{B}_t^+ \) – measurable, it follows that \( \hat{X}_t \) is \( \mathcal{B}_t^+ \) – measurable. Hence \( \{\hat{X}_t\}_{t \geq 0} \) is an adapted process. Since \( \hat{X} \) is a modification of \( X \) which is already a \( \{\mathcal{B}_t\}_{t \geq 0} \) – submartingale (see Lemma 31.15) it follows that \( \hat{X} \) is also a \( \{\mathcal{B}_t\}_{t \geq 0} \) – 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 31.26. 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

\[
\mathbb{E} [\hat{X}_t : A] = \lim_{n \to \infty} \mathbb{E} [X_{t_n} : A] \geq \lim_{n \to \infty} \mathbb{E} [X_t : A].
\]

Since \( A \in \mathcal{B}_t \) was arbitrary, it follows that

\[
X_t \leq \mathbb{E} [\hat{X}_t | \mathcal{B}_t] = \mathbb{E} [\hat{X}_t | \mathcal{B}_t] \text{ a.s.}\tag{31.22}
\]

If equality holds in Eq. (31.22), \( \mathbb{E} X_t = \mathbb{E} \hat{X}_t \) which is right continuous by Exercise 31.2. Conversely if \( t \to \mathbb{E} X_t \) is right continuous, it follows from item 6. of Theorem 31.26 that \( \hat{X}_t = X_t \) a.s.

3. Since \( X_t = \hat{X}_t \) a.s. and \( \hat{X}_t \) is \( \mathcal{B}_t^+ = \mathcal{B}_t \) – measurable, we find

\[
\hat{X}_t = \bar{X}_t = \mathbb{E} [X_t | \mathcal{B}_t] = \mathbb{E} [\bar{X}_t | \mathcal{B}_t] \text{ a.s.}
\]

With this observation (i.e. \( \bar{X}_t = \mathbb{E} [X_t | \mathcal{B}_t] \) a.s.) the assertions in item 3. follow directly from those in item 2.

Now suppose that \( \{X_t\}_{t \geq 0} \) is right continuous in probability. By Proposition 31.16 and Lemma 31.15 \( \{\bar{X}_t\}_{t \geq 0} \) is a \( \{\mathcal{B}_t\}_{t \geq 0} \) – submartingale such that, by Exercise 31.3, \( t \to \mathbb{E} X_t \) is right continuous. Therefore, by items 1. and 3. of the theorem, \( X \) defined above is the desired modification of \( X \).

**Example 31.27.** Let be a Poisson process, \( \{N_t\}_{t \geq 0} \), with parameter \( \lambda \) as described in Example 28.9. Since \( \{N_t\}_{t \geq 0} \) has independent increments, it follows that \( \{N_t\}_{t \geq 0} \) is a \( \{\mathcal{B}_t := \sigma (N_s : s \leq t)\}_{t \geq 0} \) – martingale. By example 25.14, 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 31.26 that there is a modification, \( \tilde{N} \) and \( \bar{N} \) such that \( \{\tilde{N}_t\}_{t \geq 0} \) is a \( \{\mathcal{B}_t\}_{t \geq 0} \) – 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 [3, Theorem 1.9.4 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 31.28 (Optional sampling II).** Suppose \( \{X_t\}_{t \geq 0} \) is a right continuous submartingale on a filtered probability space, \( (\Omega, \mathcal{B}, \{\mathcal{B}_t\}_{t \geq 0}, \mathbb{P}) \) such that \( \{X^+_t\}_{t \geq 0} \) is uniformly integrable. Then for any two optional times \( \sigma, \tau \in L^1 (\mathbb{P}) \) and \( X_{\sigma \land \tau} \leq \mathbb{E} [X_\tau | \mathcal{B}_\sigma^+] \). 

In particular if \( \{M_t\}_{t \geq 0} \) is a right continuous uniformly integrable martingale, then

\[
M_{\sigma \land \tau} = \mathbb{E} [M_\tau | \mathcal{B}_\sigma^+] . \tag{31.23}
\]

**Proof.** Let \( X_\infty := \lim_{t \to \infty} X_t \in L^1 (\mathbb{P}) \) as in Theorem 31.24 so that \( X_t \leq \mathbb{E} [X_\infty | \mathcal{B}_t] \) for all \( t \in \mathbb{D} \). For \( t \in \mathbb{R}_+ \), let \( \{t_n\}_{n=1} \subset \mathbb{D} \cap (t, \infty) \) be such that \( t_n \downarrow t \), then by Corollary 21.81

\[
X_t = \lim_{n \to \infty} X_{t_n} \leq \lim_{n \to \infty} \mathbb{E} [X_\infty | \mathcal{B}_t] = \mathbb{E} [X_\infty | \mathcal{B}_t^+] \text{ a.s.}
\]

Conditioning this inequality on \( \mathcal{B}_t \) also allows us to conclude that \( X_t \leq \mathbb{E} [X_\infty | \mathcal{B}_t] \). We may now reduce the inequality in Eq. (31.23) to the case in Theorem 31.9 where \( \tau \) is a bounded stopping time by simply identifying \([0, \pi/2]\) with \([0, \infty)\) via the map, \( t \mapsto \tan t \). More precisely, let \( \{Y_t\}_{0 \leq t \leq \pi/2} \) be the right continuous \( \mathbb{B} := \mathbb{B}_{\tan t} \) submartingale defined by \( Y_t := X_{\tan t} \). As \( \tan^{-1} (\sigma) \) and \( \tan^{-1} (\tau) \) are two bounded \( \{\mathbb{B}_t\}_{0 \leq t \leq \pi/2} \) – optional times, we may apply Theorem 31.9 to find;

\[
X_{\sigma \land \tau} = Y_{\tan^{-1} (\sigma) \land \tan^{-1} (\tau)} \leq \mathbb{E} [Y_{\tan^{-1} (\tau)} | \mathcal{B}_{\tan^{-1} (\sigma)}^+] = \mathbb{E} [X_\tau | \mathcal{B}_\sigma^+] \text{ a.s.}
\]

For the martingale assertions, simply apply Eq. (31.23) with \( X_t = M_t \) and \( X_t = -M_t \).

\[\footnote{According to Exercise 31.2, \( \{X_t\}_{t \geq 0} \) is also a \( \{\mathcal{B}_t^+\} \) – submartingale. Therefore for the purposes of this Theorem, there is no loss in generality in assuming that \( \mathcal{B}_t^+ = \mathcal{B}_t \).} \]

---

Page: 493 job: prob macro: svmonob.cls date/time: 2-Nov-2018/7:31
References


Homework Problem Solutions

32.1 Resnik repeats