Bruce K. Driver

Math 285 Stochastic Processes Spring 2016

June 3, 2016 *File:285notes.tex*

Contents

Part

Homework Problems

Math 285 Homework Problem List for S2016

Note: solutions to Lawler Problems will appear after all of the Lecture Note Solutions.

-1.1 Homework 1. Due Friday, April 8, 2016

- **Look at** from lecture note exercises: 3.1
- **Hand in** lecture note exercises: 3.2, 3.3, 4.1, 4.2, 4.3
- **Hand in** from **Lawler** §5.1 on page 125.

-1.2 Homework 2. Due Friday, April 15, 2016

- **Look at** from lecture note exercises: 6.3, 6.4, 6.6
- **Hand in** lecture note exercises: 6.1, 6.5, 6.7
- **Hand in** from **Lawler** §1.1, 1.4, 1.19

-1.3 Homework 3. Due Friday, April 22, 2016

- **Look at** from **Lawler** §1.10,
- **Hand in** from **Lawler** §1.5, 1.8, 1.9, 1.14, 1.18*, 2.3

*Hint: show the invariant distribuiton is uniform.

-1.4 Homework 4. Due Friday, April 29, 2016

- **Look at** lecture note exercises: 10.1
- **Hand in** from **Lawler** problems: §7.6, 7.7, 7.8.
- Please use the result in 7.8 to verify your numerical approximations found in 7.7.

***Hints for 7.8.** Recall that for $n \in \mathbb{N}$ that

$$S_n = \left\{ \mathbf{k} = (k_0, \ldots, k_n) \in \{0,1\}^{n+1} : \; k_{i-1} + k_i \le 1 \text{ for } 1 \le i \le n \right\}$$

and your goal is to compute

$$p_n(m) := \frac{\#\{\mathbf{k} \in S_n : k_m = 1\}}{\#(S_n)} \text{ for } 0 \le m \le n.$$

As Lawler suggests, for $i, j \in \{0,1\}$ and $m \in \mathbb{N}$, let[1]

$$r_m(ij) = \#\{\mathbf{k} = (k_0, \ldots, k_m) \in S_m : k_0 = i \text{ and } k_m = j\}.$$

Notice that if $\mathbf{k} = (k_0, \ldots, k_n) \in S_n$ with $k_m = 1$, then

$$(0, k_0, \ldots, k_{m-1}, 1) \in S_{m+1} \text{ and } (1, k_{m+1}, \ldots, k_n, 0) \in S_{n-m+1}$$

(and visa versa) where k_{m-1} and k_{m+1} must both be zero so that

$$(0, k_0, \ldots, k_{m-2}, 0) \in S_m \text{ and } (0, k_{m+2}, \ldots, k_n, 0) \in S_{n-m}.$$

From these considerations we find

$$\#\{\mathbf{k} \in S_n : k_m = 1\} = r_{m+1}(01) \cdot r_{n-m+1}(10) = r_m(00) \cdot r_{n-m}(00).$$

Similarly $\mathbf{k} = (k_0, \ldots, k_m) \in S_m$ then $(0, k_0, \ldots, k_n, 0) \in S_{n+2}$ and visa versa from which we learn $\#(S_n) = r_{n+2}(00)$. Combining these results shows

$$p_n(m) = \frac{r_m(00) \cdot r_{n-m}(00)}{r_{n+2}(00)}.$$

A little thought shows this formula is correct at $m = 0$ and $m = n - m$ provided we use the convention that $r_0(00) = 1$. Now follow the outline in the Lawler in order to find, $y_m = y(m) := r_m(00)$.

-1.5 Homework 5. Due Friday, May 6, 2016

- **Look at** from lecture note exercises: 13.1, 13.2
- **Hand in** lecture note exercises: 13.3, 13.5
- **Hand in** from **Lawler** § 5.4, 5.7a, 5.8a, 5.12

[1] I have changed the indexing a bit since Lawler's choices are a bit confusing.

-1.6 Homework 6. Due Friday, May 13, 2016

- **Look at** from lecture note exercises: 13.7, 13.8, 13.9, 13.14
- **Look at from Lawler** § 5.13
- **Hand in** lecture note exercises: 13.11, 13.13
- **Hand in** from **Lawler** § 5.7b, 5.9*, 5.14

***Correction to 5.9** The condition, $\mathbf{P}f(x) = g(x)$ for $x \in S \setminus A$, should read $\mathbf{P}f(x) - f(x) = g(x)$ for $x \in S \setminus A$.

-1.7 Homework 7. Due Friday, May 20, 2016

- **Look at** from lecture note exercises: 15.4, 15.6
- **Hand in** lecture note exercises: 15.1, 15.2, 15.3, 15.5

-1.8 Homework 8. Due Friday, May 27, 2016

- **Look at** from lecture note exercises: 15.7, 15.8, 15.9, 19.1
- **Hand in** lecture note exercises: 15.12, 16.1, 16.3, 19.5

-1.9 Homework 9. Due Friday, June 3, 2016

- **Look at** from lecture note exercises: 16.4, 16.5, 19.2
- **Hand in** lecture note exercises: 16.6, 16.7

Part I

Background Material

0

Introduction

Definition 0.1 (Stochastic Process via Wikipedia). *..., a **stochastic process**, or often **random process**, is a collection of random variables representing the evolution of some system of random values over time. This is the probabilistic counterpart to a deterministic process (or deterministic system). Instead of describing a process which can only evolve in one way (as in the case, for example, of solutions of an ordinary differential equation), in a stochastic, or random process, there is some indeterminacy: even if the initial condition (or starting point) is known, there are several (often infinitely many) directions in which the process may evolve.*

0.1 Deterministic Modeling

In deterministic modeling one often has a dynamical system on a state space S. The dynamical system often takes on one of the two forms;

1. there exists $f : S \to S$ and a state x_n then evolves according to the rule $x_{n+1} = f(x_n)$. [More generally one might allow $x_{n+1} = f_n(x_0, \ldots, x_n)$ where $f_n : S^{n+1} \to S$ is a given function for each n.
2. There exists a vector field f on S (where now $S = \mathbb{R}^d$ or a manifold) such that $\dot{x}(t) = f(x(t))$. [More generally, we might allow for $\dot{x}(t) = f\left(t; x|_{[0,t]}\right)$, a functional differential equation.]

Goals: the goals in this case then have to do with deriving the properties of the trajectories given the properties of the driving dynamics incorporated in f. For example, think of a golfer trying to make a put or a hot-air balloonist trying to find a path from point A to point B.

0.2 Stochastic Modeling

Much of our time in this course will be to explore the above two situations where some extra randomness is added at each state of the game. The point being that in many situations the exact nature of the dynamics is not known or is rapidly changing. What is known are statistical properties of the dynamics – i.e. likely hoods that the dynamics will be of a certain form. This amounts to replacing f above by some sort of random f and then resolving the problems. However, now rather than trying to find the properties of a given trajectory we instead try to find properties of the statistics of the now random trajectories. Typically when comparing theory to experiment one has to now average experimental results (hoping to use the law of large numbers) to make contact with the mathematical theory. Here is a little more detail on the typical sort of scenarios that we will consider in this course.

1. We may now have that $X_{n+1} \in S$ is random and evolves according to

$$X_{n+1} = f(X_n, \xi_n)$$

where $\{\xi_n\}_{n=0}^{\infty}$ is a sequence of i.i.d. random variables. Alternatively put, we might simply let $f_n := f(\cdot, \xi_n)$ so that $f_n : S \to S$ is a sequence of i.i.d. random functions from S to S. Then $\{X_n\}_{n=0}^{\infty}$ is defined recursively by

$$X_{n+1} = f_n(X_n) \text{ for } n = 0, 1, 2, \ldots. \tag{0.1}$$

This is the typical example of a time-homogeneous Markov chain. We assume that $X_0 \in S$ is given with an initial condition which is either deterministic or is independent of the $\{f_n\}_{n=0}^{\infty}$.

2. Later in the course we will study the continuous time analogue,

$$\text{“}\dot{X}_t = f_t(X_t)\text{”}$$

where $\{f_t\}_{t\geq 0}$ are again i.i.d. random vector-fields. The continuous time case will require substantially more technical care. For example, one often considers the controlled differential equation,

$$\dot{X}_t = f(X_t)\dot{B}_t \tag{0.2}$$

where $\{B_t\}_{t\geq 0}$ is Brownian motion or equivalently $\dot{B}_t$ is “white noise” or B_t is a Poisson process. The Poisson noise is often used to model arrival times in networks or in queues (i.e. service lines) or appear in electrical circuits due to “thermal fluctuations” to name a few. See for example, Johnson Noise and Shot Noise by Dennis V. Perepelitsa, November 27, 2006.) Here are two quotes from this article.

"The thermal agitation of the charge carriers in any circuit causes a small, yet detectable, current to flow. J.B. Johnson was the first to present a quantitative analysis of this phenomenon, which is unaffected by the geometry and material of the circuit."

"The quantization of charge carried by electrons in a circuit also contributes to a small amount of noise. Consider a photoelectric circuit in which current caused by the photoexcitation of electrons flow to the anode."

3. We will also consider a class of processes known as (Sub/Super) martingales which encode information about fair (or not so fair) games of chance amongst many other applications.

1

Probabilities, Expectations, and Distributions

1.1 Basics of Probabilities and Expectations

Our first goal in this course is to describe modern probability with "sufficient" precision to allow us to do the required computations for this course. We will thus be neglecting some technical details involving measures and $\sigma-$ algebras. The knowledgeable reader should be able to fill in the missing hypothesis while the less knowledgeable readers should not be too harmed by the omissions to follow.

1. $(\Omega, \mathbb{P})$ will denote a **probability space** and S will denote a set which is called **state space**. Informally put, Ω is a set (often the sample space) and $\mathbb{P}$ is a function on all[1] subsets of Ω (subsets of Ω are called **events)** with the following properties;
   a) $\mathbb{P}(A) \in [0,1]$ for all $A \subset \Omega$,
   b) $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$.
   c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ is $A \cap B = \emptyset$. More generally, if $A_n \subset \Omega$ for all n with $A_n \cap A_m = \emptyset$ for $m \neq n$ we have
   $$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$
2. A **random variable,** Z, is a function from Ω to $\mathbb{R}$ or perhaps some other range space. For example if $A \subset \Omega$ is an event then the **indicator function of** A,
   $$1_A(\omega) := \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A, \end{cases}$$
   is a random variable.
3. Note that every real value random variable, Z, may be approximated by the discrete random variables
   $$Z_\varepsilon := \sum_{n \in \mathbb{Z}} n\varepsilon \cdot 1_{\{n\varepsilon \le Z < (n+1)\varepsilon\}} \text{ for all } \varepsilon > 0. \tag{1.1}$$
   As we usually do in probability, $\{n\varepsilon \le Z < (n+1)\varepsilon\}$, stands for the event more precisely written as;
   $$\{\omega \in \Omega : n\varepsilon \le Z(\omega) < (n+1)\varepsilon\}.$$
4. $\mathbb{E}Z$ will denote the **expectation** of a random variable, $Z : \Omega \to \mathbb{R}$ which is defined as follows. If Z only takes on a finite number of real values $\{z_1, \ldots, z_m\}$ we define
   $$\mathbb{E}Z = \sum_{i=1}^{m} z_i \mathbb{P}(Z = z_i).$$
   For general $Z \ge 0$ we set $\mathbb{E}Z = \lim_{n\to\infty} \mathbb{E}Z_n$ where $\{Z_n\}_{n=1}^{\infty}$ is any sequence of discrete random variables such that $0 \le Z_n \uparrow Z$ as $n \uparrow \infty$. Finally if Z is real valued with $\mathbb{E}|Z| < \infty$ (in which case we say Z is **integrable)** we set $\mathbb{E}Z = \mathbb{E}\mathbb{Z}_+ - \mathbb{E}Z_-$ where $Z_\pm = \max(\pm Z, 0)$. With these definition one eventually shows via the dominated convergence theorem below; if $f : \mathbb{R} \to \mathbb{R}$ is a bounded continuous function, then
   $$\mathbb{E}[f(Z)] = \lim_{\Delta \to 0} \sum_{n \in \mathbb{Z}} f(n\Delta) \mathbb{P}(n\Delta < Z \le (n+1)\Delta).$$
   We summarize this informally[2] by writing;
   $$\mathbb{E}[f(Z)] = \text{“} \int_{\mathbb{R}} f(z) \mathbb{P}(z < Z \le z + dz) \text{.”}$$
5. The expectation has the following basic properties;
   a) **Expectations of indicator functions:** $\mathbb{E}1_A = \mathbb{P}(A)$ for all events $A \subset \Omega$.
   b) **Linearity**: if X and Y are integrable random variables and $c \in \mathbb{R}$, then
   $$\mathbb{E}[X + cY] = \mathbb{E}X + c\mathbb{E}Y.$$
   c) **Monotinicity:** if $X, Y : \Omega \to \mathbb{R}$ are integrable with $\mathbb{P}(X \le Y) = 1$, then $\mathbb{E}X \le \mathbb{E}Y$. In particular if $X = Y$ **almost surely** (a.s.) (i.e. $\mathbb{P}(X = Y) = 1$), then $\mathbb{E}X = \mathbb{E}Y$. [What happens on sets of probability 0 are typically irrelevant.]

[1] This is often a lie! Nevertheless, for our purposes it will be reasonably safe to ignore this lie.

[2] Think of $z = n\Delta$ and $dz = \Delta$.

d) **Finite expectation $\implies$ finite random variable.** If $Z : \Omega \to [0,\infty]$ is a random variable such that $\mathbb{E}Z < \infty$ then $\mathbb{P}(Z = \infty) = 0$, i.e. $\mathbb{P}(Z < \infty) = 1$.

e) **MCT**: the **monotone convergence theorem** holds; if $0 \le Z_n \uparrow Z$ then

$$\uparrow \lim_{n\to\infty} \mathbb{E}[Z_n] = \mathbb{E}[Z] \text{ (with } \infty \text{ allowed as a possible value).}$$

Example 1: If $\{A_n\}_{n=1}^\infty$ is a sequence of events such that $A_n \uparrow A$ (i.e. $A_n \subset A_{n+1}$ for all n and $A = \cup_{n=1}^\infty A_n$), then

$$\mathbb{P}(A_n) = \mathbb{E}[1_{A_n}] \uparrow \mathbb{E}[1_A] = \mathbb{P}(A) \text{ as } n \to \infty$$

Example 2: If $X_n : \Omega \to [0,\infty]$ for $n \in \mathbb{N}$ then

$$\mathbb{E}\sum_{n=1}^\infty X_n = \mathbb{E}\lim_{N\to\infty}\sum_{n=1}^N X_n = \lim_{N\to\infty}\mathbb{E}\sum_{n=1}^N X_n = \lim_{N\to\infty}\sum_{n=1}^N \mathbb{E}X_n = \sum_{n=1}^\infty \mathbb{E}X_n.$$

Example 3: Suppose S is a finite or countable set and $X : \Omega \to S$ is a random function. Then for any $f : S \to [0,\infty]$,

$$\mathbb{E}[f(X)] = \sum_{s\in S} f(s)\,\mathbb{P}(X = s).$$

Indeed, we have

$$f(X) = \sum_{s\in S} f(s)\,1_{\{X=s\}}$$

and so by Example 2. above,

$$\begin{aligned}\mathbb{E}[f(X)] &= \sum_{s\in S}\mathbb{E}\left[f(s)\,1_{\{X=s\}}\right] \\ &= \sum_{s\in S} f(s)\,\mathbb{E}\left[1_{\{X=s\}}\right] = \sum_{s\in S} f(s)\,\mathbb{P}(X = s).\end{aligned}$$

f) **DCT**: the **dominated convergence theorem** holds, if

$$\mathbb{E}\left[\sup_n |Z_n|\right] < \infty \text{ and } \lim_{n\to\infty} Z_n = Z, \text{ then}$$
$$\mathbb{E}\left[\lim_{n\to\infty} Z_n\right] = \mathbb{E}Z = \lim_{n\to\infty}\mathbb{E}Z_n.$$

Example 1: If $\{A_n\}_{n=1}^\infty$ is a sequence of events such that $A_n \downarrow A$ (i.e. $A_n \supset A_{n+1}$ for all n and $A = \cap_{n=1}^\infty A_n$), then

$$\mathbb{P}(A_n) = \mathbb{E}[1_{A_n}] \downarrow \mathbb{E}[1_A] = \mathbb{P}(A) \text{ as } n \to \infty.$$

The dominating function is 1 here.

Example 2: If $\{X_n\}_{n=1}^\infty$ is a sequence of real valued random variables such that

$$\mathbb{E}\sum_{n=1}^\infty |X_n| = \sum_{n=1}^\infty \mathbb{E}|X_n| < \infty,$$

then; 1) $Z := \sum_{n=1}^\infty |X_n| < \infty$ a.s. and hence $\sum_{n=1}^\infty X_n = \lim_{N\to\infty}\sum_{n=1}^N X_n$ exist a.s., 2) $\left|\sum_{n=1}^N X_n\right| \le Z$ and $\mathbb{E}Z < \infty$, and so 3) by DCT,

$$\mathbb{E}\sum_{n=1}^\infty X_n = \mathbb{E}\lim_{N\to\infty}\sum_{n=1}^N X_n = \lim_{N\to\infty}\mathbb{E}\sum_{n=1}^N X_n = \lim_{N\to\infty}\sum_{n=1}^N \mathbb{E}X_n = \sum_{n=1}^\infty \mathbb{E}X_n.$$

g) **Fatou's Lemma: Fatou's lemma** holds; if $0 \le Z_n \le \infty$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} Z_n\right] \le \liminf_{n\to\infty}\mathbb{E}[Z_n].$$

This may be proved as an application of MCT.

6. **Discrete distributions.** If S is a discrete set, i.e. finite or countable and $X : \Omega \to S$ we let

$$\rho_X(s) := \mathbb{P}(X = s).$$

Notice that if $f : S \to \mathbb{R}$ is a function, then $f(X) = \sum_{s\in S} f(s)\,1_{\{X=s\}}$ and therefore,

$$\mathbb{E}f(X) = \sum_{s\in S} f(s)\,\mathbb{E}1_{\{X=s\}} = \sum_{s\in S} f(s)\,\mathbb{P}(X = s) = \sum_{s\in S} f(s)\,\rho_X(s).$$

More generally if $X_i : \Omega \to S_i$ for $1 \le i \le n$ we let

$$\rho_{X_1,\dots,X_n}(\mathbf{s}) := \mathbb{P}(X_1 = s_1, \dots, X_n = s_n)$$

for all $\mathbf{s} = (s_1,\dots,s_n) \in S_1 \times \cdots \times S_n$ and

$$\mathbb{E}f(X_1,\dots,X_n) = \sum_{\mathbf{s}=(s_1,\dots,s_n)} f(\mathbf{s})\,\rho_{X_1,\dots,X_n}(\mathbf{s}).$$

7. **Continuous density functions.** If S is $\mathbb{R}$ or $\mathbb{R}^n$, we say $X : \Omega \to S$ is a "**continuous random variable,**" if there exists a **probability density function**, $\rho_X : S \to [0,\infty)$ such that for all bounded (or positive) functions, $f : S \to \mathbb{R}$, we have

$$\mathbb{E}[f(X)] = \int_S f(x)\,\rho_X(x)\,dx.$$

8. Given random variables X and Y we let;
   a) $\mu_X := \mathbb{E}X$ be the **mean** of X.
   b) $\mathrm{Var}(X) := \mathbb{E}\left[(X-\mu_X)^2\right] = \mathbb{E}X^2 - \mu_X^2$ be the **variance** of X.
   c) $\sigma_X = \sigma(X) := \sqrt{\mathrm{Var}(X)}$ be the **standard deviation** of X.
   d) $\mathrm{Cov}(X,Y) := \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y$ be the **covariance** of X and Y.
   e) $Corr(X,Y) := \mathrm{Cov}(X,Y)/(\sigma_X\sigma_Y)$ be the **correlation** of X and Y.
9. **Tonelli's theorem;** if $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}_+$, then
$$\int_{\mathbb{R}^k} dx \int_{\mathbb{R}^l} dy f(x,y) = \int_{\mathbb{R}^l} dy \int_{\mathbb{R}^k} dx f(x,y) \text{ (with } \infty \text{ being allowed).}$$
10. **Fubini's theorem;** if $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}$ is a function such that
$$\int_{\mathbb{R}^k} dx \int_{\mathbb{R}^l} dy |f(x,y)| = \int_{\mathbb{R}^l} dy \int_{\mathbb{R}^k} dx |f(x,y)| < \infty,$$
then
$$\int_{\mathbb{R}^k} dx \int_{\mathbb{R}^l} dy f(x,y) = \int_{\mathbb{R}^l} dy \int_{\mathbb{R}^k} dx f(x,y).$$

1.2 Some Discrete Distributions

Definition 1.1 (Generating Function). *Suppose that $N : \Omega \to \mathbb{N}_0$ is an integer valued random variable on a probability space, $(\Omega, \mathcal{B}, \mathbb{P})$. The generating function associated to N is defined by*
$$G_N(z) := \mathbb{E}\left[z^N\right] = \sum_{n=0}^{\infty} \mathbb{P}(N=n) z^n \text{ for } |z| \le 1. \tag{1.2}$$

By standard power series considerations, it follows that $\mathbb{P}(N=n) = \frac{1}{n!} G_N^{(n)}(0)$ so that G_N can be used to completely recover the distribution of N.

Proposition 1.2 (Generating Functions). *The generating function satisfies,*
$$G_N^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right] \text{ for } |z| < 1$$
and
$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}[N(N-1)\dots(N-k+1)],$$
where it is possible that one and hence both sides of this equation are infinite. In particular, $G'(1) := \lim_{z\uparrow 1} G'(z) = \mathbb{E}N$ and if $\mathbb{E}N^2 < \infty$,
$$\mathrm{Var}(N) = G''(1) + G'(1) - [G'(1)]^2. \tag{1.3}$$

Proof. By standard power series considerations, for $|z| < 1$,
$$\begin{aligned} G_N^{(k)}(z) &= \sum_{n=0}^{\infty} \mathbb{P}(N=n) \cdot n(n-1)\dots(n-k+1) z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right]. \end{aligned} \tag{1.4}$$
Since, for $z \in (0,1)$,
$$0 \le N(N-1)\dots(N-k+1) z^{N-k} \uparrow N(N-1)\dots(N-k+1) \text{ as } z \uparrow 1,$$
we may apply the MCT to pass to the limit as $z \uparrow 1$ in Eq. (1.4) to find,
$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}[N(N-1)\dots(N-k+1)].$$
■

Exercise 1.1 (Some Discrete Distributions). Let $p \in (0,1]$ and $\lambda > 0$. In the four parts below, the distribution of N will be described. You should work out the generating function, $G_N(z)$, in each case and use it to verify the given formulas for $\mathbb{E}N$ and $\mathrm{Var}(N)$.

1. Bernoulli(p) : $\mathbb{P}(N=1) = p$ and $\mathbb{P}(N=0) = 1-p$. You should find $\mathbb{E}N = p$ and $\mathrm{Var}(N) = p - p^2$.
2. $\mathrm{Bin}(n,p)$: $\mathbb{P}(N=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0,1,\dots,n$. ($\mathbb{P}(N=k)$ is the probability of k successes in a sequence of n independent yes/no experiments with probability of success being p.) You should find $\mathbb{E}N = np$ and $\mathrm{Var}(N) = n\left(p - p^2\right)$.
3. Geometric(p) : $\mathbb{P}(N=k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$. ($\mathbb{P}(N=k)$ is the probability that the kth – trial is the first time of success out a sequence of independent trials with probability of success being p.) You should find $\mathbb{E}N = 1/p$ and $\mathrm{Var}(N) = \frac{1-p}{p^2}$.
4. Poisson(λ) : $\mathbb{P}(N=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for all $k \in \mathbb{N}_0$. You should find $\mathbb{E}N = \lambda = \mathrm{Var}(N)$.

Exercise 1.2. Let $S_{n,p} \overset{d}{=} \mathrm{Bin}(n,p)$, $k \in \mathbb{N}$, $p_n = \lambda_n/n$ where $\lambda_n \to \lambda > 0$ as $n \to \infty$. Show that
$$\lim_{n\to\infty} \mathbb{P}(S_{n,p_n} = k) = \frac{\lambda^k}{k!} e^{-\lambda} = \mathbb{P}(\mathrm{Poi}(\lambda) = k).$$
Thus we see that for $p = O(1/n)$ and k not too large relative to n that for large n,
$$\mathbb{P}(\mathrm{Bin}(n,p) = k) \cong \mathbb{P}(\mathrm{Poi}(pn) = k) = \frac{(pn)^k}{k!} e^{-pn}.$$
(We will come back to the Poisson distribution and the related Poisson process later on.)

Independence

Definition 2.1. *We say that an event, A, is* ***independent*** *of an event, B, iff*[1]

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

We further say a collection of events $\{A_j\}_{j\in J}$ are ***independent*** *iff*

$$\mathbb{P}(\cap_{j\in J_0} A_j) = \prod_{j\in J_0} \mathbb{P}(A_j)$$

for any finite subset, J_0, of J.

Lemma 2.2. *If $\{A_j\}_{j\in J}$ is an independent collection of events then so is $\left\{A_j, A_j^c\right\}_{j\in J}$.*

Proof. First consider the case of two independent events, A and B. By assumption, $\mathbb{P}(A\cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$. Since A is the disjoint union of $A\cap B$ and $A\cap B^c$, the additivity of $\mathbb{P}$ implies,

$$\mathbb{P}(A) = \mathbb{P}(A\cap B) + \mathbb{P}(A\cap B^c) = \mathbb{P}(A)\,\mathbb{P}(B) + \mathbb{P}(A\cap B^c).$$

Solving this identity for $\mathbb{P}(A\cap B^c)$ gives,

$$\mathbb{P}(A\cap B^c) = \mathbb{P}(A)\,[1-\mathbb{P}(B)] = \mathbb{P}(A)\,\mathbb{P}(B^c).$$

Thus if $\{A, B\}$ are independent then so is $\{A, B^c\}$. Similarly we may show $\{A^c, B\}$ are independent and then that $\{A^c, B^c\}$ are independent. That is $\mathbb{P}\left(A^\varepsilon \cap B^\delta\right) = \mathbb{P}(A^\varepsilon)\,\mathbb{P}\left(B^\delta\right)$ where ε, δ is either "nothing" or "c."

The general case now easily follows similarly. Indeed, if $\{A_1, \dots, A_n\} \subset \{A_j\}_{j\in J}$ we must show that

$$\mathbb{P}(A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n}) = \mathbb{P}(A_1^{\varepsilon_1}) \dots \mathbb{P}(A_n^{\varepsilon_n})$$

where $\varepsilon_j = c$ or $\varepsilon_j =$ " ". But this follows from above. For example, $\{A_1 \cap \dots \cap A_{n-1}, A_n\}$ are independent implies that $\{A_1 \cap \dots \cap A_{n-1}, A_n^c\}$ are independent and hence

$$\begin{aligned}\mathbb{P}(A_1 \cap \dots \cap A_{n-1} \cap A_n^c) &= \mathbb{P}(A_1 \cap \dots \cap A_{n-1})\,\mathbb{P}(A_n^c) \\ &= \mathbb{P}(A_1) \dots \mathbb{P}(A_{n-1})\,\mathbb{P}(A_n^c).\end{aligned}$$

Thus we have shown it is permissible to add A_j^c to the list for any $j \in J$. ■

Lemma 2.3. *If $\{A_n\}_{n=1}^\infty$ is a sequence of independent events, then*

$$\mathbb{P}(\cap_{n=1}^\infty A_n) = \prod_{n=1}^\infty \mathbb{P}(A_n) := \lim_{N\to\infty} \prod_{n=1}^N \mathbb{P}(A_n).$$

Proof. Since $\cap_{n=1}^N A_n \downarrow \cap_{n=1}^\infty A_n$, it follows that

$$\mathbb{P}(\cap_{n=1}^\infty A_n) = \lim_{N\to\infty} \mathbb{P}\left(\cap_{n=1}^N A_n\right) = \lim_{N\to\infty} \prod_{n=1}^N \mathbb{P}(A_n),$$

where we have used the independence assumption for the last equality.

The convergence assertion used above follows from DCT Indeed, $1_{\cap_{n=1}^N A_n} \downarrow 1_{\cap_{n=1}^\infty A_n}$ and all functions are dominated by 1 and therefore,

$$\mathbb{P}(\cap_{n=1}^\infty A_n) = \mathbb{E}\left[1_{\cap_{n=1}^\infty A_n}\right] = \lim_{N\to\infty} \mathbb{E}\left[1_{\cap_{n=1}^N A_n}\right] = \lim_{N\to\infty} \mathbb{P}\left(\cap_{n=1}^N A_n\right).$$

■

2.1 Borel Cantelli Lemmas

Definition 2.4 (A_n i.o.). *Suppose that $\{A_n\}_{n=1}^\infty$ is a sequence of events. Let*

$$\{A_n \text{ i.o.}\} := \left\{\sum_{n=1}^\infty 1_{A_n} = \infty\right\}$$

denote the event where infinitely many of the events, A_n, occur. The abbreviation, "i.o." stands for ***infinitely often****.*

For example if X_n is H or T depending on whether a heads or tails is flipped at the nth step, then $\{X_n = H \text{ i.o.}\}$ is the event where an infinite number of heads was flipped.

[1] Shortly we will consider conditional probabilities, $\mathbb{P}(\cdot|B)$. With this notation, A is independent of B iff $\mathbb{P}(A|B) = \mathbb{P}(A)$, i.e. given the information gained by B occurring does not affect the likelihood that A occurred.

Lemma 2.5 (The First Borell – Cantelli Lemma). *If $\{A_n\}$ is a sequence of events such that $\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$, then*

$$\mathbb{P}(\{A_n \ i.o.\}) = 0.$$

Proof. Since

$$\infty > \sum_{n=0}^{\infty} \mathbb{P}(A_n) = \sum_{n=0}^{\infty} \mathbb{E}1_{A_n} = \mathbb{E}\left[\sum_{n=0}^{\infty} 1_{A_n}\right]$$

it follows that $\sum_{n=0}^{\infty} 1_{A_n} < \infty$ almost surely (a.s.), i.e. with probability 1 only finitely many of the $\{A_n\}$ can occur. ■

Under the additional assumption of independence we have the following strong converse of the first Borel-Cantelli Lemma.

Lemma 2.6 (Second Borel-Cantelli Lemma). *If $\{A_n\}_{n=1}^{\infty}$ are independent events, then*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad \Longrightarrow \quad \mathbb{P}(\{A_n \ i.o.\}) = 1. \tag{2.1}$$

Proof. We are going to show $\mathbb{P}(\{A_n \text{ i.o.}\}^c) = 0$. Since,

$$\{A_n \text{ i.o.}\}^c = \left\{\sum_{n=1}^{\infty} 1_{A_n} = \infty\right\}^c = \left\{\sum_{n=1}^{\infty} 1_{A_n} < \infty\right\},$$

we see that $\omega \in \{A_n \text{ i.o.}\}^c$ iff there exists $n \in \mathbb{N}$ such that $\omega \notin A_m$ for all $m \geq n$. Thus we have shown, if $\omega \in \{A_n \text{ i.o.}\}^c$ then $\omega \in B_n := \cap_{m \geq n} A_m^c$ for some n and therefore,

$$\{A_n \text{ i.o.}\}^c = \cup_{n=1}^{\infty} B_n.$$

As $B_n \uparrow \{A_n \text{ i.o.}\}^c$ we have

$$\mathbb{P}(\{A_n \text{ i.o.}\}^c) = \lim_{n \to \infty} \mathbb{P}(B_n).$$

But making use of the independence (see Lemmas 2.2 and 2.3) and the estimate, $1 - x \leq e^{-x}$, see Figure 2.1 below, we find

$$\mathbb{P}(B_n) = \mathbb{P}(\cap_{m \geq n} A_m^c) = \prod_{m \geq n} \mathbb{P}(A_m^c) = \prod_{m \geq n} [1 - \mathbb{P}(A_m)]$$

$$\leq \prod_{m \geq n} e^{-\mathbb{P}(A_m)} = \exp\left(-\sum_{m \geq n} \mathbb{P}(A_m)\right) = e^{-\infty} = 0.$$

■

Combining the two Borel Cantelli Lemmas gives the following Zero-One Law.

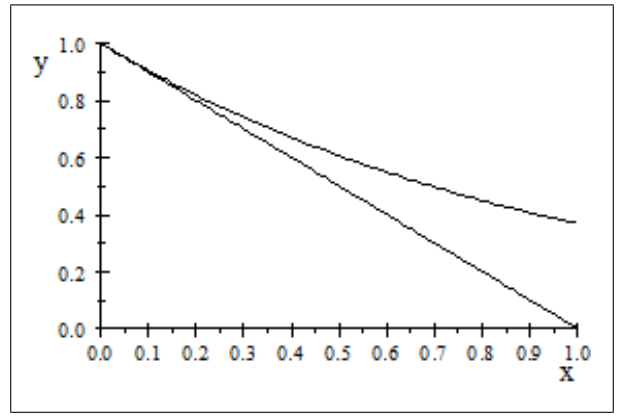


Fig. 2.1. Comparing e^{-x} and $1 - x$.

Corollary 2.7 (Borel's Zero-One law). *If $\{A_n\}_{n=1}^{\infty}$ are independent events, then*

$$\mathbb{P}(A_n \ i.o.) = \begin{cases} 0 \ \text{if} \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \\ 1 \ \text{if} \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \end{cases}.$$

Example 2.8. If $\{X_n\}_{n=1}^{\infty}$ denotes the outcomes of the toss of a coin such that $\mathbb{P}(X_n = H) = p > 0$, then $\mathbb{P}(X_n = H \text{ i.o.}) = 1$.

Example 2.9. If a monkey types on a keyboard with each stroke being independent and identically distributed with each key being hit with positive probability. Then eventually the monkey will type the text of the bible if she lives long enough. Indeed, let S be the set of possible key strokes and let $(s_1, \ldots, s_N)$ be the strokes necessary to type the bible. Further let $\{X_n\}_{n=1}^{\infty}$ be the strokes that the monkey types at time n. Then group the monkey's strokes as $Y_k := \left(X_{kN+1}, \ldots, X_{(k+1)N}\right)$. We then have

$$\mathbb{P}(Y_k = (s_1, \ldots, s_N)) = \prod_{j=1}^{N} \mathbb{P}(X_j = s_j) =: p > 0.$$

Therefore,

$$\sum_{k=1}^{\infty} \mathbb{P}(Y_k = (s_1, \ldots, s_N)) = \infty$$

and so by the second Borel-Cantelli lemma,

$$\mathbb{P}(\{Y_k = (s_1, \ldots, s_N)\} \text{ i.o. } k) = 1.$$

2.2 Independent Random Variables

Definition 2.10. *We say a collection of discrete random variables, $\{X_j\}_{j\in J}$, are **independent** if*

$$\mathbb{P}(X_{j_1}=x_1,\dots,X_{j_n}=x_n)=\mathbb{P}(X_{j_1}=x_1)\cdots\mathbb{P}(X_{j_n}=x_n) \tag{2.2}$$

for all possible choices of $\{j_1,\dots,j_n\}\subset J$ and all possible values x_k of X_{j_k}.

Proposition 2.11. *A sequence of discrete random variables, $\{X_j\}_{j\in J}$, is independent iff*

$$\mathbb{E}[f_1(X_{j_1})\dots f_n(X_{j_n})]=\mathbb{E}[f_1(X_{j_1})]\dots\mathbb{E}[f_n(X_{j_n})] \tag{2.3}$$

for all choices of $\{j_1,\dots,j_n\}\subset J$ and all choice of bounded (or non-negative) functions, $f_1,\dots,f_n$. Here n is arbitrary.

Proof. ($\Longrightarrow$) If $\{X_j\}_{j\in J}$, are independent then

$$\begin{aligned}\mathbb{E}[f(X_{j_1},\dots,X_{j_n})]&=\sum_{x_1,\dots,x_n}f(x_1,\dots,x_n)\,\mathbb{P}(X_{j_1}=x_1,\dots,X_{j_n}=x_n)\\&=\sum_{x_1,\dots,x_n}f(x_1,\dots,x_n)\,\mathbb{P}(X_{j_1}=x_1)\cdots\mathbb{P}(X_{j_n}=x_n).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[f_1(X_{j_1})\dots f_n(X_{j_n})]&=\sum_{x_1,\dots,x_n}f_1(x_1)\dots f_n(x_n)\,\mathbb{P}(X_{j_1}=x_1)\cdots\mathbb{P}(X_{j_n}=x_n)\\&=\left(\sum_{x_1}f_1(x_1)\,\mathbb{P}(X_{j_1}=x_1)\right)\cdots\left(\sum_{x_n}f(x_n)\,\mathbb{P}(X_{j_n}=x_n)\right)\\&=\mathbb{E}[f_1(X_{j_1})]\dots\mathbb{E}[f_n(X_{j_n})].\end{aligned}$$

($\Longleftarrow$) Now suppose that Eq. (2.3) holds. If $f_j:=\delta_{x_j}$ for all j, then

$$\mathbb{E}[f_1(X_{j_1})\dots f_n(X_{j_n})]=\mathbb{E}[\delta_{x_1}(X_{j_1})\dots\delta_{x_n}(X_{j_n})]=\mathbb{P}(X_{j_1}=x_1,\dots,X_{j_n}=x_n)$$

while

$$\mathbb{E}[f_k(X_{j_k})]=\mathbb{E}[\delta_{x_k}(X_{j_k})]=\mathbb{P}(X_{j_k}=x_k).$$

Therefore it follows from Eq. (2.3) that Eq. (2.2) holds, i.e. $\{X_j\}_{j\in J}$ is an independent collection of random variables. ■

Using this as motivation we make the following definition.

Definition 2.12. *A collection of arbitrary random variables, $\{X_j\}_{j\in J}$, are **independent** iff*

$$\mathbb{E}[f_1(X_{j_1})\dots f_n(X_{j_n})]=\mathbb{E}[f_1(X_{j_1})]\dots\mathbb{E}[f_n(X_{j_n})]$$

for all choices of $\{j_1,\dots,j_n\}\subset J$ and all choice of bounded (or non-negative) functions, $f_1,\dots,f_n$.

Fact 2.13 *To check independence of a collection of real valued random variables, $\{X_j\}_{j\in J}$, it suffices to show*

$$\mathbb{P}(X_{j_1}\le t_1,\dots,X_{j_n}\le t_n)=\mathbb{P}(X_{j_1}\le t_1)\dots\mathbb{P}(X_{j_n}\le t_n)$$

for all possible choices of $\{j_1,\dots,j_n\}\subset J$ and all possible $t_k\in\mathbb{R}$. Moreover, one can replace $\le$ by $<$ or reverse these inequalities in the the above expression.

Theorem 2.14 (Groupings of independent RVs). *If $\{X_j\}_{j\in J}$, are **independent** random variables and J_0, J_1 are finite disjoint subsets in J, then*

$$\mathbb{E}\left[f_0\left(\{X_j\}_{j\in J_0}\right)\cdot f_1\left(\{X_j\}_{j\in J_1}\right)\right]=\mathbb{E}\left[f_0\left(\{X_j\}_{j\in J_0}\right)\right]\cdot\mathbb{E}\left[f_1\left(\{X_j\}_{j\in J_1}\right)\right]. \tag{2.4}$$

This holds more generally for any $\{J_k\}_{k=0}^n\subset J$ with $J_k\cap J_l=\emptyset$ and $\#(J_k)<\infty$.

***In words;** disjoint groupings of independent random variables are still independent random vectors.*

Proof. Discrete case example. Suppose $\{X_1,\dots,X_5\}$ are independent discrete random variables. Then

$$\begin{aligned}&\mathbb{P}(X_1=s_1,X_2=s_2,X_3=s_3,X_4=s_4,X_5=s_5)\\&\quad=\mathbb{P}(X_1=s_1)\,\mathbb{P}(X_2=s_2)\,\mathbb{P}(X_3=s_3)\,\mathbb{P}(X_4=s_4)\,\mathbb{P}(X_5=s_5)\\&\quad=\mathbb{P}(X_1=s_1,X_2=s_2)\,\mathbb{P}(X_3=s_3,X_4=s_4,X_5=s_5)\\&\quad=:\rho_{1,2}(s_1,s_2)\cdot\rho_{3,4,5}(s_3,s_4,s_5)\end{aligned}$$

and therefore,

$$\begin{aligned}&\mathbb{E}[f(X_1,X_2)\,g(X_3,X_4,X_5)]\\&\quad=\sum_{\mathbf{s}=(s_1,\dots,s_5)}f(s_1,s_2)\,g(s_3,s_4,s_5)\,\mathbb{P}(X_1=s_1,\dots,X_5=s_5)\\&\quad=\sum_{\mathbf{s}=(s_1,\dots,s_5)}f(s_1,s_2)\,g(s_3,s_4,s_5)\cdot\rho_{1,2}(s_1,s_2)\cdot\rho_{3,4,5}(s_3,s_4,s_5)\\&\quad=\sum_{\mathbf{s}=(s_1,s_2)}f(s_1,s_2)\,\rho_{1,2}(s_1,s_2)\cdot\sum_{\mathbf{s}=(s_3,s_4,s_5)}g(s_3,s_4,s_5)\,\rho_{3,4,5}(s_3,s_4,s_5)\\&\quad=\mathbb{E}[f(X_1,X_2)]\cdot\mathbb{E}[g(X_3,X_4,X_5)].\end{aligned}$$

General Case. Equation (2.4) is easy to verify when f_0 and f_1 are themselves product functions. The general result is then deduced from this observation along with measure theoretic arguments which go under the name of Dynkin's multiplicative systems theorem. ■

Proposition 2.15 (Disintegration I.). *Suppose that X is an $\mathbb{R}^k$ – valued random variable, Y is an $\mathbb{R}^l$ – valued random variable independent of X, and $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}_+$ then (assuming X and Y have continuous distributions $\rho_X(x)$ and $\rho_Y(y)$ respectively),*

$$\mathbb{E}[f(X,Y)] = \int_{\mathbb{R}^k} \mathbb{E}[f(x,Y)]\,\rho_X(x)\,dx \text{ and}$$
$$\mathbb{E}[f(X,Y)] = \int_{\mathbb{R}^l} \mathbb{E}[f(X,y)]\,\rho_Y(y)\,dy.$$

Proof. It is a fact that independence implies that the joint probability distribution, $\rho_{(X,Y)}(x,y)$, for (X,Y) must be given by

$$\rho_{(X,Y)}(x,y) = \rho_X(x)\,\rho_Y(y).$$

Therefore,

$$\begin{aligned}\mathbb{E}[f(X,Y)] &= \int_{\mathbb{R}^k\times\mathbb{R}^l} f(x,y)\,\rho_X(x)\,\rho_Y(y)\,dxdy \\ &= \int_{\mathbb{R}^k}\left[\int_{\mathbb{R}^l} dy f(x,y)\,\rho_Y(y)\right]\rho_X(x)\,dx \\ &= \int_{\mathbb{R}^k} \mathbb{E}[f(x,Y)]\,\rho_X(x)\,dx.\end{aligned}$$

■

One of the key theorems involving independent random variables is the strong law of large numbers. The other is the central limit theorem.

Theorem 2.16 (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 $\mathbb{E}X_n = \mu$.*

Remark 2.17. If $\mathbb{E}|X_1| = \infty$ but $\mathbb{E}X_1^- < \infty$, then $\frac{1}{n}S_n \to \infty$ a.s. To prove this, for $M > 0$ let

$$X_n^M := \min(X_n, M) = \begin{cases} X_n \text{ if } X_n \le M \\ M \text{ if } X_n \ge M\end{cases}$$

and $S_n^M := \sum_{i=1}^n X_i^M$. It follows from Theorem 2.16 that $\frac{1}{n}S_n^M \to \mu^M := \mathbb{E}X_1^M$ a.s.. Since $S_n \ge S_n^M$, we may conclude that

$$\liminf_{n\to\infty}\frac{S_n}{n} \ge \liminf_{n\to\infty}\frac{1}{n}S_n^M = \mu^M \text{ a.s.}$$

Since $\mu^M \to \infty$ as $M \to \infty$, it follows that $\liminf_{n\to\infty}\frac{S_n}{n} = \infty$ a.s. and hence that $\lim_{n\to\infty}\frac{S_n}{n} = \infty$ a.s.

Here is a crude special case of Theorem 2.16 which however does come with a rate estimate. We will do considerably better later in Corollary 13.59.

Proposition 2.18. *Let $k \in \mathbb{N}$ with $k \ge 2$ and $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables with $\mathbb{E}X_n = 0$ and $\mathbb{E}X_n^{2k} < \infty$. Then for every $p > \frac{1}{2} + \frac{1}{2k}$,*

$$\lim_{n\to\infty}\frac{S_n}{n^p} = 0 \text{ a.s.}$$

In other words for any $\varepsilon > 0$ small we have

$$\frac{S_n}{n} = \frac{S_n}{n^p}\frac{1}{n^{1-p}} = O\left(\frac{1}{n^{1-p}}\right) = O\left(\frac{1}{n^{\frac{1}{2}\left(1-\frac{1}{k}-\varepsilon\right)}}\right).$$

Proof. We start with the identity,

$$\mathbb{E}\left[\frac{1}{n}S_n\right]^{2k} = \frac{1}{n^{2k}}\sum_{j_1,\dots j_{2k}=1}^n \mathbb{E}[X_{j_1}\dots X_{j_{2k}}].$$

Using $\mathbb{E}[X_{j_1}\dots X_{j_{2k}}] = 0$ if there is any one index, j_l, distinct from the others, we conclude that the above sum can contain at most $C_k n^k$ non-zero terms for some $C_k < \infty$ and all of these terms are bounded by a constant C depending on $\mathbb{E}X_n^{2k}$. For example if $k = 2$ we have $\mathbb{E}[X_{j_1}X_{j_2}X_{j_3}X_{j_4}] = 0$ unless $j_1 = j_2 = j_3 = j_4$ (of which there are n such terms) or $j_1 = j_2$ and $j_3 = j_4$ (or similar with permuted indices) of which there are $3n^2$ – terms.

From the previous observations it follows that

$$\mathbb{E}\left[\frac{1}{n}S_n\right]^{2k} \le \frac{Cn^k}{n^{2k}} = C\frac{1}{n^k}.$$

Therefore if $0 < \alpha < 1$, then

$$\begin{aligned}\mathbb{E}\left(\sum_{n=1}^\infty\left[n^\alpha\frac{1}{n}S_n\right]^{2k}\right) &= \sum_{n=1}^\infty \mathbb{E}\left[n^\alpha\frac{1}{n}S_n\right]^{2k} \\ &\le \sum_{n=1}^\infty C\frac{1}{n^k}n^{\alpha 2k} = \sum_{n=1}^\infty C\frac{1}{n^{k(1-2\alpha)}} < \infty\end{aligned}$$

provided $k(1-2\alpha) > 1$, i.e. $1 - 2\alpha > \frac{1}{k}$, i.e. $\alpha < \frac{1}{2}\left(1 - \frac{1}{k}\right)$. For such an α we have

$$\sum_{n=1}^{\infty}\left[n^{\alpha}\frac{1}{n}S_n\right]^{2k}<\infty \text{ a.s.} \quad\Longrightarrow\quad \lim_{n\to\infty}\frac{1}{n^{1-\alpha}}S_n=0 \text{ a.s..}$$

Tracing through the inequalities shows $p:=1-\alpha>1-\frac{1}{2}\left(1-\frac{1}{k}\right)=\frac{1}{2}+\frac{1}{2k}$ is the required restriction on p. ■

Often times for practical importance, the following weak law of large numbers is in fact more useful. For the proof we will need the following simple but very useful inequality.

Lemma 2.19 (Chebyshev's Inequality). *If X is a random variable, $\delta>0$, and $p>0$, then*

$$P\left(\{|X|\geq\delta\}\right)=\mathbb{E}\left[1_{|X|\geq\delta}\right]\leq\mathbb{E}\left[\frac{|X|^p}{\delta^p}1_{|X|\geq\delta}\right]\leq\delta^{-p}\mathbb{E}\left|X\right|^p. \tag{2.5}$$

Proof. Taking expectations of the following pointwise inequalities,

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

immediately gives Eq. (2.5). ■

Theorem 2.20. *Let $\{X_n\}_{n=1}^{\infty}$ be uncorrelated random square integrable random variables, then*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^{n}\left(X_m-\mathbb{E}X_m\right)\right|\geq\delta\right)\leq\frac{1}{\delta^2n^2}\sum_{m=1}^{n}\operatorname{Var}\left(X_m\right).$$

If we further assume that $\mathbb{E}X_m=\mu$ and $\operatorname{Var}\left(X_m\right)=\sigma^2$ are independent of m, then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^{n}X_m-\mu\right|\geq\delta\right)\leq\frac{\sigma^2}{\delta^2}\frac{1}{n}.$$

Proof. By Chebyshev's inequality and the assumption that $\operatorname{Cov}\left(X_m,X_k\right)=\delta_{mk}\operatorname{Var}\left(X_m\right)$, we find

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^{n}\left(X_m-\mathbb{E}X_m\right)\right|\geq\delta\right)\\
&\quad\leq\frac{1}{\delta^2}\mathbb{E}\left|\frac{1}{n}\sum_{m=1}^{n}\left(X_m-\mathbb{E}X_m\right)\right|^2\\
&\quad=\frac{1}{\delta^2n^2}\mathbb{E}\sum_{m.k=1}^{n}\left[\left(X_m-\mathbb{E}X_m\right)\left(X_k-\mathbb{E}X_k\right)\right]\\
&\quad=\frac{1}{\delta^2n^2}\sum_{m.k=1}^{n}\operatorname{Cov}\left(X_m,X_k\right)=\frac{1}{\delta^2n^2}\sum_{m=1}^{n}\operatorname{Var}\left(X_m\right).
\end{aligned}$$

■

3

Conditional Expectation

3.1 σ – algebras (partial information)

Definition 3.1 (σ - algebra of X). *Let Ω – be a sample space, W be a set, and $X : \Omega \to W$ be a function. The X – **measurable events in** Ω, $\sigma(X)$, are those events of the form,*

$$\sigma(X) := \left\{A := \{X \in B\} = X^{-1}(B) : B \subset W\right\}.$$

Put another way, $A \in \sigma(X)$ iff $A = \{X \in B\}$ for some $B \subset W$.

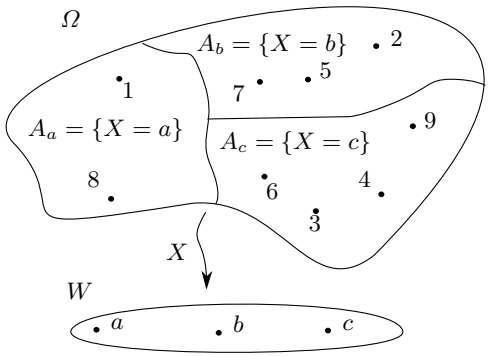


Fig. 3.1. In this figure $\Omega = \{1, \ldots, 9\}$, $A_a = \{1, 8\}$, $A_b = \{2, 5, 7\}$, $A_c = \{3, 4, 6, 9\}$, and $X : \Omega \to S = \{a, b, c\}$ is the function that satisfies, $X = x$ on A_x for $x \in S$. With this notation we find $A \in \sigma(X)$ iff $A = \emptyset$ or A is a union of elements from the list of events, $\{A_a, A_b, A_c\}$.

Lemma 3.2. *An event $A \subset \Omega$ is in $\sigma(X)$ iff there exists a function, $f : W \to \mathbb{R}$, such that $f(X) = 1_A$. [The function f is unique iff $X(\Omega) = W$.]*

Proof. If $A \in \sigma(X)$, then $A = X^{-1}(B)$ for some $B \subset W$ and in this case,

$$1_A = 1_{X^{-1}(B)} = 1_B(X).$$

Conversely if $f(X) = 1_A$, then let $B := \{w \in W : f(w) = 1\}$. We then have,

$$\omega \in X^{-1}(B) \iff X(\omega) \in B \iff 1 = f(X(\omega)) = 1_A(\omega) \iff \omega \in A$$

which show $A = X^{-1}(B) \in \sigma(X)$. ■

Remark 3.3. Notice that if $A_i \in \sigma(X)$ for $i \in \mathbb{N}$ then A_1^c, $\cup_{i=1}^\infty A_i$, $\cap_{i=1}^\infty A_i$, and $A_2 \setminus A_1 = A_2 \cap A_1^c$ are in $\sigma(X)$. In other words, $\sigma(X)$ is stable under all of the usual set operations. Also notice that $\emptyset, \Omega \in \sigma(X)$.

The reader should interpret $\sigma(X)$ as the events in Ω which can be distinguished (measured) by knowing X. Now let S be another set and $Y : \Omega \to S$ be another function.

Definition 3.4 (X – measurable). *We say Y is X – **measurable** iff $\sigma(Y) \subset \sigma(X)$, i.e. the events that can be measure by Y can also be measured by X.*

For our purposes you can forget the previous definition and use the criteria for X – measurability given in the following proposition.

Proposition 3.5 (X – measurability). *Let $Y : \Omega \to S$ and $X : \Omega \to W$ be functions. Then Y is X – measurable iff there exists $f : W \to S$ such that $Y = f(X)$, i.e. $Y(\omega)$ is completely determined by knowing $X(\omega)$.*

Proof. If $Y = f(X)$ for some function, $f : W \to S$, then for $A \subset S$ we have

$$\{Y \in A\} = \{f(X) \in A\} = \left\{X \in f^{-1}(A)\right\} \in \sigma(X)$$

from which it follows that $\sigma(Y) \subset \sigma(X)$. You are asked to prove the converse direction in Exercise 3.1. ■

Exercise 3.1 (Optional). Let $W_0 := X(\Omega) \subset W$. Finish the proof of Proposition 3.5 using the following outline;

1. Use the fact that $\sigma(Y) \subset \sigma(X)$ to show for each $s \in S$ there exists $B_s \subset W_0 \subset W$ such that $\{Y = s\} = \{X \in B_s\}$.
2. Show $B_s \cap B_{s'} = \emptyset$ for all $s, s' \in S$ with $s \neq s'$.
3. Show $X(\Omega) = W_0 := \cup_{s \in S} B_s$.
   Now fix a point $s_* \in S$ and then define, $f : W \to S$ by setting $f(w) = s_*$ when $w \in W \setminus W_0$ and $f(w) = s$ when $w \in B_s \subset W_0$.
4. Verify that $Y = f(X)$.

3.2 Theory of Conditional Expectation

Let us now fix a probability, $\mathbb{P}$, on Ω for the rest of this subsection.

Notation 3.6 (Conditional Expectation 1) *Given $Y \in L^1(\mathbb{P})$ and $A \subset \Omega$ let*

$$\mathbb{E}[Y:A] := \mathbb{E}[1_A Y]$$

and

$$\mathbb{E}[Y|A] = \begin{cases} \mathbb{E}[Y:A]/\mathbb{P}(A) & \text{if } \mathbb{P}(A) > 0 \\ 0 & \text{if } \mathbb{P}(A) = 0 \end{cases}. \tag{3.1}$$

(In point of fact, when $\mathbb{P}(A) = 0$ we could set $\mathbb{E}[Y|A]$ to be any real number. We choose 0 for definiteness and so that $Y \to \mathbb{E}[Y|A]$ is always linear.)

Example 3.7 (Conditioning for the uniform distribution). Suppose that Ω is a finite set and $\mathbb{P}$ is the uniform distribution on $\mathbb{P}$ so that $\mathbb{P}(\{\omega\}) = \frac{1}{\#(\Omega)}$ for all $\omega \in W$. Then for non-empty any subset $A \subset \Omega$ and $Y : \Omega \to \mathbb{R}$ we have $\mathbb{E}[Y|A]$ is the expectation of Y restricted to A under the uniform distribution on A. Indeed,

$$\begin{aligned}\mathbb{E}[Y|A] &= \frac{1}{\mathbb{P}(A)}\mathbb{E}[Y:A] = \frac{1}{\mathbb{P}(A)}\sum_{\omega\in A} Y(\omega)\,\mathbb{P}(\{\omega\}) \\ &= \frac{1}{\#(A)/\#(\Omega)}\sum_{\omega\in A} Y(\omega)\frac{1}{\#(\Omega)} = \frac{1}{\#(A)}\sum_{\omega\in A} Y(\omega).\end{aligned}$$

Theorem 3.8 (Conditional Expectation via Best Approximations). *Suppose $X : \Omega \to W$ as above and $Y \in L^2(\mathbb{P})$, i.e. $Y : \Omega \to \mathbb{C}$ such that $\mathbb{E}|Y|^2 < \infty$. Then there exists an "essentially unique" function $h : W \to \mathbb{C}$ such that $h(X) \in L^2(\mathbb{P})$ which satisfies,*[1]

$$\mathbb{E}\left[|Y - h(X)|^2\right] \le \mathbb{E}|Y - f(X)|^2 \tag{3.2}$$

for all function $f : W \to \mathbb{C}$ such that $f(X) \in L^2(\mathbb{P})$. The function h may alternatively be determined by requiring,

$$\mathbb{E}[[Y - h(X)]f(X)] = 0 \;\forall\; f : W \to \mathbb{C} \ni \mathbb{E}|f(X)|^2 < \infty. \tag{3.3}$$

Proof. The existence of such an h satisfying Eq. (3.2) is a consequence of the orthogonal projection theorem in Hilbert spaces. We will simply take this result for granted. However, let us show the conditions in Eq. (3.2) and Eq. (3.3) are equivalent.

[1] Note that h depends on X, Y, and $\mathbb{P}$. However, Theorem 3.16 below asserts that h can be determined uniquely by only knowing the "joint distribution" of (X, Y).

Eq. (3.2) $\implies$ Eq. (3.3): If Eq. (3.2) holds then for any $f : W \to \mathbb{R}$ such that $f(X) \in L^2(\mathbb{P})$ and $t \in \mathbb{R}$ we have,

$$\begin{aligned}\varphi(t) &:= \mathbb{E}\left[|Y - [h(X) + tg(X)]|^2\right] \\ &= \mathbb{E}|Y - h(X)|^2 + 2t\mathbb{E}([Y - h(X)]f(X)) + t^2\mathbb{E}|f(X)|^2\end{aligned}$$

has a minimum at $t = 0$. So by the first derivative test it follows that

$$0 = \dot{\varphi}(0) = 2\mathbb{E}([Y - h(X)]f(X))$$

which shows Eq. (3.3) holds.

Eq. (3.3) $\implies$ Eq. (3.2): Assuming Eq. (3.3), it follows that

$$\begin{aligned}\mathbb{E}|Y - f(X)|^2 &= \mathbb{E}|Y - h(X) + (h - f)(X)|^2 \\ &= \mathbb{E}|Y - h(X)|^2 + 2\mathbb{E}([Y - h(X)](h - f)(X)) + \mathbb{E}|(h - f)(X)|^2 \\ &= \mathbb{E}|Y - h(X)|^2 + \mathbb{E}|(h - f)(X)|^2 \ge \mathbb{E}|Y - h(X)|^2.\end{aligned}$$

■

It should be noted that the function, h, in Theorem 3.8 depends on both X and Y and therefore $\mathbb{E}[Y|X] = h(X)$ also depends on both X and Y. However as explained in Theorem 3.16 below, the function h is only depends on (X, Y) though their "joint distribution."

Definition 3.9 (Conditional Expectation I). *We refer to the function $h(X)$ in Theorem 3.8 as the* ***conditional expectation*** *of Y given X (or $\sigma(X)$) and denote the result by $\mathbb{E}[Y|\sigma(X)]$ or by $\mathbb{E}[Y|X]$.*

Proposition 3.10 (Discrete formula). *Suppose that $X : \Omega \to W$ has finite or countable range, i.e. $X(\Omega) = \{x_i\}_{i=1}^N$ where $N \in \mathbb{N} \cup \{\infty\}$. In this case,*

$$\mathbb{E}[Y|X] = h(X) \;\text{ where } h(x) = \mathbb{E}_{\mathbb{P}}[Y|X = x], \tag{3.4}$$

where $\mathbb{E}_{\mathbb{P}}[Y|X = x]$ is as in Notation 3.6.

Proof. If $\mathbb{P}(X = x) = 0$ then $\mathbb{P}(\omega) = 0$ for all $\omega \in \{X = x\}$ and Eq. (3.5) holds no matter the value of $h(x)$. For definiteness, we choose $h(x) = 0$ in this case.

We are looking to find $h : X(\Omega) \subset W \to \mathbb{R}$ satisfying,

$$\begin{aligned}0 = \mathbb{E}[[Y - h(X)]f(X)] &= \sum_{x\in X(\Omega)} \mathbb{E}[[Y - h(X)]f(X) : X = x] \\ &= \sum_{x\in X(\Omega)} \mathbb{E}[[Y - h(x)]f(x) : X = x] \\ &= \sum_{x\in X(\Omega)} (\mathbb{E}[Y : X = x] - h(x)\mathbb{P}(X = x))f(x)\end{aligned}$$

for all f on $X(\Omega) \subset W$. This implies that we require,

$$0 = \mathbb{E}[Y : X = x] - h(x)\mathbb{P}(X = x) \text{ for all } x \in X(\Omega). \tag{3.5}$$

In other word,

$$h(x) = \mathbb{E}_{\mathbb{P}}[Y \cdot 1_{X=x}] / \mathbb{P}(X = x) = \mathbb{E}_{\mathbb{P}}[Y|X = x].$$

If $\mathbb{P}(X = x) = 0$ Eq. (3.5) holds no matter the value of $h(x)$. For definiteness, we choose $h(x) = 0$ in this case. ■

Remark 3.11. If we have already worked out the probability measure, $\mathbb{P}(\cdot|X = x)$, then $h(x)$ in Eq. (3.4) may be computed using

$$h(x) = \mathbb{E}_{\mathbb{P}}[Y|X = x] = \mathbb{E}_{\mathbb{P}(\cdot|X=x)}Y.$$

The point is that for any random variable, $Z : \Omega \to \mathbb{R}$ we always have

$$\mathbb{E}_{\mathbb{P}}[Z|X = x] = \mathbb{E}_{\mathbb{P}(\cdot|X=x)}Z. \tag{3.6}$$

For example if $Z = 1_B$, then

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[1_B|X = x] &= \frac{1}{\mathbb{P}(X = x)}\mathbb{E}_{\mathbb{P}}[1_B \cdot 1_{X=x}] \\ &= \frac{1}{\mathbb{P}(X = x)}\mathbb{P}(B \cap \{X = x\}) \\ &= \mathbb{P}(B|X = x) = \mathbb{E}_{\mathbb{P}(\cdot|X=x)}1_B. \end{aligned} \tag{3.7}$$

For general Z let Z_ε be the approximation in Eq. (1.1). From Eq. (3.7) and linearity if follows that

$$\mathbb{E}_{\mathbb{P}}[Z_\varepsilon|X = x] = \mathbb{E}_{\mathbb{P}(\cdot|X=x)}Z_\varepsilon$$

and then letting $\varepsilon \downarrow 0$ proves Eq. (3.6).

Let us pause for a moment to record a few basic general properties of conditional expectations.

Proposition 3.12 (Contraction Property). *For all $Y \in L^2(\mathbb{P})$, we have $\mathbb{E}|\mathbb{E}[Y|X]| \le \mathbb{E}|Y|$. Moreover if $Y \ge 0$ then $\mathbb{E}[Y|X] \ge 0$ (a.s.).*

Proof. Let $\mathbb{E}[Y|X] = h(X)$ (with $h : S \to \mathbb{R}$) and then define

$$f(x) = \begin{cases} 1 & \text{if } h(x) \ge 0 \\ -1 & \text{if } h(x) < 0 \end{cases}.$$

Since $h(x)f(x) = |h(x)|$, it follows from Eq. (3.3) that

$$\mathbb{E}[|h(X)|] = \mathbb{E}[Yf(X)] = |\mathbb{E}[Yf(X)]| \le \mathbb{E}[|Yf(X)|] = \mathbb{E}|Y|.$$

For the second assertion take $f(x) = 1_{h(x)<0}$ in Eq. (3.3) in order to learn

$$\mathbb{E}\left[h(X)1_{h(X)<0}\right] = \mathbb{E}\left[Y1_{h(X)<0}\right] \ge 0.$$

As $h(X)1_{h(X)<0} \le 0$ we may conclude that $h(X)1_{h(X)<0} = 0$ a.s. ■

Because of this proposition we may extend the notion of conditional expectation to $Y \in L^1(\mathbb{P})$ as stated in the following theorem which we do not bother to prove here.

Theorem 3.13. *Given $X : \Omega \to W$ and $Y \in L^1(\mathbb{P})$, there exists an "essentially unique" function $h : W \to \mathbb{R}$ such that Eq. (3.3) holds for all bounded functions, $f : W \to \mathbb{R}$. (As above we write $\mathbb{E}[Y|X]$ for $h(X)$.) Moreover the contraction property, $\mathbb{E}|\mathbb{E}[Y|X]| \le \mathbb{E}|Y|$, still holds.*

Definition 3.14 (Conditional Expectation). *If $Y \in L^1(\mathbb{P})$, we let $\mathbb{E}[Y|X] = h(X)$ be the a.s. unique random variable which is $\sigma(X)$ measurable and satisfies,*

$$\mathbb{E}(Yf(X)) = \mathbb{E}(h(X)f(X)) = \mathbb{E}(\mathbb{E}[Y|X]f(X))$$

for all bounded f.

Theorem 3.15 (Basic properties). *Let Y, Y_1, and Y_2 be integrable random variables and $X : \Omega \to W$ be given. Then:*

1. $\mathbb{E}(Y_1 + Y_2|X) = \mathbb{E}(Y_1|X) + \mathbb{E}(Y_2|X)$.
2. $\mathbb{E}(aY|X) = a\mathbb{E}(Y|X)$ *for all constants a.*
3. $\mathbb{E}(g(X)Y|X) = g(X)\mathbb{E}(Y|X)$ *for all bounded functions g.*
4. $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$. ***(Law of total expectation.)***
5. *If Y and X are independent then $\mathbb{E}(Y|X) = \mathbb{E}Y$.*
6. *If Y is $\sigma(X)$ measurable, then $\mathbb{E}[Y|X] = Y$.*

Proof. 1. Let $h_i(X) = \mathbb{E}[Y_i|X]$, then for all bounded f,

$$\begin{aligned}\mathbb{E}[Y_1 f(X)] &= \mathbb{E}[h_1(X)f(X)] \text{ and} \\ \mathbb{E}[Y_2 f(X)] &= \mathbb{E}[h_2(X)f(X)]\end{aligned}$$

and therefore adding these two equations together implies

$$\begin{aligned}\mathbb{E}[(Y_1 + Y_2)f(X)] &= \mathbb{E}[(h_1(X) + h_2(X))f(X)] \\ &= \mathbb{E}[(h_1 + h_2)(X)f(X)] \\ \mathbb{E}[Y_2 f(X)] &= \mathbb{E}[h_2(X)f(X)]\end{aligned}$$

for all bounded f. Therefore we may conclude that

$$\mathbb{E}(Y_1+Y_2|X)=(h_1+h_2)(X)=h_1(X)+h_2(X)=\mathbb{E}(Y_1|X)+\mathbb{E}(Y_2|X).$$

2. The proof is similar to 1 but easier and so is omitted.

3. Let $h(X)=\mathbb{E}[Y|X]$, then $\mathbb{E}[Yf(X)]=\mathbb{E}[h(X)f(X)]$ for all bounded functions f. Replacing f by $g\cdot f$ implies

$$\mathbb{E}[Yg(X)f(X)]=\mathbb{E}[h(X)g(X)f(X)]=\mathbb{E}[(h\cdot g)(X)f(X)]$$

for all bounded functions f. Therefore we may conclude that

$$\mathbb{E}[Yg(X)|X]=(h\cdot g)(X)=h(X)g(X)=g(X)\mathbb{E}(Y|X).$$

4. Take $f\equiv 1$ in Eq. (3.3).

5. If X and Y are independent and $\mu:=\mathbb{E}[Y]$, then

$$\mathbb{E}[Yf(X)]=\mathbb{E}[Y]\mathbb{E}[f(X)]=\mu\mathbb{E}[f(X)]=\mathbb{E}[\mu f(X)]$$

from which it follows that $\mathbb{E}[Y|X]=\mu$ as desired.

6. By assumption $Y=f(X)$ for some function f and hence the best approximation to Y by a function of X is clearly $Y=f(X)$ itself. More formally,

$$\mathbb{E}[Yg(X)]=\mathbb{E}[f(X)g(X)]\ \text{ for all bounded } g$$

and hence $\mathbb{E}[Y|X]=f(X)=Y$. ■

Exercise 3.2. Suppose that X and Y are two integrable random variables such that

$$\mathbb{E}[X|Y]=18-\frac{3}{5}Y \text{ and } \mathbb{E}[Y|X]=10-\frac{1}{3}X.$$

Find $\mathbb{E}X$ and $\mathbb{E}Y$.

The next theorem says that conditional expectations essentially only depends on the distribution of (X,Y) and nothing else.

Theorem 3.16 (Dependence only on distributions). *Suppose that (X,Y) and $\left(\tilde{X},\tilde{Y}\right)$ are random vectors such that $(X,Y)\overset{d}{=}\left(\tilde{X},\tilde{Y}\right)$, i.e. $\mathbb{E}[G(X,Y)]=\mathbb{E}\left[G\left(\tilde{X},\tilde{Y}\right)\right]$ for all bounded (or non-negative) functions G. If $h(X)=\mathbb{E}[u(X,Y)|X]$, then $\mathbb{E}\left[u\left(\tilde{X},\tilde{Y}\right)|\tilde{X}\right]=h\left(\tilde{X}\right)$.*

Proof. By assumption we know that

$$\mathbb{E}[u(X,Y)f(X)]=\mathbb{E}[h(X)f(X)]\ \text{ for all bounded } f.$$

Since $(X,Y)\overset{d}{=}\left(\tilde{X},\tilde{Y}\right)$, this is equivalent to

$$\mathbb{E}\left[u\left(\tilde{X},\tilde{Y}\right)f\left(\tilde{X}\right)\right]=\mathbb{E}\left[h\left(\tilde{X}\right)f\left(\tilde{X}\right)\right]\ \text{ for all bounded } f$$

which is equivalent to $\mathbb{E}\left[u\left(\tilde{X},\tilde{Y}\right)|\tilde{X}\right]=h\left(\tilde{X}\right)$. ■

Exercise 3.3. Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with $\mathbb{E}|X_i|<\infty$ for all i and let $S_m:=X_1+\cdots+X_m$ for $m=1,2,\ldots$. Show

$$\mathbb{E}[S_m|S_n]=\frac{m}{n}S_n \text{ for all } m\le n.$$

Hint: observe by symmetry[2] that there is a function $h:\mathbb{R}\to\mathbb{R}$ such that

$$\mathbb{E}(X_i|S_n)=h(S_n)\ \text{ independent of } i.$$

Remark 3.17. If $m>n$, in Exercise 3.3, then $S_m=S_n+X_{n+1}+\cdots+X_m$. Since X_i is independent of S_n for $i>n$ (see Theorem 2.14) it follows from item 5. of Theorem 3.15 that $\mathbb{E}[X_i|S_n]=\mu:=\mathbb{E}X_i$. This result along with the basic properties of conditional expectation shows,

$$\begin{aligned}\mathbb{E}(S_m|S_n)&=\mathbb{E}(S_n+X_{n+1}+\cdots+X_m|S_n)\\&=\mathbb{E}(S_n|S_n)+\mathbb{E}(X_{n+1}|S_n)+\cdots+\mathbb{E}(X_m|S_n)\\&=S_n+(m-n)\mu \text{ if } m\ge n.\end{aligned}$$

So in general,

$$\mathbb{E}[S_m|S_n]=\begin{cases}\frac{m}{n}S_n & \text{if } m\le n\\ S_n+(m-n)\mu & \text{if } m\ge n.\end{cases}$$

item 5. of Theorem 3.15 is a rather special case the following useful tool for computing conditional expectations.

Proposition 3.18 (Disintegration II.). *Suppose that $X:\Omega\to W$ and $Y:\Omega\to W'$ are independent random functions and $U:W\times W'\to\mathbb{C}$ is a functions such that $\mathbb{E}|U(X,Y)|<\infty$. Under these assumptions,*

$$\mathbb{E}[U(X,Y)|X]=h(X)$$

where

$$h(x):=\mathbb{E}[U(x,Y)]\ \text{ for all } x\in W.$$

[2] Apply Theorem 3.16 using $(X_1,S_n)\overset{d}{=}(X_i,S_n)$ for $1\le i\le n$.

Proof. The theorem is true in general but requires measure theory in order to prove it in full generality. Here I will give the proof in the case that $X(\Omega)$ is a countable set, see Proposition 3.23 for a proof for certain continuous random functions.)

Let

$$h(x) := \mathbb{E}[U(X,Y)|X=x] = \mathbb{E}[U(x,Y)|X=x] = \mathbb{E}[U(x,Y)]$$

where in the last equality we have used the independence of X from Y which by definition means,

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)]$$

for all bounded functions, $f : W \to \mathbb{C}$ and $g : W \to \mathbb{C}$. The result now follows from Proposition 3.10 ■

Theorem 3.19 (Tower Property). *Let Y be an integrable random variable and $X_i : \Omega \to W_i$ be given functions for $i = 1, 2$. Then*

$$\mathbb{E}[\mathbb{E}[Y|(X_1,X_2)]|X_1] = \mathbb{E}[Y|X_1] \tag{3.8}$$

and

$$\mathbb{E}[\mathbb{E}[Y|X_1]|(X_1,X_2)] = \mathbb{E}[Y|X_1]. \tag{3.9}$$

Proof. If $g : W_1 \to \mathbb{R}$ is a bounded function then $g(X_1)$ is both $\sigma(X_1)$ and $\sigma(X_1,X_2)$ measurable and therefore,

$$\mathbb{E}[\mathbb{E}[Y|(X_1,X_2)] \cdot g(X_1)] = \mathbb{E}[Y \cdot g(X_1)] = \mathbb{E}[\mathbb{E}[Y|X_1] \cdot g(X_1)].$$

As this equation holds for all g we conclude that Eq. (3.8) holds.

$$\mathbb{E}[Y|X_1] = \mathbb{E}[\mathbb{E}[Y|(X_1,X_2)]|X_1].$$

For Eq. (3.9) we simply use item 6. of Theorem 3.15 and the fact that $\mathbb{E}[Y|X_1]$ is $\sigma(X_1) \subset \sigma(X_1,X_2)$ – measurable to conclude,

$$\mathbb{E}[\mathbb{E}[Y|X_1]|(X_1,X_2)] = \mathbb{E}[Y|X_1].$$

■

3.3 Conditional Expectation for Continuous Random Variables

(We will cover this section later in the course as needed.)

Suppose that Y and X are continuous random variables which have a joint density, $\rho_{(Y,X)}(y,x)$. Then by definition of $\rho_{(Y,X)}$, we have, for all bounded or non-negative, U, that

$$\mathbb{E}[U(Y,X)] = \int\int U(y,x)\rho_{(Y,X)}(y,x)\,dydx. \tag{3.10}$$

The marginal density associated to X is then given by

$$\rho_X(x) := \int \rho_{(Y,X)}(y,x)\,dy \tag{3.11}$$

and recall that the conditional density $\rho_{(Y|X)}(y,x)$ is defined by

$$\rho_{(Y|X)}(y,x) = \begin{cases} \frac{\rho_{(Y,X)}(y,x)}{\rho_X(x)} & \text{if } \rho_X(x) > 0 \\ 0 & \text{if } \rho_X(x) = 0 \end{cases}. \tag{3.12}$$

Observe that if $\rho_{(Y,X)}(y,x)$ is continuous, then

$$\rho_{(Y,X)}(y,x) = \rho_{(Y|X)}(y,x)\rho_X(x) \text{ for all } (x,y). \tag{3.13}$$

Indeed, if $\rho_X(x) = 0$, then

$$0 = \rho_X(x) = \int \rho_{(Y,X)}(y,x)\,dy$$

from which it follows that $\rho_{(Y,X)}(y,x) = 0$ for all y. If $\rho_{(Y,X)}$ is not continuous, Eq. (3.13) still holds for "a.e." (x,y) which is good enough.

Lemma 3.20. *In the notation above,*

$$\rho(x,y) = \rho_{(Y|X)}(y,x)\rho_X(x) \text{ for a.e. } (x,y). \tag{3.14}$$

Proof. By definition Eq. (3.14) holds when $\rho_X(x) > 0$ and $\rho(x,y) \geq \rho_{(Y|X)}(y,x)\rho_X(x)$ for all (x,y). Moreover,

$$\begin{aligned}\int\int \rho_{(Y|X)}(y,x)\rho_X(x)\,dxdy &= \int\int \rho_{(Y|X)}(y,x)\rho_X(x)1_{\rho_X(x)>0}dxdy \\ &= \int\int \rho(x,y)1_{\rho_X(x)>0}dxdy \\ &= \int \rho_X(x)1_{\rho_X(x)>0}dx = \int \rho_X(x)\,dx \\ &= 1 = \int\int \rho(x,y)\,dxdy,\end{aligned}$$

or equivalently,

$$\int\int \left[\rho(x,y) - \rho_{(Y|X)}(y,x)\rho_X(x)\right]dxdy = 0$$

which implies the result. ■

Theorem 3.21. *Keeping the notation above, for all or all bounded or non-negative, U, we have $\mathbb{E}[U(Y,X)|X] = h(X)$ where*

$$h(x) = \int U(y,x)\,\rho_{(Y|X)}(y,x)\,dy \tag{3.15}$$

$$= \begin{cases} \frac{\int U(y,x)\rho_{(Y,X)}(y,x)dy}{\int \rho_{(Y,X)}(y,x)dy} & \text{if} \quad \int \rho_{(Y,X)}(y,x)\,dy > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3.16}$$

In the future we will usually denote $h(x)$ informally by $\mathbb{E}[U(Y,x)|X=x]$,[3] so that

$$\mathbb{E}[U(Y,x)|X=x] := \int U(y,x)\,\rho_{(Y|X)}(y,x)\,dy. \tag{3.17}$$

Proof. We are looking for $h : S \to \mathbb{R}$ such that

$$\mathbb{E}[U(Y,X)\,f(X)] = \mathbb{E}[h(X)\,f(X)] \text{ for all bounded } f.$$

Using Lemma 3.20, we find

$$\begin{aligned}
\mathbb{E}[U(Y,X)\,f(X)] &= \int\int U(y,x)\,f(x)\,\rho_{(Y,X)}(y,x)\,dydx \\
&= \int\int U(y,x)\,f(x)\,\rho_{(Y|X)}(y,x)\,\rho_X(x)\,dydx \\
&= \int \left[\int U(y,x)\,\rho_{(Y|X)}(y,x)\,dy\right] f(x)\,\rho_X(x)\,dx \\
&= \int h(x)\,f(x)\,\rho_X(x)\,dx \\
&= \mathbb{E}[h(X)\,f(X)]
\end{aligned}$$

where h is given as in Eq. (3.15).

■

Example 3.22 (Durrett 8.15, p. 145). Suppose that X and Y have joint density $\rho(x,y) = 8xy \cdot 1_{0<y<x<1}$. We wish to compute $\mathbb{E}[u(X,Y)|Y]$. To this end we compute

$$\rho_Y(y) = \int_{\mathbb{R}} 8xy \cdot 1_{0<y<x<1}dx = 8y\int_{x=y}^{x=1} x \cdot dx = 8y \cdot \frac{x^2}{2}\Big|_y^1 = 4y \cdot \left(1-y^2\right).$$

Therefore,

$$\rho_{X|Y}(x,y) = \frac{\rho(x,y)}{\rho_Y(y)} = \frac{8xy \cdot 1_{0<y<x<1}}{4y \cdot (1-y^2)} = \frac{2x \cdot 1_{0<y<x<1}}{(1-y^2)}$$

and so

$$\mathbb{E}[u(X,Y)|Y=y] = \int_{\mathbb{R}} \frac{2x \cdot 1_{0<y<x<1}}{(1-y^2)} u(x,y)\,dx = 2\frac{1_{0<y<1}}{1-y^2}\int_y^1 u(x,y)\;xdx$$

and so

$$\mathbb{E}[u(X,Y)|Y] = 2\frac{1}{1-Y^2}\int_Y^1 u(x,Y)\,xdx.$$

is the best approximation to $u(X,Y)$ be a function of Y alone.

Proposition 3.23. *Suppose that X, Y are independent random functions, then*

$$\mathbb{E}[U(Y,X)|X] = h(X)$$

where

$$h(x) := \mathbb{E}[U(Y,x)].$$

Proof. I will prove this in the continuous distribution case and leave the discrete case to the reader. (The theorem is true in general but requires measure theory in order to prove it in full generality.) The independence assumption is equivalent to $\rho_{(Y,X)}(y,x) = \rho_Y(y)\,\rho_X(x)$. Therefore,

$$\rho_{(Y|X)}(y,x) = \begin{cases} \rho_Y(y) & \text{if } \rho_X(x) > 0 \\ 0 & \text{if } \rho_X(x) = 0 \end{cases}$$

and therefore $\mathbb{E}[U(Y,X)|X] = h_0(X)$ where

$$\begin{aligned}
h_0(x) &= \int U(y,x)\,\rho_{(Y|X)}(y,x)\,dy \\
&= 1_{\rho_X(x)>0}\int U(y,x)\,\rho_Y(y)\,dy = 1_{\rho_X(x)>0}\mathbb{E}[U(Y,x)] \\
&= 1_{\rho_X(x)>0}h(x).
\end{aligned}$$

If f is a bounded function of x, then

$$\begin{aligned}
\mathbb{E}[h_0(X)\,f(X)] &= \int h_0(x)\,f(x)\,\rho_X(x)\,dx = \int_{\{x:\rho_X(x)>0\}} h_0(x)\,f(x)\,\rho_X(x)\,dx \\
&= \int_{\{x:\rho_X(x)>0\}} h(x)\,f(x)\,\rho_X(x)\,dx = \int h(x)\,f(x)\,\rho_X(x)\,dx \\
&= \mathbb{E}[h(X)\,f(X)].
\end{aligned}$$

So for all practical purposes, $h(X) = h_0(X)$, i.e. $h(X) = h_0(X)$ – a.s. (Indeed, take $f(x) = \operatorname{sgn}(h(x) - h_0(x))$ in the above equation to learn that $\mathbb{E}|h(X) - h_0(X)| = 0$. ■

[3] **Warning:** this is **not** consistent with Eq. (3.1) as $P(X=x) = 0$ for continuous distributions.

Theorem 3.24 (Iterated conditioning 1). *Let X, Y, and Z be random vectors and suppose that (X,Y) is distributed according to $\rho_{(X,Y)}(x,y)\,dxdy$. Then*

$$\mathbb{E}[Z|Y=y] = \int \mathbb{E}[Z|Y=y, X=x]\,\rho_{Y|X}(y,x)\,dx \qquad \rho_Y(y)\,dy \text{ – a.s.}$$

Proof. Let $h(x,y) := \mathbb{E}[Z|Y=y, X=x]$ so that

$$\mathbb{E}[Zv(X,Y)] = \mathbb{E}[h(X,Y)\,v(X,Y)] \text{ for all } v.$$

Taking $v(x,y) = g(y)$ to be a function of Y alone shows,

$$\mathbb{E}[Zg(Y)] = \mathbb{E}[h(X,Y)\,g(Y)] \text{ for all } g.$$

Thus it follows that

$$\begin{aligned}\mathbb{E}[Z|Y=y] &= \mathbb{E}[h(X,Y)\,|Y=y] \\ &= \int h(x,y)\,\rho_{X|Y}(x,y)\,dx \\ &= \int \mathbb{E}[Z|Y=y, X=x]\,\rho_{X|Y}(x,y)\,dx.\end{aligned}$$

■

Remark 3.25. Often, $\mathbb{E}[Z|Y=y_0]$ may be computed as;

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}[Z|\,|Y-y_0|<\varepsilon].$$

To understand this formula, suppose that $h(y) := \mathbb{E}[Z|Y=y]$ and $\rho_Y(y)$ are continuous near y_0 and $\rho_Y(y_0)>0$. Then

$$\begin{aligned}\mathbb{E}[Z|\,|Y-y_0|<\varepsilon] &= \frac{\mathbb{E}[Z : |Y-y_0|<\varepsilon]}{\mathbb{P}(|Y-y_0|<\varepsilon)} \\ &= \frac{\mathbb{E}\left[h(Y)\,1_{|Y-y_0|<\varepsilon}\right]}{\mathbb{P}(|Y-y_0|<\varepsilon)} \\ &= \frac{\int h(y)\,1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy}{\int 1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy} \to h(y_0) \text{ as } \varepsilon\downarrow 0,\end{aligned}$$

wherein we have used, $h(y)\cong h(y_0)$ for y near y_0 and therefore,

$$\frac{\int h(y)\,1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy}{\int 1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy} \cong \frac{\int h(y_0)\,1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy}{\int 1_{|y-y_0|<\varepsilon}\rho_Y(y)\,dy} = h(y_0).$$

Here is a consequence of this result.

Theorem 3.26 (Iterated conditioning 2). *Suppose that Ω is partitioned into disjoint sets $\{A_i\}_{i=1}^n$ and y_0 is given such that the following limits exists;*

$$\mathbb{P}(A_i|Y=y_0) := \lim_{\varepsilon\downarrow 0}\mathbb{P}(A_i|\,|Y-y_0|<\varepsilon) \text{ and}$$
$$\mathbb{E}[Z|Y=y_0, A_i] := \lim_{\varepsilon\downarrow 0}\mathbb{E}[Z|\,|Y-y_0|<\varepsilon, A_i].$$

(In particular we are assuming that and $\mathbb{P}(|Y-y_0|<\varepsilon, A_i)>0$ for all $\varepsilon>0$.) Then

$$\mathbb{E}[Z|Y=y_0] = \sum_{i=1}^n \mathbb{E}[Z|Y=y_0, A_i]\,\mathbb{P}(A_i|Y=y_0).$$

Proof. Since,

$$\begin{aligned}\mathbb{E}[Z|\,|Y-y_0|<\varepsilon, A_i] &= \frac{\mathbb{E}\left[Z1_{A_i}1_{|Y-y_0|<\varepsilon}\right]}{\mathbb{P}(A_i\cap\{|Y-y_0|<\varepsilon\})} \\ &= \frac{\mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon]\,\mathbb{P}(|Y-y_0|<\varepsilon)}{\mathbb{P}(A_i\cap\{|Y-y_0|<\varepsilon\})} \\ &= \frac{\mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon]}{\mathbb{P}(A_i|\{|Y-y_0|<\varepsilon\})},\end{aligned}$$

it follows that

$$\begin{aligned}\lim_{\varepsilon\downarrow 0}\mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon] &= \lim_{\varepsilon\downarrow 0}\left(\mathbb{E}[Z|\,|Y-y_0|<\varepsilon, A_i]\cdot\mathbb{P}(A_i|\{|Y-y_0|<\varepsilon\})\right) \\ &= \mathbb{E}[Z|Y=y_0, A_i]\,\mathbb{P}(A_i|Y=y_0).\end{aligned}$$

Moreover,

$$\begin{aligned}\sum_{i=1}^n \lim_{\varepsilon\downarrow 0}\mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon] &= \lim_{\varepsilon\downarrow 0}\sum_{i=1}^n \mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon] \\ &= \lim_{\varepsilon\downarrow 0}\mathbb{E}\left[Z\sum_{i=1}^n 1_{A_i}|\,|Y-y_0|<\varepsilon\right] \\ &= \lim_{\varepsilon\downarrow 0}\mathbb{E}[Z|\,|Y-y_0|<\varepsilon] = \mathbb{E}[Z|Y=y_0],\end{aligned}$$

and therefore

$$\begin{aligned}\mathbb{E}[Z|Y=y_0] &= \sum_{i=1}^n \lim_{\varepsilon\downarrow 0}\mathbb{E}[Z1_{A_i}|\,|Y-y_0|<\varepsilon] \\ &= \sum_{i=1}^n \mathbb{E}[Z|Y=y_0, A_i]\,\mathbb{P}(A_i|Y=y_0)\end{aligned}$$

as claimed. ■

Remark 3.27 (An advanced remark). The last result is in fact a special case of Theorem 3.24 wherein we take $X = \sum_i i 1_{A_i}$. In this case, any function $u(X,Y)$ may be written as,

$$u(X,Y) = \sum_i 1_{X=i} u(i,Y).$$

Thus if we let $h(x,y) := \mathbb{E}[Z|(X=x,Y=y)]$, i.e. $h(X,Y) = \mathbb{E}[Z|(X,Y)]$ a.s. then on one hand;

$$\mathbb{E}[h(X,Y)u(X,Y)] = \sum_i \mathbb{E}[h(X,Y)1_{X=i}u(i,Y)] = \sum_i \mathbb{E}[h(i,Y)u(i,Y)1_{X=i}]$$

while on the other,

$$\mathbb{E}[h(X,Y)u(X,Y)] = \mathbb{E}[Zu(X,Y)] = \sum_i \mathbb{E}[Z1_{X=i}u(i,Y)].$$

Taking $u(i,Y) = \delta_{i,j}v(Y)$ and comparing the resulting expressions shows,

$$\mathbb{E}[h(j,Y)v(Y)1_{X=j}] = \mathbb{E}[Z1_{X=j}v(Y)] \text{ for all } v$$

and therefore that

$$\mathbb{E}[Z1_{X=j}|Y] = \mathbb{E}[h(j,Y)1_{X=j}|Y] = h(j,Y)\cdot\mathbb{E}[1_{X=j}|Y].$$

Summing this equation on j then shows,

$$\mathbb{E}[Z|Y] = \sum_j \mathbb{E}[Z1_{X=j}|Y] = \sum_j h(j,Y)\cdot\mathbb{E}[1_{X=j}|Y],$$

which reads,

$$\begin{aligned}\mathbb{E}[Z|Y=y] &= \sum_j \mathbb{E}[Z|X=j,Y=y]\cdot\mathbb{E}[1_{X=j}|Y=y] \\ &= \sum_j \mathbb{E}[Z|X=j,Y=y]\cdot\mathbb{P}[X=j|Y=y] \\ &= \sum_j \mathbb{E}[Z|A_j,Y=y]\cdot\mathbb{P}[A_j|Y=y] \qquad (\mu_Y - \text{a.s.}).\end{aligned}$$

where μ_Y is the law of Y, i.e. $\mu_Y(A) := \mathbb{P}(Y\in A)$.

Example 3.28. Suppose that $\{T_k\}_{k=1}^n$ are independent random times such that $\mathbb{P}(T_k > t) = e^{-\lambda_k t}$ for all $t \ge 0$ for some $\lambda_k > 0$. Let $\left\{\tilde{T}_k\right\}_{k=1}^n$ be the order statistics of the sequence, i.e. $\left\{\tilde{T}_k\right\}_{k=1}^n$ is the sequence $\{T_k\}_{k=1}^n$ in increasing order, i.e. $\tilde{T}_1 < \tilde{T}_2 < \cdots < \tilde{T}_n$. Further let $K = i$ on $\tilde{T}_1 = T_i$. Then

$$\mathbb{E}\left[f\left(\tilde{T}_2 - \tilde{T}_1\right)|\tilde{T}_1 = t\right] = \mathbb{E}\left[f\left(\tilde{T}_2 - \tilde{T}_1\right)|\tilde{T}_1 = t, K = i\right]\mathbb{P}\left(K = i|\tilde{T}_1 = t\right)$$

where $\left\{\tilde{T}_1 = t, K = i\right\} = \{t = T_i < T_j \text{ for } j \ne i\}$ and therefore,

$$\begin{aligned}\mathbb{E}\left[f\left(\tilde{T}_2 - \tilde{T}_1\right)|\tilde{T}_1 = t, K = i\right] &= \mathbb{E}\left[f\left(\tilde{T}_2 - t\right)|\tilde{T}_1 = t, K = i\right] \\ &= \mathbb{E}\left[f\left(\min_{j\ne i} T_j - t\right)|t = T_i < \min_{j\ne i} T_j\right] \\ &= \mathbb{E}\left[f\left(\min_{j\ne i} T_j - t\right)|t < \min_{j\ne i} T_j\right] \\ &= \mathbb{E}\left[f\left(\min_{j\ne i} T_j\right)\right],\end{aligned}$$

wherein we have used T_i is independent of $S := \min_{j\ne i} T_j \overset{d}{=} E(\lambda - \lambda_i)$, where $\lambda = \lambda_1 + \cdots + \lambda_n$. Since S is an exponential random variable, $\mathbb{P}(S > t + s|S > t) = \mathbb{P}(S > s)$, i.e. $S - t$ under $\mathbb{P}(\cdot|S > t)$ is the same in distribution as S under $\mathbb{P}$. Thus we have shown,

$$\mathbb{E}\left[f\left(\tilde{T}_2 - \tilde{T}_1\right)|\tilde{T}_1 = t\right] = \sum_i \mathbb{E}\left[f\left(\min_{j\ne i} T_j\right)\right]\mathbb{P}\left(K = i|\tilde{T}_1 = t\right).$$

We now compute informally,

$$\begin{aligned}\mathbb{P}\left(K = i|\tilde{T} = t\right) &= \frac{\mathbb{P}(\{t = T_i < T_j \text{ for } j \ne i\})}{\mathbb{P}\left(\left\{t = \tilde{T}_1\right\}\right)} = \frac{e^{-(\lambda-\lambda_i)t}\cdot\mathbb{P}(T_i = t)}{\mathbb{P}\left(\left\{t = \tilde{T}_1\right\}\right)} \\ &= \frac{e^{-(\lambda-\lambda_i)t}\cdot\lambda_i e^{-\lambda_i t}dt}{\lambda e^{-\lambda t}dt} = \frac{\lambda_i}{\lambda}.\end{aligned}$$

Here is the above computation done more rigorously;

$$\begin{aligned}\mathbb{P}\left(K = i|t < \tilde{T} \le t + \varepsilon\right) &= \frac{\mathbb{P}(\{T_i < T_j \text{ for } j \ne i, t < T_i \le t + \varepsilon\})}{\mathbb{P}\left(t < \tilde{T} \le t + \varepsilon\right)} \\ &= \frac{\int_t^{t+\varepsilon}\mathbb{P}(\{\tau < T_j \text{ for } j \ne i\})\lambda_i e^{-\lambda_i\tau}d\tau}{\int_t^{t+\varepsilon}\lambda e^{-\lambda\tau}d\tau} \\ &\overset{\varepsilon\downarrow 0}{\to} \frac{\mathbb{P}(\{t < T_j \text{ for } j \ne i\})\lambda_i e^{-\lambda_i t}}{\lambda e^{-\lambda t}} \\ &= \frac{e^{-(\lambda-\lambda_i)t}\lambda_i e^{-\lambda_i t}}{\lambda e^{-\lambda t}} = \frac{\lambda_i}{\lambda}.\end{aligned}$$

In summary we have shown;

$$\mathbb{E}\left[f\left(\tilde{T}_2 - \tilde{T}_1\right)|\tilde{T}_1 = t\right] = \sum_i \mathbb{E}\left[f\left(\min_{j\ne i} T_j\right)\right]\frac{\lambda_i}{\lambda}.$$

3.4 Summary on Conditional Expectation Properties

Let Y and X be random variables such that $\mathbb{E}Y^2 < \infty$ and h be function from the range of X to $\mathbb{R}$. Then the following are equivalent:

1. $h(X) = \mathbb{E}(Y|X)$, i.e. $h(X)$ is the conditional expectation of Y given X.
2. $\mathbb{E}(Y - h(X))^2 \le \mathbb{E}(Y - g(X))^2$ for all functions g, i.e. $h(X)$ is the best approximation to Y among functions of X.
3. $\mathbb{E}(Y \cdot g(X)) = \mathbb{E}(h(X) \cdot g(X))$ for all functions g, i.e. $Y - h(X)$ is orthogonal to all functions of X. Moreover, this condition uniquely determines $h(X)$.

The methods for computing $\mathbb{E}(Y|X)$ are given in the next two propositions.

Proposition 3.29 (Discrete Case). *Suppose that Y and X are discrete random variables and $p(y,x) := \mathbb{P}(Y = y, X = x)$. Then $\mathbb{E}(Y|X) = h(X)$, where*

$$h(x) = \mathbb{E}(Y|X = x) = \frac{\mathbb{E}(Y : X = x)}{\mathbb{P}(X = x)} = \frac{1}{p_X(x)} \sum_y y p(y,x) \tag{3.18}$$

and $p_X(x) = \mathbb{P}(X = x)$ is the marginal distribution of X which may be computed as $p_X(x) = \sum_y p(y,x)$.

Proposition 3.30 (Continuous Case). *Suppose that Y and X are random variables which have a joint probability density $\rho(y,x)$ (i.e. $\mathbb{P}(Y \in dy, X \in dx) = \rho(y,x)dydx$). Then $\mathbb{E}(Y|X) = h(X)$, where*

$$h(x) = \mathbb{E}(Y|X = x) := \frac{1}{\rho_X(x)} \int_{-\infty}^{\infty} y\rho(y,x)dy \tag{3.19}$$

and $\rho_X(x)$ is the marginal density of X which may be computed as

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho(y,x)dy.$$

Intuitively, in all cases, $\mathbb{E}(Y|X)$ on the set $\{X = x\}$ is $\mathbb{E}(Y|X = x)$. This intuitions should help motivate some of the basic properties of $\mathbb{E}(Y|X)$ summarized in the next theorem.

Theorem 3.31. *Let Y, Y_1, Y_2 and X be random variables. Then:*

1. $\mathbb{E}(Y_1 + Y_2|X) = \mathbb{E}(Y_1|X) + \mathbb{E}(Y_2|X)$.
2. $\mathbb{E}(aY|X) = a\mathbb{E}(Y|X)$ *for all constants* a.
3. $\mathbb{E}(f(X)Y|X) = f(X)\mathbb{E}(Y|X)$ *for all functions* f.
4. $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$.
5. *If Y and X are independent then* $\mathbb{E}(Y|X) = \mathbb{E}Y$.
6. *If $Y \ge 0$ then* $\mathbb{E}(Y|X) \ge 0$.

Remark 3.32. Property 4 in Theorem 3.31 turns out to be a very powerful method for computing expectations. I will finish this summary by writing out Property 4 in the discrete and continuous cases:

$$\mathbb{E}Y = \sum_x \mathbb{E}(Y|X = x)p_X(x) \qquad (\textbf{Discrete Case})$$

where

$$\mathbb{E}(Y|X = x) = \begin{cases} \frac{\mathbb{E}(Y 1_{X=x})}{\mathbb{P}(X=x)} & \text{if} \quad \mathbb{P}(X = x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[U(Y,X)] = \int \mathbb{E}(U(Y,X)|X = x)\rho_X(x)\,dx, \qquad (\textbf{Continuous Case})$$

where

$$\mathbb{E}[U(Y,x)|X = x] := \int U(y,x)\,\rho_{(Y|X)}(y,x)\,dy$$

and

$$\rho_{(Y|X)}(y,x) = \begin{cases} \frac{\rho_{(Y,X)}(y,x)}{\rho_X(x)} & \text{if } \rho_X(x) > 0 \\ 0 & \text{if } \rho_X(x) = 0 \end{cases}.$$

4

Filtrations and stopping times

Notation 4.1 *Let* $\mathbb{N} := \{1, 2, 3, \dots\}$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, *and* $\bar{\mathbb{N}} := \mathbb{N}_0 \cup \{\infty\} = \mathbb{N} \cup \{0, \infty\}$.

In this chapter, let Ω – be a sample space and S be a set called **state space**. Further let $\mathbb{X} := \{X_n\}_{n=0}^{\infty}$ be a sequence of random variables taking values in S which we will refer to generically as a **stochastic process.**

4.1 Filtrations

Definition 4.2 (Filtration). *The* ***filtration*** *associated to* $\mathbb{X}$ *is* $\left\{\mathcal{F}_n^{\mathbb{X}} := \sigma(X_0, \dots, X_n)\right\}_{n=0}^{\infty}$. *We further let* $\mathcal{F}_\infty^{\mathbb{X}} = \sigma(X_0, \dots, X_n, \dots)$.

Notice that $g : \Omega \to S$ is $\mathcal{F}_n^{\mathbb{X}}$ – measurable iff $g = G(X_0, \dots, X_n)$ some $G : S^{n+1} \to S$. Also notice that $\mathcal{F}_m^{\mathbb{X}} \subset \mathcal{F}_n^{\mathbb{X}}$ for all $0 \le m \le n \le \infty$.

Remark 4.3 (Independence). If $\mathbb{X} := \{X_n\}_{n=0}^{\infty}$ are independent S valued random variables and $n \in \mathbb{N}$, then by Theorem 2.14 X_{n+1} is independent of $(X_0, \dots, X_n)$ and therefore by item 5. of Theorem 3.15,

$$\mathbb{E}\left[f(X_{n+1}) \,|\, (X_0, \dots, X_n)\right] = \mathbb{E} f(X_{n+1}) \;\; \forall\, n \in \mathbb{N}_0 \text{ and } f : S \to \mathbb{R}. \tag{4.1}$$

It turns out that Eq. (4.1) is equivalent to the assertion that $\mathbb{X} := \{X_n\}_{n=0}^{\infty}$ are independent.

For example if Eq. (4.1) holds, then

$$\mathbb{E}\left[f_1(X_1) \,|X_0\right] = m_1 := \mathbb{E} f_1(X_1) \;\text{ a.s.}$$

which implies

$$\begin{aligned}\mathbb{E}\left[f_1(X_1) f_0(X_0)\right] &= \mathbb{E}\left[\mathbb{E}\left[f_1(X_1) \,|X_0\right] \cdot f_0(X_0)\right] = \mathbb{E}\left[m_1 f_0(X_0)\right] = m_1 \mathbb{E}\left[f_0(X_0)\right] \\ &= \mathbb{E} f_1(X_1) \cdot \mathbb{E}\left[f_0(X_0)\right].\end{aligned}$$

Similarly, another application of Eq. (4.1) gives,

$$\mathbb{E}\left[f_2(X_2) \,|\, (X_0, X_1)\right] = m_2 := \mathbb{E}\left[f_2(X_2)\right]$$

and therefore,

$$\begin{aligned}\mathbb{E}\left[f_2(X_2) f_1(X_1) f_0(X_0)\right] &= \mathbb{E}\left[\mathbb{E}\left[f_2(X_2) \,|\, (X_0, X_1)\right] \cdot f_1(X_1) f_0(X_0)\right] \\ &= m_2 \mathbb{E}\left[f_1(X_1) f_0(X_0)\right] \\ &= \mathbb{E}\left[f_2(X_2)\right] \cdot \mathbb{E} f_1(X_1) \cdot \mathbb{E}\left[f_0(X_0)\right].\end{aligned}$$

The general result follows by induction.

Besides independence the two types of dependence structures we will study most in this course are described in the following two definitions.

Definition 4.4 (Markov Property). *Let* $(\Omega, \mathbb{P})$ *be a probability space and* $\mathbb{X} := \{X_n : \Omega \to S\}_{n=0}^{\infty}$ *be a stochastic process. We say that* $(\Omega, \mathbb{X}, \mathbb{P})$ *has the* ***Markov property*** *if*

$$\mathbb{E}\left[f(X_{n+1}) \,|\, (X_0, \dots, X_n)\right] = \mathbb{E}\left[f(X_{n+1}) \,|X_n\right]$$

for all $n \ge 0$ *and* $f : S \to \mathbb{R}$ *bounded or non-negative functions.*

Definition 4.5 (Martingales). *Let* $\{M_n\}_{n=0}^{\infty}$ *be a sequence of complex or real valued integrable random variables on a probability space* $(\Omega, \mathbb{P})$. *We say* $\{M_n\}_{n=0}^{\infty}$ *is a martingale if*

$$\mathbb{E}\left[M_{n+1} | \,(M_0, \dots, M_n)\right] = M_n \text{ for } n = 0, 1, 2, \dots.$$

In other words,

$$\mathbb{E}\left[M_{n+1} - M_n | \,(M_0, \dots, M_n)\right] = 0$$

which states the expected future increment given the past is zero.

Example 4.6. If $\{X_n\}_{n=0}^{\infty}$ are independent random variables with $\mathbb{E} X_n = 0$, then $M_n = X_0 + \dots + X_n$ is a martingale. Indeed,

$$\begin{aligned}\mathbb{E}\left[M_{n+1} - M_n | \,(M_0, \dots, M_n)\right] &= \mathbb{E}\left[M_{n+1} - M_n | \,(X_0, \dots, X_n)\right] \\ &= \mathbb{E}\left[X_{n+1} | \,(X_0, \dots, X_n)\right] = \mathbb{E} X_{n+1} = 0,\end{aligned}$$

wherein we have used item 5. of Theorem 3.15 for the second to last equality.

4.2 Stopping Times

Definition 4.7 (Stopping time). *A function $\tau : \Omega \to \bar{\mathbb{N}} := \mathbb{N}_0 \cup \{\infty\}$ is a $\{\mathcal{F}_n\}$ – **stopping time** iff*

$$\{\tau = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}_0.$$

[In words, we should be able to determine if we are going to stop at time n from observing the information about $\{X_k\}_{k=0}^{\infty}$ that we have up to time n. See Remark 4.12 below for a more precise version of the last statement.]

Remark 4.8. According to Lemma 3.2, $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff for each $n \in \bar{\mathbb{N}}$, there exists a function f_n so that

$$1_{\{\tau=n\}} = f_n(X_0, \dots, X_n).$$

More precisely there must sets $A_n \subset S^{n+1}$ for $n \in \bar{\mathbb{N}}_0$ so that

$$\{\tau = n\} = \{(X_0, \dots, X_n) \in A_n\} \ \forall\, n \in \bar{\mathbb{N}}_0.$$

Example 4.9. If $\tau : \Omega \to \bar{\mathbb{N}}$ is constant, $\tau(\omega) = k$ for some $k \in \mathbb{N}_0$, then τ is a stopping time.

Example 4.10 (First Hitting times). Let $A \subset S$ be a set and let

$$H_A := \inf\{n \geq 0 : X_n \in A\}$$

where $\inf \emptyset := \infty$. We call H_A the **first hitting time** of A. Since

$$\begin{aligned}\{H_A = n\} &= \{X_0 \notin A, \dots, X_{n-1} \notin A^c, X_n \in A\} \\ &= \{X_0 \in A^c, \dots, X_{n-1} \in A^c, X_n \in A\} \\ &= (X_0, X_1, \dots, X_n)^{-1}(A^c \times \cdots \times A^c \times A) \in \mathcal{F}_n^{\mathbb{X}}\end{aligned}$$

we see that H_A is a stopping time.

Example 4.11 (First Hitting time after 0). Let $A \subset S$ be a set and let

$$T_A := \inf\{n > 0 : X_n \in A\}$$

where $\inf \emptyset := \infty$. Since $\{T_A = 0\} = \emptyset \in \mathcal{F}_0$ and for $n \geq 1$ we have

$$\begin{aligned}\{T_A = n\} &= \{X_1 \in A^c, \dots, X_{n-1} \in A^c, X_n \in A\} \\ &= (X_0, X_1, \dots, X_n)^{-1}(S \times A^c \times \cdots \times A^c \times A) \in \mathcal{F}_n^{\mathbb{X}},\end{aligned}$$

it follows that T_A is a stopping time.

Remark 4.12 (Another stopping time test). Let $\tau : \Omega \to \bar{\mathbb{N}}_0$ be a function, then τ is stopping time iff τ satisfies the following test.

Stopping time test. To each $\omega \in \Omega$, let $n = \tau(\omega) \in \bar{\mathbb{N}}_0$. If $n = \infty$ there is nothing to check. If $n < \infty$, then τ passes the test if $\tau(\omega') = n$ for any other $\omega' \in \Omega$ such that $X_k(\omega) = X_k(\omega')$ for $0 \leq k \leq n$.

Exercise 4.1. Let $\tau : \Omega \to \bar{\mathbb{N}}$ be a function. Verify the following are equivalent;

1. τ is a stopping time.
2. $\{\tau \leq n\} \in \mathcal{F}_n^{\mathbb{X}}$ for all $n \in \mathbb{N}_0$.
3. $\{\tau > n\} \in \mathcal{F}_n^{\mathbb{X}}$ for all $n \in \mathbb{N}_0$.

Also show that if τ is a stopping time then $\{\tau = \infty\} \in \mathcal{F}_\infty^{\mathbb{X}}$.

Exercise 4.2. If τ and σ are two stopping times shows, $\sigma \wedge \tau = \min\{\sigma, \tau\}$, $\sigma \vee \tau = \max\{\sigma, \tau\}$, and $\sigma + \tau$ are all stopping times.

Exercise 4.3 (Hitting time after a stopping time). Let σ be any stopping time. Show

$$\begin{aligned}\tau_1 &= \inf\{n \geq \sigma : X_n \in B\} \text{ and} \\ \tau_2 &= \inf\{n > \sigma : X_n \in B\}.\end{aligned}$$

are both stopping times.

Definition 4.13. *Let τ be a stopping time. A function $F : \Omega \to W$ is said to be $\mathcal{F}_\tau^{\mathbb{X}}$ – measurable if*

$$F = \text{“}\sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau=n\}} F_n\text{”} \text{ i.e. } F = F_n \text{ on } \{\tau = n\} \ \forall\, n \in \bar{\mathbb{N}},$$

where $F_n : \Omega \to W$ is $\mathcal{F}_n^{\mathbb{X}}$ – measurable for all $n \in \bar{\mathbb{N}}$. In more detail, we are assuming for each $n \in \bar{\mathbb{N}}$ there exists $f_n : S^{n+1} \to W$ such that $F_n = f_n(X_0, \dots, X_n)$ and $F = F_n$ on $\{\tau = n\}$. We also say that $A \subset \Omega$ is in $\mathcal{F}_\tau^{\mathbb{X}}$ iff $A \cap \{\tau = n\} \in \mathcal{F}_n^{\mathbb{X}}$ for all $n \in \bar{\mathbb{N}}$.

Remark 4.14. A set $A \subset \Omega$ is in $\mathcal{F}_\tau^{\mathbb{X}}$ iff 1_A is $\mathcal{F}_\tau^{\mathbb{X}}$ – measurable. Indeed, if $A \in \mathcal{F}_\tau^{\mathbb{X}}$ then there exists $f_n : S^{n+1} \to \{0, 1\}$ such that $1_{A \cap \{\tau=n\}} = f_n(X_0, \dots, X_n)$. Conversely, if 1_A is $\mathcal{F}_\tau^{\mathbb{X}}$ – measurable, then there exists $f_n : S^{n+1} \to \mathbb{R}$ such that

$$1_A = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau=n\}} f_n(X_0, \dots, X_n)$$

and so $1_{A \cap \{\tau=n\}} = f_n(X_0, \dots, X_n)$. Thus it follows that

$$A \cap \{\tau = n\} = \{f_n(X_0, \dots, X_n) = 1\} = (X_0, \dots, X_n)^{-1}\left(f_n^{-1}(\{1\})\right) \in \mathcal{F}_n^{\mathbb{X}}$$

which implies $A \cap \{\tau = n\} \in \mathcal{F}_n^{\mathbb{X}}$ for all $n \in \mathbb{N}_0$ and hence $A \in \mathcal{F}_\tau^{\mathbb{X}}$.

Example 4.15. If τ is a stopping time and $f : S \to \mathbb{R}$ is a function, then $F := 1_{\tau<\infty} f(X_\tau)$ is $\mathcal{F}_\tau$ – measurable. Indeed, for $n \in \mathbb{N}_0$ we have

$$F 1_{\{\tau=n\}} = f(X_n) 1_{\{\tau=n\}}.$$

Example 4.16. If $f \geq 0$, then

$$F := \sum_{k \leq \tau} f(X_k)$$

is $\mathcal{F}_\tau$ – measurable. Indeed, for $n \in \bar{\mathbb{N}}$ we have

$$F 1_{\{\tau=n\}} = 1_{\{\tau=n\}} \sum_{k \leq n} f(X_k)$$

which is of the desired form.

Theorem 4.17. *Suppose now that $\mathbb{P}$ is a probability on Ω. If $Z \in L^1(\mathbb{P})$ and τ is a stopping time, then*

$$\mathbb{E}[Z|\mathcal{F}_\tau] = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau=n\}} \mathbb{E}[Z|\mathcal{F}_n]. \tag{4.2}$$

Proof. Let F be a bounded $\mathcal{F}_\tau$ – measurable function, i.e. $F = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau=n\}} F_n$ where $F_n = f_n(X_0, \ldots, X_n)$ are $\mathcal{F}_n$ – measurable functions as in Definition 4.13. Since $F 1_{\{\tau=n\}} = F_n 1_{\{\tau=n\}}$ is $\mathcal{F}_n$ – measurable for all n, we find

$$\begin{aligned}
\mathbb{E}[ZF] &= \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}\left[Z F_n 1_{\{\tau=n\}}\right] = \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}\left[\mathbb{E}[Z|\mathcal{F}_n] F_n 1_{\{\tau=n\}}\right] \\
&= \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}\left[\mathbb{E}[Z|\mathcal{F}_n] F 1_{\{\tau=n\}}\right] = \mathbb{E}\left[\left[\sum_{n \in \bar{\mathbb{N}}} \mathbb{E}[Z|\mathcal{F}_n] 1_{\{\tau=n\}}\right] F\right].
\end{aligned}$$

As this is true for all bounded $\mathcal{F}_\tau$ – measurable functions F we conclude that Eq. (4.2) holds. ∎

Part II

Discrete Time & Space Markov Processes

5

Markov Chain Basics

In deterministic modeling one often has a dynamical system on a state space S. The dynamical system often takes on one of the two forms;

1. there exists $f : S \to S$ and a state x_n then evolves according to the rule $x_{n+1} = f(x_n)$. [More generally one might allow $x_{n+1} = f_n(x_0, \ldots, x_n)$ where $f_n : S^{n+1} \to S$ is a given function for each n.
2. There exists a vector field f on S (where now $S = \mathbb{R}^d$ or a manifold) such that $\dot{x}(t) = f(x(t))$. [More generally, we might allow for $\dot{x}(t) = f\left(t; x|_{[0,t]}\right)$, a functional differential equation.]

Much of our time in this course will be to explore the above two situations where some extra randomness is added at each state of the game. Namely;

1. We may now have that $X_{n+1} \in S$ is random and evolves according to
$$X_{n+1} = f(X_n, \xi_n)$$
where $\{\xi_n\}_{n=0}^{\infty}$ is a sequence of i.i.d. random variables. Alternatively put, we might simply let $f_n := f(\cdot, \xi_n)$ so that $f_n : S \to S$ is a sequence of i.i.d. random functions from S to S.[1] Then $\{X_n\}_{n=0}^{\infty}$ is defined recursively by
$$X_{n+1} = f_n(X_n) \text{ for } n = 0, 1, 2, \ldots. \text{(i.e. } X_{n+1}(\omega) = f_n(\omega, X_n(\omega))) \quad (5.1)$$
This is the typical example of a time-homogeneous Markov chain. We assume that $X_0 \in S$ is given with an initial condition which is either deterministic or is independent of the $\{f_n\}_{n=0}^{\infty}$.
2. Later in the course we will study the continuous time analogue,
$$\text{“}\dot{X}_t = f_t(X_t)\text{”}$$
where $\{f_t\}_{t \geq 0}$ are again i.i.d. random vector-fields. The continuous time case will require substantially more technical care.

[1] To be more explicit, we actually have $f_n : \Omega \times S \to S$. When we write $f_n(y)$ below this is short hand for the random function,
$$\Omega \ni \omega \to f_n(\omega, y).$$
As usual in probability we do our best to suppress the ω from the notation.

5.1 Markov Chain Descriptions

Notation 5.1 *Given a random function, $f : S \to S$ we may describe its "statistics" or distribution in two ways. The first is to assign a matrix to f while the second is to assign a weighted graph to f.*

1. ***(Matrix assignment.)*** *The first way is to let*
$$p(x, y) := \mathbb{P}(f(x) = y) = \mathbb{P}(\{\omega \in \Omega : f_\omega(x) = y\}) \ \forall\ x, y \in S.$$
Here we must have that $p(x, y) \in [0, 1]$ and $\sum_{y \in S} p(x, y) = 1$ for all $x \in S$.
2. ***(Jump diagram.)*** *Given $p(x, y)$ as above, let $\langle x, y \rangle$ be an edge of a graph over S iff $p(x, y) > 0$ and weight this edge by $p(x, y)$.*
The function $p : S \times S \to [0, 1]$ is called the ***one step transition probability*** *associate to the Markov chain $\{X_n\}_{n=0}^{\infty}$. Notice that;*
   a) *$\sum_y p(x, y) = \sum_{y \in S} \mathbb{P}(y = f_n(x)) = 1$, and*
   b) *as we will see in Theorem 5.11 below, the law of $\{X_n\}_{n=0}^{\infty}$ is completely determined by the one step transition probability p and the initial distribution, $\pi(x) := \mathbb{P}(X_0 = x)$.*

Notation 5.2 *If $p(x, y) \in [0, 1]$ and $\sum_{y \in S} p(x, y) = 1$ for all $x \in S$, we let $\mathbf{P}$ be the matrix indexed by S so that $\mathbf{P}_{xy} = p(x, y)$ for all $x, y \in S$. We refer to such a matrix, $\mathbf{P}$, as a* ***Markov matrix or a transition matrix.***

Example 5.3. The transition matrix,
$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$
is represented by the jump diagram in Figure 5.1.

Example 5.4. The jump diagram for
$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$
is shown in Figure 5.2.

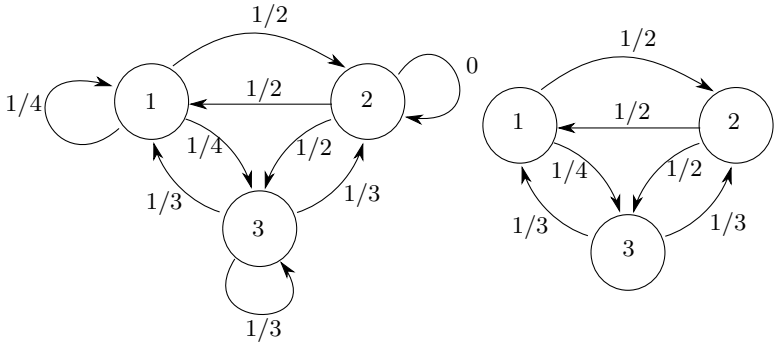


Fig. 5.1. A simple 3 state jump diagram. We typically abbreviate the jump diagram on the left by the one on the right. That is we infer by conservation of probability there has to be probability 1/4 of staying at 1, 1/3 of staying at 3 and 0 probability of staying at 2.

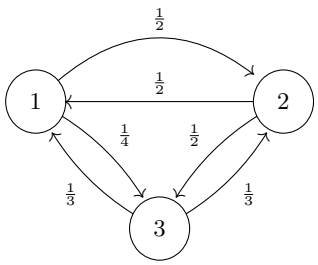


Fig. 5.2. In the above diagram there are jumps from 1 to 1 with probability 1/4 and jumps from 3 to 3 with probability 1/3 which are not explicitly shown but must be inferred by conservation of probability.

Example 5.5. Suppose that $S = \{1, 2, 3\}$, then

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

has the jump graph given by 5.3.

Example 5.6 (Realizing Random Functions I). Suppose we are given the Markov matrix

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 \end{bmatrix}$$

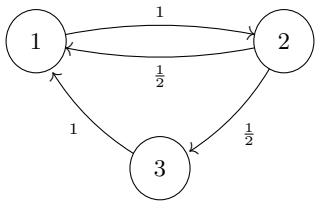


Fig. 5.3. A simple 3 state jump diagram.

and we want to simulate the associated chain by producing random functions $\{f_n\}_{n=0}^{\infty}$ such that $\mathbb{P}(f_n(i) = j) = \mathbf{P}_{ij}$ for $1 \le i, j \le 3$. One experimental way to do this is to use three independent spinners as shown in Figure 5.4.

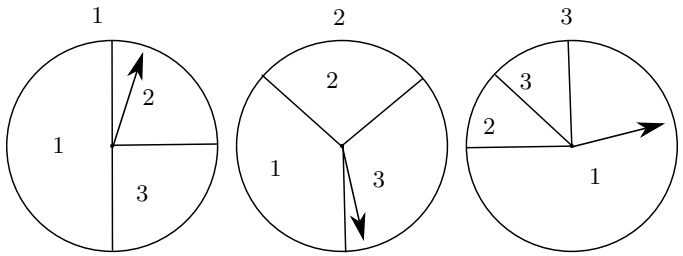


Fig. 5.4. In this figure we generate the random function, $f : \{1, 2, 3\} \to \{1, 2, 3\}$ using the results of three independent spinners. For the sample point shown here we have $f(1) = 2$, $f(2) = 3$, and $f(3) = 1$.

In the previous example we constructed our random function f so that $\{f(1), f(2), f(3)\}$ were independent. In fact at any given time n, we only use one of the values of $\{f_n(1), f_n(2), f_n(3)\}$ and so the independence and in fact the joint distribution of the three values, $\{f_n(1), f_n(2), f_n(3)\}$, is unimportant. The next example uses this fact to produce a more numerically tractable construction of the random functions $\{f_n\}$.

Example 5.7 (Realizing Random Functions II). Again suppose that $\mathbf{P}$ is the Markov matrix in Example 5.6. We now keep the same spinner faces as in Figure 5.4 but now use a "master spinner" to determine the placement of the spinner in each of the spin faces all at once, see Figure 5.5. In this case we still have $\mathbb{P}(f(i) = j) = \mathbf{P}_{ij}$ but is is now no longer true that $\{f(1), f(2), f(3)\}$ are independent. As mentioned above this is unimportant from the Markov chains point of vies.

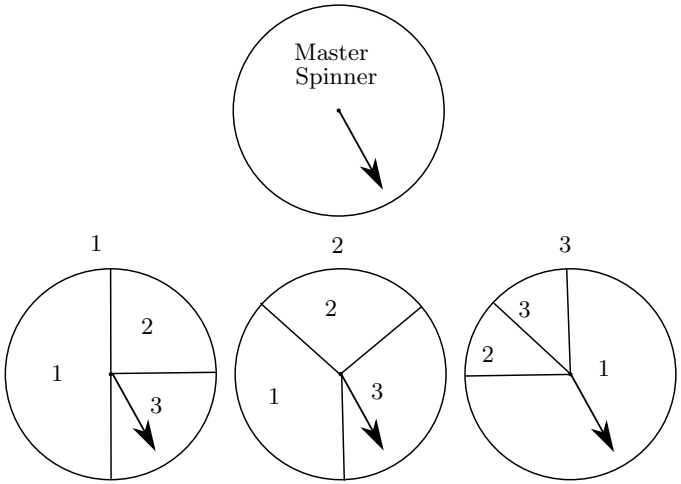


Fig. 5.5. In this figure we generate the random function, $f : \{1,2,3\} \to \{1,2,3\}$ using the result of a master spinner to determine the (same) placement of the needle on all three spin faces at once. For this sample point we have $f(1) = f(2) = 3$ and $f(3) = 1$.

Example 5.8 (Random Walks on Graphs). Let S be a set and G be a graph on S. We then take $\{f_n\}_{n=0}^{\infty}$ i.i.d. such that

$$\mathbb{P}(f_n(x) = y) = \begin{cases} 0 & \text{if } \{x,y\} \notin G \\ \frac{1}{d(x)} & \text{if } \{x,y\} \in G \end{cases}$$

where $d(x) := \#\{y \in S : \{x,y\} \in G\}$. We can give a similar definition for directed graphs, namely

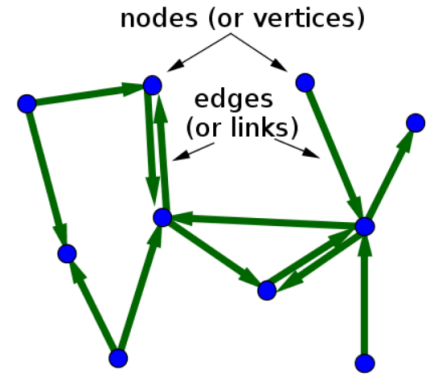
$$\mathbb{P}(f_n(x) = y) = \begin{cases} 0 & \text{if } \langle x \to y \rangle \notin G \\ \frac{1}{d(x)} & \text{if } \langle x \to y \rangle \in G \end{cases}$$

where now

$$d(x) := \#\{y \in S : \langle x \to y \rangle \in G\}.$$

A directed graph on S is a subset $G \subset S^2 \setminus \Delta$ where $\Delta := \{(s,s) : S \in S\}$. We say that G is undirected if $(s,t) \in G$ implies $(t,s) \in G$. As we have seen, every Markov chain is really determined by a weighted random walk on a graph.

Example 5.9. Suppose we flip a fair coin repeatedly and would like to find the first time the pattern HHT appears. To do this we will later examine the Markov chain, $Y_n = (X_n, X_{n+1}, X_{n+2})$ where $\{X_n\}_{n=0}^{\infty}$ is the sequence of unbiased independent coin flips with values in $\{H,T\}$. The state space for Y_n is



$$S = \{\, TTT \;\; THT \;\; TTH \;\; THH \;\; HHH \;\; HTT \;\; HTH \;\; HHT \,\}.$$

The transition matrix for recording three flips in a row of a fair coin is

$$\mathbf{P} = \frac{1}{2}\left[\begin{array}{c|cccccccc} & TTT & THT & TTH & THH & HHH & HTT & HTH & HHT \\ \hline TTT & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ THT & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ TTH & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ THH & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ HHH & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ HTT & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ HTH & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ HHT & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right].$$

Example 5.10 (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 $X_n = (\#$ 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,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i-1, i+1\} \\ \frac{i}{N} & \text{if } \quad j = i-1 \\ \frac{N-i}{N} & \text{if } \quad j = i+1 \end{cases} =: q(i,j)$$

As these probabilities do not depend on n, $\{X_n\}$ is a time homogeneous Markov chain.

Exercise 5.1. 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 7×7 transition matrix, q, with entries given by $q(i,j) := \mathbb{P}(X_{n+1} = j | X_n = i)$, where X_n denotes the room the rat is in at time n.

5.2 Joint Distributions of an MC

Theorem 5.11. *Suppose that $\{f_n\}_{n=0}^\infty$ are i.i.d. random functions*[2] *from S to S, $X_0 \in S$ is a random variable independent of the $\{f_n\}_{n=0}^\infty$, and $\{X_n\}_{n=1}^\infty$ are generated as in Eq. (5.1). Further let*

$$p(x,y) := \mathbb{P}(y = f_n(x)) = \mathbb{P}(y = f_1(x))$$

and

$$\pi(x) := \mathbb{P}(X_0 = x).$$

Then

$$\mathbb{P}_\pi(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \pi(x_0)\, p(x_0,x_1)\, p(x_1,x_2) \dots p(x_{n-1},x_n) \tag{5.2}$$

for all $x_0, \dots, x_n \in S$. In particular if $G : S^{n+1} \to \mathbb{R}$ is a function, then

$$\begin{aligned}&\mathbb{E}_\pi[G(X_0,\dots,X_n)] \\ &= \sum_{x_0,\dots,x_n \in S} G(x_0,\dots,x_n)\, \pi(x_0)\, p(x_0,x_1)\, p(x_1,x_2) \dots p(x_{n-1},x_n).\end{aligned} \tag{5.3}$$

Proof. This is straightforward to verify using

$$\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = \{X_0 = x_0\} \cap \cap_{k=0}^{n-1} \{x_{k+1} = f_k(x_k)\}$$

and hence as $(X_0, f_0(x_0), \dots, f_{n-1}(x_{n-1}))$ is a list of independent S – valued random functions it follows that

$$\begin{aligned}&\mathbb{P}_\pi(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbb{P}(X_0 = x_0) \cdot \prod_{k=0}^{n-1} \mathbb{P}(x_{k+1} = f_k(x_k)) \\ &= \pi(x_0) \cdot \prod_{k=0}^{n-1} p(x_k, x_{k+1}).\end{aligned}$$

■

[2] Actually, the functions $\{f_n\}$ do not need to be identically distributed, we really need only require that $P(y = f_n(x))$ is independent of n for all n. The correlations between events like $\{y = f_n(x)\}$ and $\{y' = f_n(x')\}$ are irrelevant.

Notation 5.12 ($\mathbb{P}_x$ and $\mathbb{P}_\pi$) *We will either write $\{X_n^\pi\}_{n=0}^\infty$ for the Markov chain constructed in Theorem 5.11 when $\pi(y) := \mathbb{P}(X_0 = y)$ or we will indicate the starting distribution π by writing $\mathbb{P}_\pi$ rather than $\mathbb{P}$ in this case. If $\pi(y) = \delta_x(y)$ for some $x \in S$, then we write simply write $\{X_n^x\}_{n=0}^\infty$ instead of $\{X_n^{\delta_x}\}_{n=0}^\infty$ and $\mathbb{P}_x$ instead of $\mathbb{P}_{\delta_x}$. We also let $\mathbb{E}_\pi$ and $\mathbb{E}_x$ denote expectations relative to $\mathbb{P}_\pi$ and $\mathbb{P}_x$ respectively. We may also write,*

$$\mathbb{P}_\pi = \sum_{x \in S} \pi(x)\, \mathbb{P}_x \text{ and } \mathbb{E}_\pi = \sum_{x \in S} \pi(x)\, \mathbb{E}_x.$$

Proposition 5.13. *Let us continue the notation above. The Markov chain $\{X_n\}_{n=0}^\infty$ has the Markov property. More precisely if $g : S \to \mathbb{R}$ is a bounded function, then*

$$\mathbb{E}[g(X_{n+1}) \,|\, (X_0,\dots,X_n)] = (\mathbf{P}g)(X_n) = \mathbb{E}[g(X_{n+1}) \,|\, X_n], \tag{5.4}$$

where

$$(\mathbf{P}g)(x) := \sum_{y \in S} p(x,y)\, g(y) = \sum_{y \in S} \mathbf{P}_{xy} g(y).$$

We also have

$$\begin{aligned}&\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= p(x_n, x_{n+1}) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).\end{aligned} \tag{5.5}$$

Proof. We will give two proofs. The first proof makes use of the construction of $\{X_n\}_{n=0}^\infty$ in in Eq. (5.1). The second proof uses the description of the distribution of $\{X_n\}$ given in Theorem 5.11.

Proof 1. By the construction of $\{X_n\}_{n=0}^\infty$ in in Eq. (5.1), the events $\{X_0 = x_0, \dots, X_n = x_n\}$ and $f_n(x_n)$ are independent and therefore,

$$\begin{aligned}h(x_0,\dots,x_n) &= \mathbb{E}[g(X_{n+1}) \,|\, X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{E}[g(f_n(x_n)) \,|\, X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{E}[g(f_n(x_n))] = \sum_{y \in S} g(y)\, \mathbb{P}(f_n(x_n) = y) \\ &= \sum_{y \in S} p(X_n, y)\, g(y).\end{aligned}$$

It now follows from Proposition 3.10 that

$$\mathbb{E}[g(X_{n+1}) \,|\, (X_0,\dots,X_n)] = h(X_0,\dots,X_n) = \sum_{y \in S} p(X_n, y)\, g(y).$$

Since the latter expression only depends on X_n, the previous equation along with the tower property (Theorem 3.19) of conditional expectations implies[3]

[3] We could alternatively make use of the fact that $\{X_n = x_n\}$ is independent of $f_n(x_n)$ and then working as before directly show the second equality in Eq. (5.4).

$$\sum_{y\in S} p(X_n,y)\,g(y) = \mathbb{E}\left[g(X_{n+1})\,|X_n\right].$$

At this point Eq. (5.4) is proved.

A similar argument (or take $g(y)=\delta_{x_{n+1},y}$ in Eq. (5.4)) also shows

$$\mathbb{P}(X_{n+1}=x_{n+1}|X_0=x_0,\dots,X_n=x_n) = p(x_n,x_{n+1}).$$

Similarly since $\{f_n(x_n)=x_{n+1}\}$ and $\{X_n=x_n\}$ are independent events we find,

$$\begin{aligned}\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n) &= \mathbb{P}(f_n(x_n)=x_{n+1}|X_n=x_n)\\ &= \mathbb{P}(f_n(x_n)=x_{n+1}) = p(x_n,x_{n+1}).\end{aligned}$$

Proof 2. We have

$$\begin{aligned}&\mathbb{E}\left[g(X_{n+1}) : X_0=x_0,\dots,X_n=x_n\right]\\ &= \mathbb{E}\left[1_{x_0}(X_0)\dots 1_{x_n}(X_n)\,g(X_{n+1})\right]\\ &= \sum_{x_{n+1}\in S}\pi(x_0)\,p(x_0,x_1)\dots p(x_{n-1},x_n)\,p(x_n,x_{n+1})\,g(x_{n+1})\\ &= \mathbb{P}(X_0=x_0,\dots,X_n=x_n)\sum_{x_{n+1}\in S}p(x_n,x_{n+1})\,g(x_{n+1})\end{aligned}$$

from which it follows that

$$\mathbb{E}\left[g(X_{n+1})\,|X_0=x_0,\dots,X_n=x_n\right] = \sum_{y\in S}p(x_n,y)\,g(y). \tag{5.6}$$

It now follows from this identity and Proposition 3.10 that

$$\mathbb{E}\left[g(X_{n+1})\,|\,(X_0,\dots,X_n)\right] = \sum_{y\in S}p(X_n,y)\,g(y)$$

which proves the first equality in Eq. (5.4). This equality along with the tower property of conditional expectation then implies,

$$\begin{aligned}\sum_{y\in S}p(X_n,y)\,g(y) &= \mathbb{E}\left[\sum_{y\in S}p(X_n,y)\,g(y)\,|X_n\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[g(X_{n+1})\,|\,(X_0,\dots,X_n)\right]|X_n\right] = \mathbb{E}\left[g(X_{n+1})\,|X_n\right].\end{aligned}$$

Taking $g=1_{\{x_{n+1}\}}$ in Eq. (5.6) shows,

$$\mathbb{P}(X_{n+1}=x_{n+1}|X_0=x_0,\dots,X_n=x_n) = p(x_n,x_{n+1})$$

or equivalently that

$$\mathbb{P}(X_{n+1}=x_{n+1},X_0=x_0,\dots,X_n=x_n) = \mathbb{P}(X_0=x_0,\dots,X_n=x_n)\,p(x_n,x_{n+1}).$$

Summing this last equation on $x_0,\dots,x_{n-1}$ gives

$$\mathbb{P}(X_{n+1}=x_{n+1},X_n=x_n) = \mathbb{P}(X_n=x_n)\,p(x_n,x_{n+1})$$

which implies,

$$\mathbb{P}(X_{n+1}=x_{n+1},X_n=x_n) = p(x_n,x_{n+1}).$$

Both equalities in Eq. (5.5) have now been verified.

■

Corollary 5.14. *For all $n\in\mathbb{N}$, $\mathbb{P}_\pi(X_n=y)=\sum_{x\in S}\pi(x)\,p^n(x,y)$ where $p^0(x,y)=\delta_x(y)$, $p^1=p$, and $\{p^n\}_{n\ge1}$ is defined inductively by*

$$p^n(x,y) := \sum_{z\in S}p(x,z)\,p^{n-1}(z,y).$$

Proof. From Eq. (5.3) with $G(x_0,\dots,x_n)=\delta_y(x_n)$ we learn that

$$\begin{aligned}\mathbb{P}_\pi(X_n=y) &= \sum_{x_0,\dots,x_n\in S}\delta_y(x_n)\,\pi(x_0)\,p(x_0,x_1)\,p(x_1,x_2)\dots p(x_{n-1},x_n)\\ &= \sum_{x\in S}\pi(x)\,p^n(x,y).\end{aligned}$$

■

Notation 5.15 *To emphasize the matrix aspect of Markov-process, we will often write $\mathbf{P}$ for the "matrix" with entries $\{p(x,y)\}_{x,y\in S}$ and correspondingly often write $\mathbf{P}_{xy}$ for $p(x,y)$ and $\mathbf{P}^n_{x,y}$ for $p^n(x,y)$.*

Definition 5.16. *We say $\pi:S\to[0,1]$ is an **invariant distribution** for the Markov chain determined by $p:S\times S\to[0,1]$ provided $\sum_{y\in S}\pi(y)=1$ and*

$$\pi(y) = \sum_{x\in S}\pi(x)\,p(x,y) \text{ for all } y\in S.$$

Alternatively put, π, should be a distribution such that $\mathbb{P}_\pi(X_n=x)=\pi(x)$ for all $n\in\mathbb{N}_0$ and $x\in S$. [The invariant distribution is found by solving the matrix equation, $\pi\mathbf{P}=\pi$ under the restrictions that $\pi(x)\ge0$ and $\sum_{x\in S}\pi(x)=1$.]

Example 5.17. If

$$\mathbf{P} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix},$$

then $\mathbf{P}^{\mathrm{tr}}$ has eigenvectors,

$$\left\{ \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} \right\} \leftrightarrow 0, \left\{ \begin{bmatrix} 1 \\ \frac{5}{6} \\ 1 \end{bmatrix} \right\} \leftrightarrow 1, \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\} \leftrightarrow -\frac{5}{12}.$$

The invariant distribution is given by

$$\begin{aligned} \pi &= \frac{1}{1+1+\frac{5}{6}} \begin{bmatrix} 1 & \frac{5}{6} & 1 \end{bmatrix} = \begin{bmatrix} \frac{6}{17} & \frac{5}{17} & \frac{6}{17} \end{bmatrix} \\ &= \begin{bmatrix} 0.352\,94 & 0.294\,12 & 0.352\,94 \end{bmatrix}. \end{aligned}$$

Notice that

$$\begin{aligned} \mathbf{P}^{100} &= \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}^{100} = \begin{bmatrix} 0.352\,94 & 0.294\,12 & 0.352\,94 \\ 0.352\,94 & 0.294\,12 & 0.352\,94 \\ 0.352\,94 & 0.294\,12 & 0.352\,94 \end{bmatrix} \\ &\cong \begin{bmatrix} \pi \\ \pi \\ \pi \end{bmatrix}. \end{aligned}$$

Exercise 5.2 (2 - step MC). Consider the following simple (i.e. no-brainer) two state "game" consisting of moving between two sites labeled 1 and 2. At each site you find a coin with sides labeled 1 and 2. The probability of flipping a 2 at site 1 is $a \in (0,1)$ and a 1 at site 2 is $b \in (0,1)$. If you are at site i at time n, then you flip the coin at this site and move or stay at the current site as indicated by coin toss. We summarize this scheme by the "jump diagram" of Figure 5.6. It is reasonable to suppose that your location, X_n, at time n is modeled by a

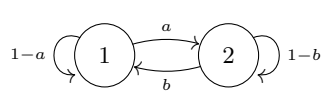


Fig. 5.6. The generic jump diagram for a two state Markov chain.

Markov process with state space, $S = \{1,2\}$. Explain (briefly) why this is a time homogeneous chain and find the one step transition probabilities,

$$p(i,j) = \mathbb{P}(X_{n+1} = j | X_n = i) \text{ for } i,j \in S.$$

Use your result and basic linear (matrix) algebra to compute, $\lim_{n\to\infty} \mathbb{P}(X_n = 1)$. Your answer should be independent of the possible starting distributions, $\nu = (\nu_1, \nu_2)$ for X_0 where $\nu_i := \mathbb{P}(X_0 = i)$.

Solution to Exercise 5.2. Writing $\mathbf{P}$ as a matrix with entry in the ith row and j^{th} column being $p(i,j)$, we have

$$\mathbf{P} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}.$$

If $\mathbb{P}(X_0 = i) = \nu_i$ for $i = 1,2$ then

$$\mathbb{P}(X_n = 1) = \sum_{k=1}^{2} \nu_k \mathbf{P}^n_{k,1} = [\nu \mathbf{P}^n]_1$$

where we now write $\nu = (\nu_1, \nu_2)$ as a row vector. A simple computation shows that

$$\begin{aligned} \det\left(\mathbf{P}^{\mathrm{tr}} - \lambda I\right) &= \det(\mathbf{P} - \lambda I) \\ &= \lambda^2 + (a+b-2)\lambda + (1-b-a) \\ &= (\lambda - 1)(\lambda - (1-a-b)). \end{aligned}$$

Note that

$$\mathbf{P}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

since $\sum_j p(i,j) = 1$ – this is a general fact. Thus we always know that $\lambda_1 = 1$ is an eigenvalue of $\mathbf{P}$. The second eigenvalue is $\lambda_2 = 1 - a - b$. We now find the eigenvectors of $\mathbf{P}^{\mathrm{tr}}$;

$$\mathrm{Nul}\left(\mathbf{P}^{\mathrm{tr}} - \lambda_1 I\right) = \mathrm{Nul}\left(\begin{bmatrix} -a & b \\ a & -b \end{bmatrix}\right) = \mathbb{R} \cdot \begin{bmatrix} b \\ a \end{bmatrix}$$

while

$$\mathrm{Nul}\left(\mathbf{P}^{\mathrm{tr}} - \lambda I_2\right) = \mathrm{Nul}\left(\begin{bmatrix} b & b \\ a & a \end{bmatrix}\right) = \mathbb{R} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

In fact by a direct check we have,

$$\begin{aligned} \begin{bmatrix} b & a \end{bmatrix} \mathbf{P} &= \begin{bmatrix} b & a \end{bmatrix} \text{ and} \\ \begin{bmatrix} 1 & -1 \end{bmatrix} \mathbf{P} &= (1-b-a) \begin{bmatrix} 1 & -1 \end{bmatrix}. \end{aligned}$$

Thus we may write

$$\nu = \alpha(b,a) + \beta(1,-1)$$

where

$$1 = \nu \cdot (1,1) = \alpha (b,a) \cdot (1,1) = \alpha (a+b).$$

Thus $\beta = \nu_1 - b = -(\nu_2 - a)$, we have

$$\nu = \frac{1}{a+b}(b,a) + \beta (1,-1).$$

and therefore,

$$\nu \mathbf{P}^n = (b,a)\mathbf{P}^n + \beta (1,-1)\mathbf{P}^n = \frac{1}{a+b}(b,a) + \beta (1,-1)\lambda_2^n.$$

By our assumptions on and $a, b \in (0,1)$ it follows that $|\lambda_2| < 1$ and therefore

$$\lim_{n\to\infty} \nu \mathbf{P}^n = \frac{1}{a+b}(b,a)$$

and we have shown

$$\lim_{n\to\infty} P(X_n = 1) = \frac{b}{a+b} \text{ and } \lim_{n\to\infty} P(X_n = 2) = \frac{a}{a+b}$$

independent of the starting distribution ν. Also observe that the convergence is exponentially fast. Notice that

$$\pi := \frac{1}{a+b}(b,a)$$

is the invariant distribution of this chain.

5.3 Hitting Times Estimates

Lemma 5.18. *For any random time $T : \Omega \to \mathbb{N} \cup \{0,\infty\}$ we have*

$$\mathbb{P}(T = \infty) = \lim_{n\to\infty} \mathbb{P}(T > n) \text{ and } \mathbb{E}T = \sum_{k=0}^{\infty} \mathbb{P}(T > k) = \sum_{k=1}^{\infty} \mathbb{P}(T \geq k).$$

Proof. The first equality is a consequence of the continuity of $\mathbb{P}$ (or use DCT) and the fact that

$$\{T > n\} \downarrow \{T = \infty\}.$$

The second equality is proved by the following simple computation;

$$\sum_{k=0}^{\infty} \mathbb{P}(T > k) = \sum_{k=0}^{\infty} \mathbb{E}1_{\{T>k\}} = \mathbb{E}\sum_{k=0}^{\infty} 1_{\{T>k\}} = \mathbb{E}T.$$

■

Let us now assume that $\{X_n\}_{n=0}^{\infty}$ is a Markov chain with values in S and transition kernel $\mathbf{P}$. (We often write $p(x,y)$ for $\mathbf{P}_{xy}$.) We are going to further assume that $B \subset S$ is non-empty proper subset of S and $A = S \setminus B$.

Definition 5.19 (Hitting times). *Given a subset $B \subset S$ we let H_B be the first time $\{X_n\}$ hits B, i.e.*

$$H_B = \min\{n : X_n \in B\}$$

with the convention that $H_B = \infty$ if $\{n : X_n \in B\} = \emptyset$. We call H_B the ***first hitting time*** *of B by $X = \{X_n\}_n$.*

Observe that

$$\begin{aligned}\{H_B = n\} &= \{X_0 \notin B, \dots, X_{n-1} \notin B, X_n \in B\} \\ &= \{X_0 \in A, \dots, X_{n-1} \in A, X_n \in B\}\end{aligned}$$

and

$$\{H_B > n\} = \{X_0 \in A, \dots, X_{n-1} \in A, X_n \in A\}$$

so that $\{H_B = n\}$ and $\{H_B > n\}$ only depends on $(X_0, \dots, X_n)$. Recall that a random time, $T : \Omega \to \mathbb{N} \cup \{0,\infty\}$, with either of these properties is called a **stopping time.**

Notation 5.20 *Let $\mathbf{Q} := \{\mathbf{P}_{x,y}\}_{x,y\in A}$ be the matrix $\mathbf{P}$ restricted to A.*

Proposition 5.21. *Using the notation above and letting $\mathbf{1}$ denote the constant function 1 on A then for all $x \in A$;*

$$\mathbb{P}_x(H_B > n) = [\mathbf{Q}^n \mathbf{1}]_x \tag{5.7}$$

$$\mathbb{P}_x(H_B = \infty) = \lim_{n\to\infty} [\mathbf{Q}^n \mathbf{1}]_x \text{ and} \tag{5.8}$$

$$\mathbb{E}_x(H_B) = \sum_{n=0}^{\infty} [\mathbf{Q}^n \mathbf{1}]_x. \tag{5.9}$$

Proof. If $x \in A$ and $n \in \mathbb{N}$, then

$$\begin{aligned}\mathbb{P}_x(H_B > n) &= \mathbb{P}_x(X_1 \in A, X_2 \in A, \dots, X_n \in A) \\ &= \sum_{x_1,\dots,x_n \in A} \mathbb{P}_x(X_0 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\ &= \sum_{x_1,\dots,x_n \in A} \mathbf{P}_{x,x_1}\mathbf{P}_{x_1,x_2}\dots\mathbf{P}_{x_{n-1},x_n} = \sum_{y\in A} \mathbf{Q}^n_{x,y} = [\mathbf{Q}^n\mathbf{1}]_x\end{aligned}$$

which proves Eq. (5.7). An application of Lemma 5.18 then gives the remaining results. ■

Lemma 5.22. *If $m, n \in \mathbb{N}$ and $x \in A$, then*

$$\mathbb{P}_x(H_B > m+n) = \sum_{y\in A} \mathbf{Q}^m_{x,y} \mathbb{P}_y(H_B > n) \tag{5.10}$$

and

$$\mathbb{P}_x(H_B=\infty)=\sum_{y\in S}\mathbf{Q}^m_{x,y}\mathbb{P}_y(H_B=\infty). \tag{5.11}$$

Proof. From Eq. (5.7),

$$\begin{aligned}\mathbb{P}_x(H_B>m+n)&=\left[\mathbf{Q}^{n+m}\mathbf{1}\right]_x=\left[\mathbf{Q}^m\mathbf{Q}^n\mathbf{1}\right]_x\\&=\sum_{y\in A}\mathbf{Q}^m_{x,y}\left[\mathbf{Q}^n\mathbf{1}\right]_y=\sum_{y\in A}\mathbf{Q}^m_{x,y}\mathbb{P}_y(H_B>n)\end{aligned}$$

which is Eq. (5.10). Letting $n\to\infty$ (using DCT) in Eq. (5.10) then leads to Eq. (5.11). ■

Corollary 5.23. *If there exists $\alpha<1$ such that $\mathbb{P}_x(H_B=\infty)\le\alpha$ for all $x\in A$, then $\mathbb{P}_x(H_B=\infty)=0$ for all $x\in A$. [In words; if there is a "uniform" chance that X hits B starting from any site, then X will surely hit B from any point in A.]*

Proof. From Eqs. (5.7) and (5.11) and the assumption that $\mathbb{P}_x(H_B=\infty)\le\alpha$ for all $x\in A$ we find,

$$\mathbb{P}_x(H_B=\infty)\le\sum_{y\in S}\mathbf{Q}^m_{x,y}\alpha=\alpha\left[\mathbf{Q}^m\mathbf{1}\right]_x=\alpha\mathbb{P}_x(H_B>m).$$

Letting $m\uparrow\infty$ then shows,

$$\mathbb{P}_x(H_B=\infty)\le\alpha\mathbb{P}_x(H_B=\infty)\implies\mathbb{P}_x(H_B=\infty)=0.$$

■

Corollary 5.24. *If there exists $\alpha\in(0,1)$ and $n\in\mathbb{N}$ such that*

$$\mathbb{P}_x(H_B>n)=\left[\mathbf{Q}^n\mathbf{1}\right]_x\le\alpha\ \forall\ x\in A, \tag{5.12}$$

then;

1. *$\mathbf{I}-\mathbf{Q}$ is invertible with*

$$(\mathbf{I}-\mathbf{Q})^{-1}_{x,y}=\sum_{n=0}^\infty\mathbf{Q}^n_{x,y}<\infty\ \forall\ x,y\in A,$$

2. $$\mathbb{E}_xH_B=\left[(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}\right]_x\le\frac{n}{1-\alpha}<\infty\ \forall\ x\in A,$$

where

$$\left[(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}\right]_x=\sum_{y\in A}(\mathbf{I}-\mathbf{Q})^{-1}_{x,y}=\sum_{y\in A}\sum_{n=0}^\infty\mathbf{Q}^n_{x,y},$$

3. *and for any $0<\beta<\sqrt[n]{\alpha}$ and $x\in A$,*

$$\mathbb{E}_x\left[\beta^{H_B}\right]\le\beta\left(1-\beta\sqrt[n]{\alpha}\right)^{-1}<\infty.$$

This last assertion implies $\sup_{x\in A}\mathbb{E}_xH_B^N<\infty$ for all $N\in\mathbb{N}$. [In words; if there is a "uniform" chance that X hits B starting from any site within a fixed number of steps, then the expected hitting time of B is finite and bounded independent of the starting distribution.]

Proof. Proof 1. The assumption in Eq. (5.12) may be stated as saying $\mathbf{Q}^n\mathbf{1}\le\alpha\mathbf{1}$. Therefore if $\ell=kn+m\in\mathbb{N}_0$ with $k\in\mathbb{N}_0$ ($k\le\ell/n$) and $0\le m<n$, then

$$\mathbb{P}_x(H_B>\ell)=\mathbf{Q}^{kn+m}\mathbf{1}=\left[\mathbf{Q}^n\right]^k\mathbf{Q}^m\mathbf{1}\le\left[\mathbf{Q}^n\right]^k\mathbf{1}\le\alpha^k\mathbf{1}. \tag{5.13}$$

Summing this last equation on $k\in\mathbb{N}_0$ and $0\le m<n$ (so that ℓ runs over $\mathbb{N}_0$) we find (with the aid of Eq. (5.9)) that

$$\mathbb{E}_xH_B=\sum_{n=0}^\infty\left[\mathbf{Q}^n\mathbf{1}\right]_x\le\sum_{k=0}^\infty\sum_{0\le m<n}\alpha^k=\frac{n}{1-\alpha}<\infty.$$

For the last assertion we see from Eq. (5.13) and observations that $k\le\ell/n$ (noted above) we find

$$\mathbb{P}_x(H_B>\ell)\le\alpha^{\ell/n}$$

and therefore,

$$\begin{aligned}\mathbb{E}_x\beta^{H_B}&=\mathbb{P}_x(H_B=0)+\sum_{\ell=0}^\infty\beta^{\ell+1}\mathbb{P}_x(H_B=\ell+1)\\&\le1+\sum_{\ell=0}^\infty\beta^{\ell+1}\mathbb{P}_x(H_B>\ell)\le1+\beta\sum_{\ell=0}^\infty\beta^\ell\alpha^{\ell/n}\\&\le1+\beta\left(1-\beta\sqrt[n]{\alpha}\right)^{-1}.\end{aligned}$$

Proof 2. From Eq. (5.10) and the assumption in Eq. (5.12),

$$\mathbb{P}_x(H_B>m+n)=\sum_{y\in A}\mathbf{Q}^m_{x,y}\mathbb{P}_y(H_B>n)\le\alpha\sum_{y\in A}\mathbf{Q}^m_{x,y}=\alpha\mathbb{P}_x(H_B>m).$$

Therefore if $\ell=kn+m\in\mathbb{N}_0$ with $k\in\mathbb{N}_0$ and $0\le m<n$, then by induction,

$$\begin{aligned}\mathbb{P}_x(H_B>\ell)&=\mathbb{P}_x(H_B>kn+m)\le\alpha\mathbb{P}_x(H_B>(k-1)n+m)\\&\le\alpha^2\mathbb{P}_x(H_B>(k-2)n+m)\cdots\le\alpha^k\mathbb{P}_x(H_B>m)\le\alpha^k\end{aligned}$$

and hence

$$\mathbb{E}_x H_B = \sum_{\ell=0}^{\infty} \mathbb{P}_x (H_B > \ell) = \sum_{k=0}^{\infty} \sum_{0 \le m < n} \mathbb{P}_x (H_B > kn + m)$$
$$\le \sum_{k=0}^{\infty} \sum_{0 \le m < n} \alpha^k = \frac{n}{1-\alpha} < \infty.$$

■

Corollary 5.25. *If $A = S \setminus B$ is a finite set and $\mathbb{P}_x (H_B = \infty) < 1$ for all $x \in A$, then $\mathbb{E}_\pi H_B < \infty$.*

Proof. Since

$$\mathbb{P}_x (H_B > m) \downarrow \mathbb{P}_x (H_B = \infty) < 1 \text{ for all } x \in A$$

we can find $M_x < \infty$ such that $\mathbb{P}_x (H_B > M_x) < 1$ for all $x \in A$. As A is a finite set

$$n := \max_{x \in A} M_x \in \mathbb{N} \text{ and } \alpha := \max_{x \in A} \mathbb{P}_x (H_B > n) < 1.$$

Corollary 5.24 now applies to complete the proof. ■

Remark 5.26 (First step analysis arguments). We can also compute $\mathbb{P}_x (H_B > n)$ using the first step analysis. (**Please note:** this remark makes use of information developed in Chapter 6 below.) To do this let $u_n (x) := \mathbb{P}_x (H_B > n)$ for $x \in A$ and observe that $\mathbb{P}_x (H_B > 0) = 0$ for $x \in B$ and $u_0 (x) = 1$. The first step analysis then gives,

$$u_{n+1} (x) = \mathbb{P}_x (H_B > n+1) = \sum_{y \in S} p (x, y) \mathbb{P}_x (H_B > n+1 | X_1 = y)$$
$$= \sum_{y \in A} p (x, y) \mathbb{P}_x (H_B > n+1 | X_1 = y) = \sum_{y \in A} p (x, y) \mathbb{P}_y (H_B > n)$$
$$= \sum_{y \in A} p (x, y) u_n (y).$$

Thus we have shown $\mathbf{u}_{n+1} = \mathbf{Q}\mathbf{u}_n$ for all n and therefore $\mathbf{u}_n = \mathbf{Q}^n \mathbf{u}_0 = \mathbf{u}_n = \mathbf{Q}^n \mathbf{1}$ as before.

Similarly using the first step analysis we find,

$$\mathbb{P}_x (H_B = \infty) = \sum_{y \in S} p (x, y) \mathbb{P}_y (H_B = \infty) = \sum_{y \in A} p (x, y) \mathbb{P}_y (H_B = \infty)$$

from which it follows that

$$\mathbb{P}_x (H_B = \infty) = \sum_{y \in S} \mathbf{Q}_{x,y} \mathbb{P}_y (H_B = \infty)$$

or in matrix notation, $\mathbf{u} = \mathbf{Q}\mathbf{u}$ where $\mathbf{u} := \{\mathbb{P}_x (H_B = \infty)\}_{x \in A}$. This equation may be iterated n - times to arrive at Eq. (5.11).

First Step Analysis

The next theorem (which is a special case of Theorem 7.5) is the basis of the first step analysis developed in this section.

Theorem 6.1 (First step analysis). *Let $F(X) = F(X_0, X_1, \dots)$ be some function of the paths $(X_0, X_1, \dots)$ of our Markov chain, then for all $x, y \in S$ with $p(x,y) > 0$ we have*

$$\mathbb{E}_x [F(X_0, X_1, \dots) | X_1 = y] = \mathbb{E}_y [F(x, X_0, X_1, \dots)] \tag{6.1}$$

and

$$\begin{aligned}\mathbb{E}_x [F(X_0, X_1, \dots)] &= \sum_{y \in S} p(x,y) \mathbb{E}_y [F(x, X_0, X_1, \dots)] \\ &= \mathbb{E}_{p(x,\cdot)} [F(x, X_0, X_1, \dots)]\end{aligned} \tag{6.2}$$

Proof. Equation (6.1) follows directly from Theorem 7.5 below,

$$\begin{aligned}\mathbb{E}_x [F(X_0, X_1, \dots) | X_1 = y] &= \mathbb{E}_x [F(X_0, X_1, \dots) | X_0 = x, X_1 = y] \\ &= \mathbb{E}_y [F(x, X_0, X_1, \dots)].\end{aligned}$$

Equation (6.2) now follows from Eq. (6.1), the law of total expectation (i.e. $\mathbb{E}_x [F(X_0, X_1, \dots)] = \mathbb{E}_x \mathbb{E} [F(X_0, X_1, \dots) | X_1]$), and the fact that $\mathbb{P}_x (X_1 = y) = p(x,y)$.

Alternative direct proof. For this proof we use the measure theoretic fact that it suffice to prove the theorem under the restriction that $F(X_0, X_1, \dots)$ is actually of the form,

$$F(X_0, X_1, \dots) = F(X_0, X_1, \dots, X_n) \text{ for some } n < \infty.$$

Under this additional assumption,

$$\begin{aligned}&\mathbb{E}_x [F(X_0, X_1, \dots, X_n) 1_{X_1 = y}] \\ &= \sum_{x_1, \dots, x_n \in S} F(x, x_1, \dots, x_n) 1_{x_1 = y} p(x, x_1) p(x_1, x_2) \dots p(x_{n-1}, x_n) \\ &= p(x,y) \sum_{x_2, \dots, x_n \in S} F(x, y, x_2, \dots, x_n) p(y, x_2) \dots p(x_{n-1}, x_n) \\ &= p(x,y) \sum_{x_1, \dots, x_{n-1} \in S} F(x, y, x_1, \dots, x_{n-1}) p(y, x_1) \dots p(x_{n-2}, x_{n-1}) \\ &= \mathbb{P}_x (X_1 = y) \mathbb{E}_y F(x, X_0, \dots, X_{n-1})\end{aligned}$$

which upon dividing by $\mathbb{P}_x (X_1 = y)$ gives Eq. (6.1). ■

Let us now suppose for until further notice that B is a non-empty proper subset of S, $A = S \setminus B$, and $H_B = H_B(X)$ is the first hitting time of B by X.

Notation 6.2 *Given a transition matrix $\mathbf{P} = (p(x,y))_{x,y \in S}$ we let $\mathbf{Q} = \mathbf{P}_{A \times A} = (p(x,y))_{x,y \in A}$ and $\mathbf{R} = \mathbf{P}_{A \times B} = (p(x,y))_{x \in A, y \in B}$ so that, schematically,*

$$\mathbf{P} = \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} \mathbf{P}_{A \times A} & \mathbf{P}_{A \times B} \\ * & * \end{bmatrix} \end{array} \begin{array}{c} A \\ B \end{array} = \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ * & * \end{bmatrix} \end{array} \begin{array}{c} A \\ B \end{array}.$$

Remark 6.3. To construct the matrix $\mathbf{Q}$ and $\mathbf{R}$ from $\mathbf{P}$, let $\mathbf{P}'$ be $\mathbf{P}$ with the rows corresponding to B omitted. To form $\mathbf{Q}$ from $\mathbf{P}'$, remove the columns of $\mathbf{P}'$ corresponding to B and to form $\mathbf{R}$ from $\mathbf{P}'$, remove the columns of $\mathbf{P}'$ corresponding to A.

Example 6.4. If $S = \{1,2,3,4,5,6,7\}$, $A = \{1,2,4,5,6\}$, $B = \{3,7\}$, and

$$\mathbf{P} = \left[\begin{array}{c|ccccccc} r \backslash c & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 2 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 3 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 4 & 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 5 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 6 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right],$$

then

$$\mathbf{P}' = \left[\begin{array}{c|ccccccc} r \backslash c & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 2 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 4 & 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 5 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 6 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \end{array}\right].$$

Deleting the 3 and 7 columns of $\mathbf{P}'$ gives

$$\mathbf{Q} = \mathbf{P}_{A\times A} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}$$

and deleting the $1, 2, 4, 5$, and 6 columns of $\mathbf{P}'$ gives

$$\mathbf{R} = \mathbf{P}_{A\times B} = \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Theorem 6.5 (Hitting distributions). *Let $h : B \to \mathbb{R}$ be a bounded or non-negative function and let $u : S \to \mathbb{R}$ be defined by*

$$u(x) := \mathbb{E}_x \left[h\left(X_{H_B}\right) : H_B < \infty \right] \text{ for } x \in S.$$

Then $u = h$ on B and $u|_A$ satisfies,

$$u(x) = \sum_{y\in A} p(x,y)\, u(y) + \sum_{y\in B} p(x,y)\, h(y) \text{ for all } x \in A. \tag{6.3}$$

In matrix notation this becomes (where $\mathbf{u} := u|_A$),

$$\mathbf{u} = \mathbf{Q}u + \mathbf{R}\mathbf{h} \implies \mathbf{u} = (I - \mathbf{Q})^{-1}\mathbf{R}\mathbf{h},$$

i.e. for all $x \in A$ we have[1]

$$\mathbb{E}_x \left[h\left(X_{H_B}\right) : H_B < \infty \right] = \left[(I - \mathbf{Q})^{-1}\mathbf{R}\mathbf{h} \right]_x = \left[(I - \mathbf{P}_{A\times A})^{-1}\mathbf{P}_{A\times B}\mathbf{h} \right]_x. \tag{6.4}$$

As a special case, if $y \in B$ and $h(s) = \delta_y(s)$, then Eq. (6.4) becomes,

$$\mathbb{P}_x \left(X_{H_B} = y : H_B < \infty \right) = \left[(I - \mathbf{Q})^{-1}\mathbf{R} \right]_{x,y} = \left[(I - \mathbf{P}_{A\times A})^{-1}\mathbf{P}_{A\times B}\delta_y \right]_x \tag{6.5}$$

[1] It is possible that $I - \mathbf{Q}$ is not invertible. In this case it turns out that (at least when $h \geq 0$) that the correct interpretation of $(I - \mathbf{Q})^{-1}$ is;

$$(I - \mathbf{Q})^{-1} = \sum_{n=0}^{\infty} \mathbf{Q}^n.$$

More generally if $B_0 \subset B$ and $h(s) := 1_{B_0}(s)$ we learn that

$$\mathbb{P}_x \left(X_{H_B} \in B_0 : H_B < \infty \right) = \left[(I - \mathbf{P}_{A\times A})^{-1}\mathbf{P}_{A\times B} 1_{B_0} \right]_x. \tag{6.6}$$

Proof. Let

$$F(X_0, X_1, \dots) = h\left(X_{H_B(X)}\right) 1_{H_B(X)<\infty},$$

then for $x \in A$ we have $F(x, X_0, X_1, \dots) = F(X_0, X_1, \dots)$. Therefore by the first step analysis (Theorem 6.1) we learn

$$\begin{aligned} u(x) &= \mathbb{E}_x \left[h\left(X_{H_B(X)}\right) 1_{H_B(X)<\infty} \right] = \mathbb{E}_x F(X_0, X_1, \dots) \\ &= \mathbb{E}_x F(x, X_1, \dots) = \sum_{y\in S} p(x,y)\, \mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \sum_{y\in S} p(x,y)\, \mathbb{E}_y F(X_0, X_1, \dots) \\ &= \sum_{y\in S} p(x,y)\, \mathbb{E}_y \left[h\left(X_{H_B(X)}\right) 1_{H_B(X)<\infty} \right] \\ &= \sum_{y\in A} p(x,y)\, u(y) + \sum_{y\in B} p(x,y)\, h(y). \end{aligned}$$

wherein the last equality we have used,

$$\mathbb{E}_y \left[h\left(X_{H_B(X)}\right) 1_{H_B(X)<\infty} \right] = \begin{cases} u(y) \text{ if } y \in A \\ h(y) \text{ if } y \in B \end{cases}.$$

■

Theorem 6.6 (Travel averages). *Given $g : A \to [0, \infty]$, let $w(x) := \mathbb{E}_x \left[\sum_{n<H_B} g(X_n) \right]$. Then $w(x)$ satisfies*

$$w(x) = \sum_{y\in A} p(x,y)\, w(y) + g(x) \text{ for all } x \in A. \tag{6.7}$$

In matrix notation this becomes,

$$\mathbf{w} = \mathbf{Q}\mathbf{w} + \mathbf{g} \implies \mathbf{w} = (I - \mathbf{Q})^{-1}\mathbf{g}$$

so that

$$\mathbb{E}_x \left[\sum_{n<H_B} g(X_n) \right] = \left[(I - \mathbf{Q})^{-1}\mathbf{g} \right]_x = \left[(I - \mathbf{P}_{A\times A})^{-1}\mathbf{g} \right]_x$$

The following two special cases are of most interest;

1. *Suppose* $g(x) = \delta_y(x)$ *for some* $y \in A$, *then* $\sum_{n<H_B} g(X_n) = \sum_{n<H_B} \delta_y(X_n)$ *is the number of visits of the chain to* y *and*

$$\begin{aligned}\mathbb{E}_x(\#\text{ visits to } y \text{ before hitting } B) \\ = \mathbb{E}_x\left[\sum_{n<H_B} \delta_y(X_n)\right] = (I-\mathbf{Q})^{-1}_{x,y}. \end{aligned} \tag{6.8}$$

2. *Suppose that* $g(x) = 1$, *then* $\sum_{n<H_B} g(X_n) = H_B$ *and we may conclude that*

$$\mathbb{E}_x[H_B] = \left[(I-\mathbf{Q})^{-1}\mathbf{1}\right]_x \tag{6.9}$$

where $\mathbf{1}$ *is the column vector consisting of all ones.*

Proof. Let $F(X_0, X_1, \dots) = \sum_{n<H_B(X_0,X_1,\dots)} g(X_n)$ be the sum of the values of g along the chain before its first exit from A, i.e. entrance into B. With this interpretation in mind, if $x \in A$, it is easy to see that

$$\begin{aligned} F(x, X_0, X_1, \dots) &= \begin{cases} g(x) & \text{if } X_0 \in B \\ g(x) + F(X_0, X_1, \dots) & \text{if } X_0 \in A \end{cases} \\ &= g(x) + 1_{X_0 \in A} \cdot F(X_0, X_1, \dots). \end{aligned}$$

Therefore by the first step analysis (Theorem 6.1) it follows that

$$\begin{aligned} w(x) = \mathbb{E}_x F(X_0, X_1, \dots) &= \sum_{y\in S} p(x,y)\, \mathbb{E}_y F(x, X_0, X_1, \dots) \\ &= \sum_{y\in S} p(x,y)\, \mathbb{E}_y\left[g(x) + 1_{X_0\in A} \cdot F(X_0, X_1, \dots)\right] \\ &= g(x) + \sum_{y\in A} p(x,y)\, \mathbb{E}_y\left[F(X_0, X_1, \dots)\right] \\ &= g(x) + \sum_{y\in A} p(x,y)\, w(y). \end{aligned}$$

■

Remark 6.7. We may combine Theorems 6.1 and 6.6 into one theorem as follows. Suppose that $h : S \to \mathbb{R}$ is a given function and let

$$w(x) := \mathbb{E}_x\left[\sum_{n<H_B} h(X_n) + h(X_{H_B})\right],$$

then for $w(x) = h(x)$ for $x \in B$ and

$$w(x) = \sum_{y\in A} p(x,y)\, w(y) + h(x) + \sum_{y\in B} p(x,y)\, h(y) \text{ for } x \in A.$$

In matrix format we have

$$\mathbf{w} = \mathbf{P}_{A\times A}\mathbf{w} + \mathbf{h}_A + \mathbf{P}_{A\times B}\mathbf{h}_B$$

where $\mathbf{h}_A = \{h(x)\}_{x\in A}$ and $\mathbf{h}_B = \{h(x)\}_{x\in B}$.

6.1 Finite state space examples

Example 6.8. Consider the Markov chain determined by

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left[\begin{array}{cccc} 0 & 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

whose hitting diagram is given in Figure 6.1.Notice that 3 and 4 are absorbing

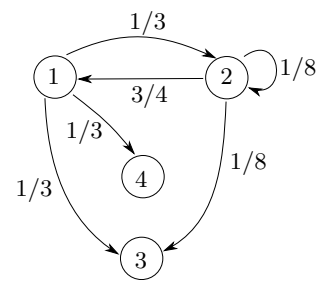


Fig. 6.1. For this chain the states 3 and 4 are absorbing.

states. Taking $A = \{1,2\}$ and $B = \{3,4\}$, we find

$$\mathbf{P}' = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left[\begin{array}{cccc} 0 & 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \end{array},$$

$$\mathbf{Q} = \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left[\begin{array}{cc} 0 & 1/3 \\ 3/4 & 1/8 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \end{array}, \text{ and } \mathbf{R} = \begin{array}{c} \begin{array}{cc} 3 & 4 \end{array} \\ \left[\begin{array}{cc} 1/3 & 1/3 \\ 1/8 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \end{array}.$$

Matrix manipulations now show,

$\mathbb{E}_i\left(\#\text{ visits to } j \text{ before hitting } \{3,4\}\right)$

$$= (I-\mathbf{Q})^{-1} = \begin{array}{c|cc} i\backslash j & 1 & 2 \\ \hline 1 & \frac{7}{5} & \frac{8}{15} \\ 2 & \frac{6}{5} & \frac{8}{5} \end{array} = \begin{bmatrix} 1.4 & 0.53333 \\ 1.2 & 1.6 \end{bmatrix},$$

$$\mathbb{E}_i H_{\{3,4\}} = (I-\mathbf{Q})^{-1}\begin{bmatrix}1\\1\end{bmatrix} = \begin{array}{c|c} i & \\ \hline 1 & \frac{29}{15} \\ 2 & \frac{14}{5} \end{array} = \begin{bmatrix} 1.933\,3 \\ 2.8 \end{bmatrix} \text{ and}$$

$$P_i\left(X_{H_{\{3,4\}}} = j\right) = (I-\mathbf{Q})^{-1}\mathbf{R} = \begin{array}{c|cc} i\backslash j & 3 & 4 \\ \hline 1 & \frac{8}{15} & \frac{7}{15} \\ 2 & \frac{3}{5} & \frac{2}{5} \end{array} = \begin{bmatrix} 0.534 & 0.467 \\ 0.6 & 0.4 \end{bmatrix}.$$

The output of one simulation from `www.zweigmedia.com/RealWorld/markov/markov.html` is in Figure 6.2 below.

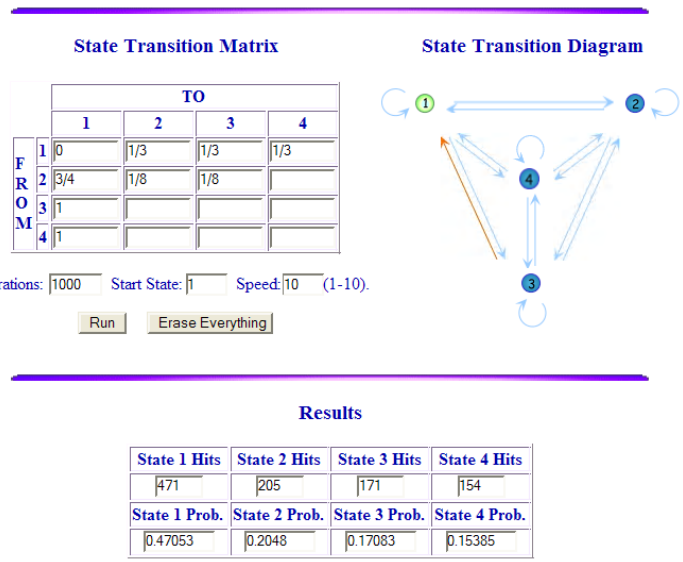


Fig. 6.2. In this run, rather than making sites 3 and 4 absorbing, we have made them transition back to 1. I claim now to get an approximate value for $P_1\left(X_n \text{ hits } 3\right)$ we should compute: (State 3 Hits)/(State 3 Hits + State 4 Hits). In this example we will get $171/(171+154) = 0.526\,15$ which is a little lower than the predicted value of 0.533. You can try your own runs of this simulator.

Remark 6.9. As an aside in the above simulation we really used the matrix,

$$\hat{\mathbf{P}} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

for which has invariant distribution,

$$\begin{aligned} \pi &= \frac{1}{21+8+8+7}\left[21\ 8\ 8\ 7\right] \\ &= \left[\tfrac{21}{44}\ \tfrac{2}{11}\ \tfrac{2}{11}\ \tfrac{7}{44}\right] \\ &= \left[0.477\ 0.182\ 0.182\ 0.159\right]. \end{aligned}$$

Notice that

$$\frac{2}{11}\Big/\left(\frac{2}{11}+\frac{7}{44}\right) = \frac{8}{15}!$$

Lemma 6.10. *Suppose that B is a subset of S and $x \in A := S \setminus B$. Then*

$$\mathbb{P}_x\left(H_B < \infty\right) = \left[\left(I - \mathbf{P}_{A\times A}\right)^{-1}\mathbf{P}_{A\times B}\mathbf{1}\right]_x$$

where

$$\left(I-\mathbf{P}_{A\times A}\right)^{-1} := \sum_{n=0}^{\infty}\mathbf{P}_{A\times A}^n.$$

[See the optional Section 6.6 below for more analysis of this type.]

Proof. We work this out by first principles,

$$\begin{aligned} \mathbb{P}_x\left(H_B<\infty\right) &= \sum_{n=1}^{\infty}\mathbb{P}_x\left(H_B = n\right) = \sum_{n=1}^{\infty}\mathbb{P}_x\left(X_1\in A,\dots,X_{n-1}\in A, X_n\in B\right) \\ &= \sum_{n=1}^{\infty}\sum_{x_1,\dots,x_{n-1}\in A}\sum_{y\in B} p\left(x,x_1\right)p\left(x_1,x_2\right)\dots p\left(x_{n-2},x_{n-1}\right)p\left(x_{n-1},y\right) \\ &= \left[\sum_{n=1}^{\infty}\mathbf{P}_{A\times A}^{n-1}\mathbf{P}_{A\times B}\mathbf{1}\right]_x = \left[\sum_{n=0}^{\infty}\mathbf{P}_{A\times A}^{n}\mathbf{P}_{A\times B}\mathbf{1}\right]_x. \end{aligned}$$

■

Example 6.11. Let us continue the rat in the maze Exercise 5.1 and now suppose that room 3 contains food while room 7 contains a mouse trap.

$$\begin{bmatrix} 1 & 2 & 3\ (\text{food}) \\ 4 & 5 & 6 \\ 7\ (\text{trap}) & & \end{bmatrix}.$$

Recall that the transition matrix for this chain with sites 3 and 7 absorbing is given by,

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array},$$

see Figure 6.3 for the corresponding jump diagram for this chain.

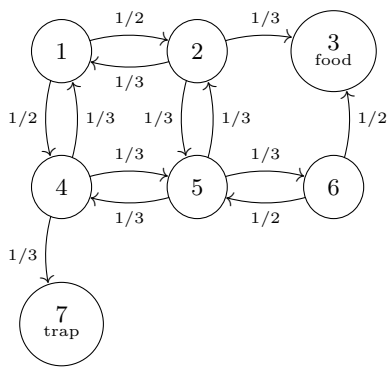


Fig. 6.3. The jump diagram for our proverbial rat in the maze. Here we assume the rat is "absorbed" at sites 3 and 7

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 $T := H_B$ be the first hitting time of B. Then deleting the 3 and 7 rows of $\mathbf{P}$ leaves the matrix,

$$\mathbf{P}' = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Deleting the 3 and 7 columns of $\mathbf{P}'$ gives

$$\mathbf{Q} = \mathbf{P}_{A\times A} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}$$

and deleting the $1,2,4,5$, and 6 columns of $\mathbf{P}'$ gives

$$\mathbf{R} = \mathbf{P}_{A\times B} = \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Therefore,

$$I - \mathbf{Q} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix},$$

and using a computer algebra package we find

$$\mathbb{E}_i\left[\#\text{ visits to } j \text{ before hitting } \{3,7\}\right] = (I-\mathbf{Q})^{-1} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} j \\ \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} \end{array} \begin{array}{c} \\ i \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

In particular we may conclude,

$$\begin{bmatrix} \mathbb{E}_1 T \\ \mathbb{E}_2 T \\ \mathbb{E}_4 T \\ \mathbb{E}_5 T \\ \mathbb{E}_6 T \end{bmatrix} = (I-\mathbf{Q})^{-1}\mathbf{1} = \begin{bmatrix} \frac{17}{3} \\ \frac{14}{3} \\ \frac{14}{3} \\ \frac{16}{3} \\ \frac{11}{3} \end{bmatrix},$$

and

$$\begin{bmatrix} \mathbb{P}_1(X_T=3) & \mathbb{P}_1(X_T=7) \\ \mathbb{P}_2(X_T=3) & \mathbb{P}_2(X_T=3) \\ \mathbb{P}_4(X_T=3) & \mathbb{P}_4(X_T=3) \\ \mathbb{P}_5(X_T=3) & \mathbb{P}_5(X_T=3) \\ \mathbb{P}_6(X_T=3) & \mathbb{P}_6(X_T=7) \end{bmatrix} = (I-\mathbf{Q})^{-1}\mathbf{R} = \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Since the event of hitting 3 before 7 is the same as the event $\{X_T = 3\}$, the desired hitting probabilities are

$$\begin{bmatrix} \mathbb{P}_1(X_T=3) \\ \mathbb{P}_2(X_T=3) \\ \mathbb{P}_4(X_T=3) \\ \mathbb{P}_5(X_T=3) \\ \mathbb{P}_6(X_T=3) \end{bmatrix} = \begin{bmatrix} \frac{7}{12} \\ \frac{3}{4} \\ \frac{5}{12} \\ \frac{2}{3} \\ \frac{5}{6} \end{bmatrix}.$$

Example 6.12 (A modified rat maze). Here is the modified maze,

$$\begin{bmatrix} 1 & 2 & 3(\text{food}) \\ 4 & 5 & \\ 6(\text{trap}) & & \end{bmatrix}.$$

We now let $T = H_{\{3,6\}}$ be the first time to absorption – we assume that 3 and 6 made are absorbing states.[2] The transition matrix is given by

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}.$$

The corresponding $\mathbf{Q}$ and $\mathbf{R}$ matrices in this case are;

$$\mathbf{Q} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \end{array}, \text{ and } \qquad \mathbf{R} = \begin{array}{c} \begin{array}{cc} 3 & 6 \end{array} \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \end{array}.$$

[2] It is not necessary to make states 3 and 6 absorbing. In fact it does matter at all what the transition probabilities are for the chain for leaving either of the states 3 or 6 since we are going to stop when we hit these states. This is reflected in the fact that the first thing we will do in the first step analysis is to delete rows 3 and 6 from P. Making 3 and 6 absorbing simply saves a little ink.

After some matrix manipulation we then learn,

$$\mathbb{E}_i[\#\text{ visits to } j] = (I_4-\mathbf{Q})^{-1} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 4 & 5 \end{array} \\ \begin{bmatrix} 2 & \frac{3}{2} & \frac{3}{2} & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & \frac{3}{2} & \frac{3}{2} & 2 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \end{array},$$

$$\mathbb{P}_i[X_T = j] = (I_4-\mathbf{Q})^{-1}\mathbf{R} = \begin{array}{c} \begin{array}{cc} 3 & 6 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \end{array},$$

$$\mathbb{E}_i[T] = (I_4-\mathbf{Q})^{-1}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ 5 \\ 6 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \end{array}.$$

So for example, $\mathbb{P}_4(X_T = 3(\text{food})) = 1/3$, $\mathbb{E}_4(\text{Number of visits to } 1) = 1$, $\mathbb{E}_5(\text{Number of visits to } 2) = 3/2$ and $\mathbb{E}_1 T = \mathbb{E}_5 T = 6$ and $\mathbb{E}_2 T = \mathbb{E}_4 T = 5$.

Exercise 6.1 (Invariant distributions and expected return times). Suppose that $\{X_n\}_{n=0}^\infty$ is a Markov chain on a finite state space S determined by one step transition probabilities, $p(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x)$. For $x \in S$, let $R_x := \inf\{n > 0 : X_n = x\}$ be the first **passage time**[3] to x. [We will assume here that $\mathbb{E}_y R_x < \infty$ for all $x, y \in S$.] Use the first step analysis to show,

$$\mathbb{E}_x R_y = \sum_{z: z \neq y} p(x,z)\,\mathbb{E}_z R_y + 1. \tag{6.10}$$

Now further assume that $\pi : S \to [0,1]$ is an invariant distributions for p, that is $\sum_{x\in S}\pi(x) = 1$ and $\sum_{x\in S}\pi(x)\,p(x,y) = \pi(y)$ for all $y \in S$, i.e. $\pi\mathbf{P} = \pi$. By multiplying Eq. (6.10) by $\pi(x)$ and summing on $x \in S$, show,

$$\pi(y)\,\mathbb{E}_y R_y = 1 \text{ for all } y \in S \implies \pi(y) = \frac{1}{\mathbb{E}_y R_y} > 0. \tag{6.11}$$

Example 6.13 (Two heads in a row). On average, how many times do we need to toss a coin to get two consecutive heads? To answer this question, we need to construct an appropriate Markov chain. There are different ways to approach this; probably the simplest is to let

$$X_n = 2 \wedge \left(\#\text{ of heads tossed in a row after the } n^{\text{th}}\text{ -toss}\right).$$

[3] R_x is the first **return time** to x when the chain starts at x.

Here is an example of how X_n is computed;

Time	0	1	2	3	4	5	6	7	8	9	10	11	...
Tosses		H	T	H	T	T	H	H	T	H	H	H	...
$\mathbb{X}$	0	1	0	1	1	0	1	2	0	1	2	2	...

This means X_n is a Markov chain with state space $\{0,1,2\}$ and transition matrix

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \end{array}.$$

We then take $B = \{2\}$ and we wish to compute

$$\mathbb{E}_0 H_B = \left[(I - \mathbf{Q})^{-1} \mathbf{1} \right]_0$$

where

$$\mathbf{Q} = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \end{array}.$$

The linear algebra gives,

$$(I - \mathbf{Q})^{-1} = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \end{array} \text{ and } (I - \mathbf{Q})^{-1} \mathbf{1} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array}$$

and hence $\mathbb{E}_0 H_B = 6$.

Remark 6.14. It is unimportant how we make assignments in the bottom row of $\mathbf{P}$ in Example 6.13 since when $B = \{2\}$ we will immediately delete this row in the computations above. So we could have made 2 an absorbing state, i.e.

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \end{array}$$

Actually a more interesting choice is to send 2 back to 0 and restart the chain, i.e. take

$$\hat{\mathbf{P}} = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \end{array}.$$

This represents the chain where we restart the game once we have reached two heads in a row. The invariant distribution for $\hat{\mathbf{P}}$ is

$$\pi = \frac{1}{7} \begin{bmatrix} 4 & 2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{7} & \frac{2}{7} & \frac{1}{7} \end{bmatrix}.$$

Note that on average we expect 4 visits to 0 and 2 visits to 2 for every time the game is played and hence the expected number of times to get two heads in a row is $4 + 2 = 6$ in agreement with Example 6.13. (Alternatively we have $1/7 = 1/\mathbb{E}_2 R_2$ where R_2 is the return time to 2. Thus $\mathbb{E}_2 R_2 = 7$. Now $R_2 = 1 + H$ above and hence we see that $\mathbb{E}H = 7 - 1 = 6$ as we saw before.) Let us further

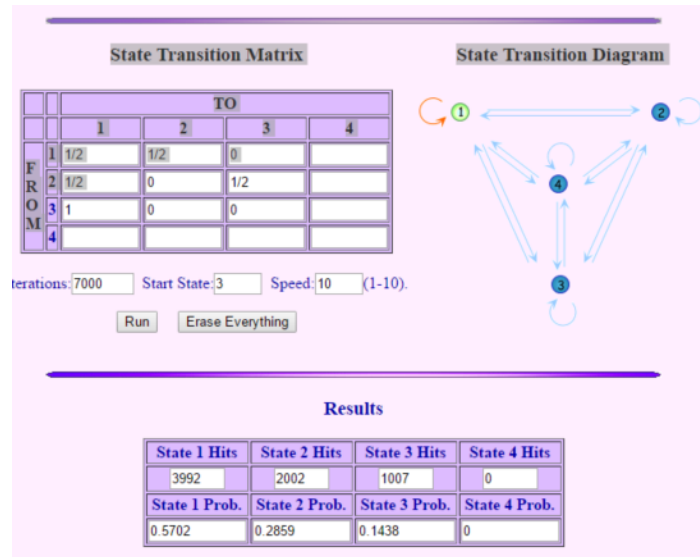


Fig. 6.4. Here is a numerical verification of the above statements. In this case the chain was run for $N = 7000$ and in that time there were 1007 returns to 3 (corresponds to site 2) for an expected return time of about 7.

observe that

$$\hat{\mathbf{P}}^{100} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix}^{100} = \begin{bmatrix} 0.571\,43 & 0.285\,71 & 0.142\,86 \\ 0.571\,43 & 0.285\,71 & 0.142\,86 \\ 0.571\,43 & 0.285\,71 & 0.142\,86 \end{bmatrix}$$

while

$$\pi = \begin{bmatrix} 0.571\,43 & 0.285\,71 & 0.142\,86 \end{bmatrix}.$$

6.2 More first step analysis examples

Example 6.15 (Computing $\mathbb{E}H_B^2$). Let $H_B = \inf\{n \geq 0 : X_n \in B\}$. Here we compute $w(x) = \mathbb{E}_x H_B^2$ as follows,

$$\begin{aligned} w(x) &= \sum_{y\in S} p(x,y)\,\mathbb{E}_x\left[H_B^2 | X_1 = y\right] = \sum_{y\in S} p(x,y)\,\mathbb{E}_y\left[(H_B+1)^2\right] \\ &= \sum_{y\in S} p(x,y)\,\mathbb{E}_y\left[H_B^2 + 2H_B + 1\right] = 1 + \sum_{y\in A} p(x,y)\,\mathbb{E}_y\left[H_B^2 + 2H_B\right] \\ &= 1 + \sum_{y\in A} p(x,y)\,w(y) + 2\sum_{y\in A} p(x,y)\,\mathbb{E}_y H_B. \end{aligned}$$

Thus in matrix notation with $\mathbf{Q} = \mathbf{P}_{A\times A}$ this may be written as,

$$\mathbb{E}_{(\cdot)} H_B^2 = \mathbf{w} = (I-\mathbf{Q})^{-1}\left[\mathbf{1} + 2\mathbf{Q}(I-\mathbf{Q})^{-1}\mathbf{1}\right]$$

wherein we have used $\mathbb{E}_{(\cdot)} H_B = (I-\mathbf{Q})^{-1}\mathbf{1}$. In other words,

$$\begin{aligned} \mathbb{E}_{(\cdot)} H_B^2 = \mathbf{w} &= (I-\mathbf{Q})^{-1}\left[I + 2\mathbf{Q}(I-\mathbf{Q})^{-1}\right]\mathbf{1} \\ &= (I-\mathbf{Q})^{-2}(I+\mathbf{Q})\mathbf{1}. \end{aligned}$$

Example 6.16 (Generating function for H_B). Let us now suppose that $|z| < 1$ and let us set $w(x) := \mathbb{E}_x\left[z^{H_B}\right]$ and notice that $w(y) = 1$ for $y \in B$. Then by the first step analysis we have,

$$\begin{aligned} w(x) &:= \sum_{y\in S} p(x,y)\,\mathbb{E}_x\left[z^{H_B} | X_1 = y\right] = \sum_{y\in S} p(x,y)\,\mathbb{E}_y\left[z^{H_B+1}\right] \\ &= z\sum_{y\in S} p(x,y)\,\mathbb{E}_y\left[z^{H_B}\right] = z\sum_{y\in S} p(x,y)\,w(y). \end{aligned}$$

So let $\mathbf{w} := w|_A$, we find,

$$\begin{aligned} \mathbf{w} &= z\left[\mathbf{Q}\mathbf{w} + p(\cdot,B)\right] = z\left[\mathbf{Q}\mathbf{w} + \mathbf{1} - p(\cdot,A)\right] \\ &= z\left[\mathbf{Q}\mathbf{w} + \mathbf{1} - \mathbf{Q}\mathbf{1}\right]. \end{aligned}$$

Thus solving this equation for $\mathbf{w}$ gives,

$$\mathbb{E}_{(\cdot)}\left[z^{H_B}\right] = \mathbf{w} = z(I - z\mathbf{Q})^{-1}(I-\mathbf{Q})\mathbf{1}.$$

Differentiating this equation with respect to z at $z = 1$ shows

$$\begin{aligned} \mathbb{E}_{(\cdot)}[H_B] &= \mathbf{1} + \frac{d}{dz}|_1 (I - z\mathbf{Q})^{-1}(I-\mathbf{Q})\mathbf{1} \\ &= \mathbf{1} + (I-\mathbf{Q})^{-1}\mathbf{Q}(I-\mathbf{Q})^{-1}(I-\mathbf{Q})\mathbf{1} \\ &= \left[I + (I-\mathbf{Q})^{-1}\mathbf{Q}\right]\mathbf{1} = (I-\mathbf{Q})^{-1}\mathbf{1} \end{aligned}$$

which reproduces are old formula for $\mathbb{E}_{(\cdot)}[H_B]$.

Example 6.17 (Generating function for $\sum_{n<H_B} g(X_n)$). Suppose that $g : A \to \mathbb{R}$ is a given function, $F := \sum_{n<H_B} g(X_n)$, $|z| \leq 1$, and let

$$w(x) := \mathbb{E}_x z^F.$$

Then by the first step analysis,

$$w(x) = \sum_{y\in S} p(x,y)\,\mathbb{E}_y z^{F(x,X_0,X_1,\dots)}$$

where

$$\begin{aligned} &\mathbb{E}_y z^{F(x,X_0,X_1,\dots)} = z^{g(x)} \text{ if } y \in B \text{ and} \\ &\mathbb{E}_y z^{F(x,X_0,X_1,\dots)} = z^{g(x)}\mathbb{E}_y z^{F(X_0,X_1,\dots)} = z^{g(x)} w(y) \text{ if } y \in A. \end{aligned}$$

Thus we learn that w satisfies,

$$w(x) = z^{g(x)}\left[\sum_{y\in A}\mathbf{Q}_{xy} w + \sum_{y\in B}\mathbf{P}_{xy} 1\right].$$

Thus let $D(z) = \operatorname{diag}\left(z^{g(x)} : x \in A\right)$, we have

$$\mathbf{w} = D(z)\mathbf{Q}\mathbf{w} + D(z)\mathbf{R}\mathbf{1} \tag{6.12}$$

and hence

$$\mathbf{w} = (I - D(z)\mathbf{Q})^{-1} D(z)\mathbf{R}\mathbf{1}.$$

Sanity checks.

1. If $z = 1$ then $\mathbf{w} = \mathbf{1}$ which solves Eq. (6.12) since $\mathbf{Q}\mathbf{1} + \mathbf{R}\mathbf{1} = \mathbf{P}_{A\times S}\mathbf{1} = \mathbf{1}$ since the row sums of $\mathbf{P}$ add to 1. Consequently we now know that

$$(I - Q)\mathbf{1} = \mathbf{R}\mathbf{1} \implies (I-\mathbf{Q})^{-1}\mathbf{R}\mathbf{1} = \mathbf{1}.$$

2. Suppose that we differentiate this expression in z one time,

$$\begin{aligned}\frac{d}{dz}|_{z=1}\mathbf{w} &= (I-\mathbf{Q})^{-1} D'(1)\,\mathbf{Q}\,(I-\mathbf{Q})^{-1}\mathbf{R1} + (I-\mathbf{Q})^{-1} D'(1)\,\mathbf{R1} \\ &= (I-\mathbf{Q})^{-1} D'(1)\,\mathbf{Q1} + (I-\mathbf{Q})^{-1} D'(1)\,\mathbf{R1} \\ &= (I-\mathbf{Q})^{-1} D'(1)\,[\mathbf{Q1}+\mathbf{R1}] = (I-\mathbf{Q})^{-1} D'(1)\,\mathbf{1} \\ &= (I-\mathbf{Q})^{-1}\mathbf{g}\end{aligned}$$

which is consistent with the fact that

$$\frac{d}{dz}|_{z=1}\mathbf{w} = \mathbb{E}_{(\cdot)}\left[\sum_{n<H_B} g(X_n)\right]$$

and our previous computations.

Exercise 6.2 (III.4.P11 on p.132). An urn contains two red and two green balls. The balls are chosen at random, one by one, and removed from the urn. The selection process continues until all of the green balls have been removed from the urn. What is the probability that a single red ball is in the urn at the time that the last green ball is chosen?

Theorem 6.18. *Let $h : B \to [0,\infty]$ and $g : A \to [0,\infty]$ be given and for $x \in S$. If we let* [4]

$$\begin{aligned} w(x) &:= \mathbb{E}_x\left[h(X_{H_B})\cdot\sum_{n<H_B} g(X_n) : H_B<\infty\right] \text{ and} \\ g_h(x) &= g(x)\,\mathbb{E}_x\left[h(X_{H_B}) : H_B<\infty\right],\end{aligned}$$

then

$$w(x) = \mathbb{E}_x\left[\sum_{n<H_B} g_h(X_n) : H_B<\infty\right]. \tag{6.13}$$

Remark 6.19. Recall that we can find $u_h(x) := \mathbb{E}_x[h(X_{H_B}) : H_B<\infty]$ using Theorem 6.5 and then we can solve for $w(x)$ using Theorem 6.6 with g replaced by $g_h(x) = g(x)\,u_h(x)$. So in the matrix language we solve for $w(x)$ as follows;

$$\begin{aligned}\mathbf{u}_h &:= (I-\mathbf{Q})^{-1}\mathbf{Rh}, \\ \mathbf{g_h} &:= \mathbf{g} * \mathbf{u}_h, \text{ and} \\ \mathbf{w} &= (I-\mathbf{Q})^{-1}\mathbf{g}_h,\end{aligned}$$

where $[\mathbf{a}*\mathbf{b}]_x := \mathbf{a}_x \cdot \mathbf{b}_x$ – the entry by entry product of column vectors.

[4] Recall from Theorem 6.5 that $\mathbf{u}_h = (I-\mathbf{Q})^{-1}\mathbf{Rh}$, i.e. $u = h$ on B and u satisfies

$$u(x) = \sum_{y\in A} p(x,y)\,u(y) + \sum_{y\in B} p(x,y)\,h(y) \text{ for all } x\in A.$$

Proof. First proof. Let $H(X) := h(X_{H_B})\,1_{H_B<\infty}$, then using $1_{n<H_B(X)} = 1_{X_0\in A,\dots,X_n\in A}$ and

$$H(X_0,\dots,X_{n-1},X_n,\dots) = H(X_n,X_{n+1},\dots) \text{ when } X_0,\dots,X_n\in A$$

along with the strong Markov property in Theorem 7.11 shows;

$$\begin{aligned} w(x) &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[H(X)\cdot 1_{n<H_B} g(X_n)\right] \\ &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[H(X)\cdot 1_{X_0\in A,\dots,X_n\in A} g(X_n)\right] \\ &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[\mathbb{E}^{(Y)}_{X_n}\left[H(X_0,\dots,X_{n-1},Y)\right]\cdot 1_{X_0\in A,\dots,X_n\in A} g(X_n)\right] \\ &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[\mathbb{E}^{(Y)}_{X_n} H(Y)\cdot 1_{X_0\in A,\dots,X_n\in A} g(X_n)\right] \\ &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[u_h(X_n)\cdot 1_{X_0\in A,\dots,X_n\in A} g(X_n)\right] \\ &= \sum_{n=0}^{\infty}\mathbb{E}_x\left[u_h(X_n)\cdot g(X_n)\,1_{n<H_B}\right] \\ &= \mathbb{E}_x\left[\sum_{n<H_B} u_h(X_n)\cdot g(X_n)\right] = \mathbb{E}_x\left[\sum_{n<H_B} g_h(X_n)\right].\end{aligned}$$

Second proof. Let $G(X) := \sum_{n<H_B} g(X_n)$ and observe that

$$\begin{aligned} H(x,Y) &= \begin{cases} H(Y) & \text{if } x\in A \\ h(x) & \text{if } x\in B\end{cases} \text{ and} \\ G(x,Y) &= g(x) + G(Y)\end{aligned}$$

and so by the first step analysis we find,

$$\begin{aligned} w(x) &= \mathbb{E}_x[H(X)\,G(X)] = \mathbb{E}_{p(x,\cdot)}[H(x,Y)\,G(x,Y)] \\ &= \mathbb{E}_{p(x,\cdot)}[H(x,Y)\,(g(x)+G(Y))] \\ &= g(x)\,\mathbb{E}_{p(x,\cdot)}[H(x,Y)] + \mathbb{E}_{p(x,\cdot)}[H(x,Y)\,G(Y)].\end{aligned}$$

The first step analysis also shows (see the proof of Theorem 6.5)

$$u_h(x) := \mathbb{E}_x[h(X_{H_B})\,1_{H_B<\infty}] = \mathbb{E}_x[H(X)] = \mathbb{E}_{p(x,\cdot)}[H(x,Y)].$$

and therefore,

$$w(x) = g(x)\, u_h(x) + \mathbb{E}_{p(x,\cdot)}\left[H(x,Y)\, G(Y)\right]$$

Since $G(Y) = 0$ if $Y_0 \in B$ and $H(x,Y) = H(Y)$ if $Y_0 \in A$ we find,

$$\begin{aligned}
\mathbb{E}_{p(x,\cdot)}\left[H(x,Y)\, G(Y)\right] &= \sum_{x\in S} p(x,y)\, \mathbb{E}_y\left[H(x,Y)\, G(Y)\right] \\
&= \sum_{x\in A} p(x,y)\, \mathbb{E}_y\left[H(x,Y)\, G(Y)\right] \\
&= \sum_{x\in A} p(x,y)\, \mathbb{E}_y\left[H(Y)\, G(Y)\right] \\
&= \sum_{x\in A} p(x,y)\, w(y)
\end{aligned}$$

and hence

$$w(x) = g(x)\, u_h(x) + \sum_{x\in A} p(x,y)\, w(y) = g_h(x) + \sum_{x\in A} p(x,y)\, w(y).$$

But Theorem 6.6 with g replaced by g_h then shows w is given by Eq. (6.13). ■

Example 6.20 (A possible carnival game). Suppose that B is the disjoint union of L and W and suppose that you win $\sum_{n<H_B} g(X_n)$ if you end in W and win nothing when you end in L. What is the least we can expect to have to pay to play this game and where in $A := S \setminus B$ should we choose to start the game. To answer these questions we should compute our expected winnings ($w(x)$) for each starting point $x \in A$;

$$w(x) = \mathbb{E}_x\left[1_W(X_{H_B}) \sum_{n<H_B} g(X_n)\right].$$

Once we find w we should expect to pay at least $C := \max_{x\in A} w(x)$ and we should start at a location $x_0 \in A$ where $w(x_0) = \max_{x\in A} w(x) = C$. As an application of Theorem 6.18 we know that

$$w(x) = \left[(\mathbf{I} - \mathbf{Q})^{-1}\, \mathbf{g}_h\right]_x$$

where[5]

$$g_h(x) = g(x)\, \mathbb{E}_x\left[1_W(X_{H_B})\right] = g(x)\, \mathbb{P}_x\left(X_{H_B} \in W\right).$$

Let us now specialize these results to the chain in Example 6.8 where

[5] Intuitively, the effective pay off for a visit to site x is $g(x) \cdot P_x(\text{ we win}) + 0 \cdot P_x(\text{we loose})$.

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left[\begin{array}{cccc} 0 & 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

Let us make 4 the winning state and 3 the losing state (i.e. $h(3) = 0$ and $h(4) = 1$) and let $g = (g(1), g(2))$ be the payoff function. We have already seen that

$$\begin{bmatrix} u_h(1) \\ u_h(2) \end{bmatrix} = \begin{bmatrix} \mathbb{P}_1(X_{H_B} = 4) \\ \mathbb{P}_2(X_{H_B} = 4) \end{bmatrix} = \begin{bmatrix} \frac{7}{15} \\ \frac{2}{5} \end{bmatrix}$$

so that $\mathbf{g} * \mathbf{u}_h = \begin{bmatrix} \frac{7}{15} g_1 \\ \frac{2}{5} g_2 \end{bmatrix}$ and therefore

$$\begin{aligned}
\begin{bmatrix} w(1) \\ w(2) \end{bmatrix} &= (I - \mathbf{Q})^{-1} \begin{bmatrix} \frac{7}{15} g_1 \\ \frac{2}{5} g_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{7}{5} & \frac{8}{15} \\ \frac{6}{5} & \frac{8}{5} \end{bmatrix} \begin{bmatrix} \frac{7}{15} g_1 \\ \frac{2}{5} g_2 \end{bmatrix} = \begin{bmatrix} \frac{49}{75} g_1 + \frac{16}{75} g_2 \\ \frac{14}{25} g_1 + \frac{16}{25} g_2 \end{bmatrix}.
\end{aligned}$$

Let us examine a few different choices for g.

1. When $g(1) = 32$ and $g(2) = 7$, we have

$$\begin{bmatrix} w(1) \\ w(2) \end{bmatrix} = \begin{bmatrix} \frac{7}{5} & \frac{8}{15} \\ \frac{6}{5} & \frac{8}{5} \end{bmatrix} \begin{bmatrix} \frac{7}{15} 32 \\ \frac{2}{5} 7 \end{bmatrix} = \begin{bmatrix} \frac{112}{5} \\ \frac{112}{5} \end{bmatrix} = \begin{bmatrix} 22.4 \\ 22.4 \end{bmatrix}$$

   and so it does not matter where we start and we are going to have to pay at least \$22.40 to play.
2. When $g(1) = 10 = g(2)$, then

$$\begin{bmatrix} w(1) \\ w(2) \end{bmatrix} = \begin{bmatrix} \frac{7}{5} & \frac{8}{15} \\ \frac{6}{5} & \frac{8}{5} \end{bmatrix} \begin{bmatrix} \frac{7}{15} 10 \\ \frac{2}{5} 10 \end{bmatrix} = \begin{bmatrix} \frac{26}{3} \\ 12 \end{bmatrix} = \begin{bmatrix} 8.666\,7 \\ 12.0 \end{bmatrix}$$

   and we should enter the game at site 2. We are going to have to pay at least \$12 to play.
3. If $g(1) = 20$ and $g(2) = 7$,

$$\begin{bmatrix} w(1) \\ w(2) \end{bmatrix} = \begin{bmatrix} \frac{7}{5} & \frac{8}{15} \\ \frac{6}{5} & \frac{8}{5} \end{bmatrix} \begin{bmatrix} \frac{7}{15} 20 \\ \frac{2}{5} 7 \end{bmatrix} = \begin{bmatrix} \frac{364}{25} \\ \frac{392}{25} \end{bmatrix} = \begin{bmatrix} 14.56 \\ 15.68 \end{bmatrix}$$

   and again we should enter the game at site 2. We are going to have to pay at least \$15.68 to play.

6.3 Random Walk Exercises

Exercise 6.3 (Uniqueness of solutions to 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 $g : \mathbb{Z} \cap (\alpha, \beta) \to \mathbb{R}$ be a given function. Show that there is exactly one function $u : [\alpha, \beta] \cap \mathbb{Z} \to \mathbb{R}$ with prescribed values on two two consecutive points in $[\alpha, \beta] \cap \mathbb{Z}$ which satisfies the second order recurrence relation:

$$au(x+1) + bu(x) + cu(x-1) = f(x) \text{ for all } x \in \mathbb{Z} \cap (\alpha, \beta). \tag{6.14}$$

are for $\alpha < x < \beta$. Show; if u and w both satisfy Eq. (6.14) and $u = w$ on two consecutive points in $(\alpha, \beta) \cap \mathbb{Z}$, then $u(x) = w(x)$ for all $x \in [\alpha, \beta] \cap \mathbb{Z}$.

Exercise 6.4 (General homogeneous solutions). 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

$$au(x+1) + bu(x) + cu(x-1) = 0 \text{ for all } x \in \mathbb{Z} \cap (\alpha, \beta), \tag{6.15}$$

i.e. Eq. (6.14) with $f(x) \equiv 0$. Show:

1. for any $\lambda \in \mathbb{C}$,
$$a\lambda^{x+1} + b\lambda^x + c\lambda^{x-1} = \lambda^{x-1} p(\lambda) \tag{6.16}$$
where $p(\lambda) = a\lambda^2 + b\lambda + c$ is the **characteristic polynomial** associated to Eq. (6.14).
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. (6.14) it follows that for any choice of $A_\pm \in \mathbb{R}$, the function,
$$w(x) := A_+ \lambda_+^x + A_- \lambda_-^x, \tag{6.17}$$
solves Eq. (6.14) for all $x \in \mathbb{Z}$.
2. Show there is a unique choice of constants, $A_\pm \in \mathbb{R}$, such that the function $u(x)$ is given by
$$u(x) := A_+ \lambda_+^x + A_- \lambda_-^x \text{ for all } \alpha \le x \le \beta.$$
3. Now suppose that $b^2 = 4ac$ and $\lambda_0 := -b/(2a)$ is the double root of $p(\lambda)$. Show for any choice of A_0 and A_1 in $\mathbb{R}$ that
$$w(x) := (A_0 + A_1 x)\lambda_0^x \tag{6.18}$$
solves Eq. (6.14) for all $x \in \mathbb{Z}$. **Hint:** Differentiate Eq. (6.16) with respect to λ and then set $\lambda = \lambda_0$.
4. Again show that any function u solving Eq. (6.14) is of the form $u(x) = (A_0 + A_1 x)\lambda_0^x$ for $\alpha \le x \le \beta$ for some unique choice of constants $A_0, A_1 \in \mathbb{R}$.

In the next group 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

$$\mathbb{P}(X_{n+1} = x \pm 1 | X_n = x) = 1/2 \text{ for all } n \in \mathbb{N}_0 \text{ and } x \in \mathbb{Z}.$$

Further let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and

$$H_{a,b} := \min\{n : X_n \in \{a, b\}\} \text{ and } H_b := \inf\{n : X_n = b\}.$$

We know by Corollary 5.25 that $\mathbb{E}_0[H_{a,b}] < \infty$ from which it follows that $\mathbb{P}(H_{a,b} < \infty) = 1$ for all $a < 0 < b$. For these reasons we will ignore the event $\{H_{a,b} = \infty\}$ in what follows below.

Exercise 6.5. Let $w(x) := \mathbb{P}_x\left(X_{H_{a,b}} = b\right) := \mathbb{P}\left(X_{H_{a,b}} = b | X_0 = x\right)$.

1. Use first step analysis to show for $a < x < b$ that
$$w(x) = \frac{1}{2}(w(x+1) + w(x-1)) \tag{6.19}$$
provided we define $w(a) = 0$ and $w(b) = 1$.
2. Use the results of Exercises 6.3 and 6.4 to show
$$\mathbb{P}_x\left(X_{H_{a,b}} = b\right) = w(x) = \frac{1}{b-a}(x-a). \tag{6.20}$$
3. Let
$$H_b := \begin{cases} \min\{n : X_n = b\} & \text{if } \{X_n\} \text{ hits } b \\ \infty & \text{otherwise} \end{cases}$$
be the first time $\{X_n\}$ hits b. Explain why, $\left\{X_{H_{a,b}} = b\right\} \subset \{H_b < \infty\}$ and use this along with Eq. (6.20) to conclude that $\mathbb{P}_x(H_b < \infty) = 1$ for all $x < b$. (By symmetry this result holds true for all $x \in \mathbb{Z}$.)

Exercise 6.6. The goal of this exercise is to give a second proof of the fact that $\mathbb{P}_x(H_b < \infty) = 1$. Here is the outline:

1. Let $w(x) := \mathbb{P}_x(H_b < \infty)$. Again use first step analysis to show that $w(x)$ satisfies Eq. (6.19) for all x with $w(b) = 1$.
2. Use Exercises 6.3 and 6.4 to show that there is a constant, c, such that
$$w(x) = c \cdot (x - b) + 1 \text{ for all } x \in \mathbb{Z}.$$
3. Explain why c must be zero to again show that $\mathbb{P}_x(H_b < \infty) = 1$ for all $x \in \mathbb{Z}$.

Exercise 6.7. Let $H = H_{a,b}$ and $u(x) := \mathbb{E}_x H := \mathbb{E}[H | X_0 = x]$.

1. Use first step analysis to show for $a < x < b$ that
$$u(x) = \frac{1}{2}(u(x+1) + u(x-1)) + 1 \tag{6.21}$$
with the convention that $u(a) = 0 = u(b)$.
2. Show that
$$u(x) = A_0 + A_1 x - x^2 \tag{6.22}$$
solves Eq. (6.21) for any choice of constants A_0 and A_1.
3. Choose A_0 and A_1 so that $u(x)$ satisfies the boundary conditions, $u(a) = 0 = u(b)$. Use this to conclude that
$$\mathbb{E}_x H_{a,b} = -ab + (b+a)x - x^2 = -a(b-x) + bx - x^2. \tag{6.23}$$

Remark 6.21. Notice that $H_{a,b} \uparrow H_b = \inf\{n : X_n = b\}$ as $a \downarrow -\infty$, and so passing to the limit as $a \downarrow -\infty$ in Eq. (6.23) shows
$$\mathbb{E}_x H_b = \infty \text{ for all } x < b.$$

Combining the last couple of exercises together shows that $\{X_n\}$ is "**null - recurrent.**"

Exercise 6.8. Let $H = H_b$. The goal of this exercise is to give a second proof of the fact and $u(x) := \mathbb{E}_x H = \infty$ for all $x \neq b$. Here is the outline. Let $u(x) := \mathbb{E}_x H \in [0,\infty] = [0,\infty) \cup \{\infty\}$.

1. Note that $u(b) = 0$ and, by a first step analysis, that $u(x)$ satisfies Eq. (6.21) for all $x \neq b$ – allowing for the possibility that some of the $u(x)$ may be infinite.
2. Argue, using Eq. (6.21), that if $u(x) < \infty$ for some $x < b$ then $u(y) < \infty$ for all $y < b$. Similarly, if $u(x) < \infty$ for some $x > b$ then $u(y) < \infty$ for all $y > b$.
3. If $u(x) < \infty$ for all $x > b$ then $u(x)$ must be of the form in Eq. (6.22) for some A_0 and A_1 in $\mathbb{R}$ such that $u(b) = 0$. However, this would imply, $u(x) = \mathbb{E}_x H \to -\infty$ as $x \to \infty$ which is impossible since $\mathbb{E}_x H \geq 0$ for all x. Thus we must conclude that $\mathbb{E}_x H = u(x) = \infty$ for all $x > b$. (A similar argument works if we assume that $u(x) < \infty$ for all $x < b$.)

1. The first step analysis shows $u(x)$ satisfies Eq. (6.21) for all $x \neq b$ even if $u(x) = \infty$ for some x. Note that $u(b) = \mathbb{E}_b H_b = 0$.
2. If $u(x) < \infty$ for some $x < b$, then Eq. (6.21) can only hold if $u(x+1) < \infty$ and $u(x-1) < \infty$. In particular we may continue this line of reasoning to see that $u(x-2) < \infty$, and then $u(x-3) < \infty, \ldots$, i.e. $u(y) < \infty$ for all $y < b$. A similar argument shows that if $u(x) < \infty$ for some $x > b$ then $u(y) < \infty$ for all $y > b$.
3. If $u(x) < \infty$ for all $x > b$ then $u(x)$ must be of the form in Eq. (6.22) for some A_0 and A_1 in $\mathbb{R}$ such that $u(b) = 0$. However, this would imply,
$$\mathbb{E}_x H = u(x) = A_0 + A_1 x - x^2 \to -\infty \text{ as } x \to \infty$$
which is impossible since $\mathbb{E}_x H \geq 0$ for all x. Thus we must conclude that in fact $u(x) = \infty$ for all $x > b$. A similar argument shows that $u(x) = \infty$ for all $x < b$ as well.

For the remaining exercises in this section we will assume that $p \in (1/2, 1)$ and $q = 1 - p$ so that $p/q > 1$.

Exercise 6.9 (Biased random walks I). Let $p \in (1/2, 1)$ and consider the biased random walk $\{X_n\}_{n\geq 0}$ on the $S = \mathbb{Z}$ where $X_n = \xi_0 + \xi_1 + \cdots + \xi_n$, $\{\xi_i\}_{i=1}^\infty$ are i.i.d. with $\mathbb{P}(\xi_i = 1) = p \in (0,1)$ and $\mathbb{P}(\xi_i = -1) = q := 1 - p$, and $\xi_0 = x$ for some $x \in \mathbb{Z}$. Let $H = H_{\{0\}}$ be the first hitting time of $\{0\}$ and $u(x) := \mathbb{P}_x(H < \infty)$.

a) Use the first step analysis to show
$$u(x) = pu(x+1) + qu(x-1) \text{ for } x \neq 0 \text{ and } u(0) = 1. \tag{6.24}$$

b) Use Eq. (6.24) along with Exercises 6.3 and 6.4 to show for some $a_\pm \in \mathbb{R}$ that
$$u(x) = (1 - a_+) + a_+ (q/p)^x \text{ for } x \geq 0 \text{ and} \tag{6.25}$$
$$u(x) = (1 - a_-) + a_- (q/p)^x \text{ for } x \leq 0. \tag{6.26}$$

c) By considering the limit as $x \to -\infty$ conclude that $a_- = 0$ and $u(x) = 1$ for all $x < 0$, i.e. $\mathbb{P}_x(H_0 < \infty) = 1$ for all $x \leq 0$.

Exercise 6.10 (Biased random walks II). The goal of this exercise is to evaluate $\mathbb{P}_x(H_0 < \infty)$ for $x \geq 0$. To do this let $B_n := \{0, n\}$ and $H_n := H_{\{0,n\}}$. Let $h(x) := \mathbb{P}_x(X_{H_n} = 0)$ where $\{X_{H_n} = 0\}$ is the event of hitting 0 before n.

a) Use the first step analysis to show
$$h(x) = ph(x+1) + qh(x-1) \text{ with } h(0) = 1 \text{ and } h(n) = 0.$$

b) Show the unique solution to this equation is given by
$$\mathbb{P}_x(X_{H_n} = 0) = h(x) = \frac{(q/p)^x - (q/p)^n}{1 - (q/p)^n}.$$

c) Argue that
$$\mathbb{P}_x(H < \infty) = \lim_{n\to\infty} \mathbb{P}_x(\{X_{H_n} = 0\}) = (q/p)^x < 1 \text{ for all } x \geq 0.$$

The following formula summarizes Exercises 6.9 and 6.10; for $\frac{1}{2} < p < 1$,

$$\mathbb{P}_x(H < \infty) = \begin{cases} (q/p)^x & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases}. \tag{6.27}$$

Example 6.22 (Biased random walks III). Continue the notation in Exercise 6.9. Let us start to compute $\mathbb{E}_x H$. Since $\mathbb{P}_x(H = \infty) > 0$ for $x > 0$ we already know that $\mathbb{E}_x H = \infty$ for all $x > 0$. Nevertheless we will deduce this fact again here. Letting $u(x) = \mathbb{E}_x H$ it follows by the first step analysis that, for $x \neq 0$,

$$\begin{aligned} u(x) &= p[1 + u(x+1)] + q[1 + u(x-1)] \\ &= pu(x+1) + qu(x-1) + 1 \end{aligned} \tag{6.28}$$

with $u(0) = 0$. Notice $u(x) = \infty$ is a solution to this equation while if $u(n) < \infty$ for some $n \neq 0$ then Eq. (6.28) implies that $u(x) < \infty$ for all $x \neq 0$ with the same sign as n. 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. (6.28) is therefore,

$$u(x) = (q-p)^{-1} x + a + b(q/p)^x. \tag{6.29}$$

Using the boundary condition, $u(0) = 0$ allows us to conclude that $a + b = 0$ and therefore,

$$u(x) = (q-p)^{-1} x + a[1 - (q/p)^x]. \tag{6.30}$$

Notice that $u(x) \to -\infty$ as $x \to +\infty$ no matter how a is chosen and therefore we must conclude that the desired solution to Eq. (6.28) is $u(x) = \infty$ for $x > 0$ as we already mentioned. In the next exercise you will compute $\mathbb{E}_x H$ for $x < 0$.

Exercise 6.11 (Biased random walks IV). Continue the notation in Example 6.22. Using the outline below, show

$$\mathbb{E}_x H = \frac{|x|}{p-q} \text{ for } x \leq 0. \tag{6.31}$$

In the following outline n is a negative integer, H_n is the first hitting time of n so that $H_{\{n,0\}} = H_n \wedge H = \min\{H, H_n\}$ is the first hitting time of $\{n, 0\}$. By Corollary 5.25 we know that $u(x) := \mathbb{E}_x\left[H_{\{n,0\}}\right] < \infty$ for all $n \leq x \leq 0$ and by a first step analysis one sees that $u(x)$ still satisfies Eq. (6.28) for $n < x < 0$ and has boundary conditions $u(n) = 0 = u(0)$.

a) From Eq. (6.30) we know that, for some $a \in \mathbb{R}$,

$$\mathbb{E}_x\left[H_{\{n,0\}}\right] = u(x) = (q-p)^{-1} x + a[1 - (q/p)^x].$$

Use $u(n) = 0$ in order to show

$$a = a_n = \frac{n}{(1 - (q/p)^n)(p-q)}$$

and therefore,

$$\mathbb{E}_x\left[H_{\{n,0\}}\right] = \frac{1}{p-q}\left[|x| + n\frac{1 - (q/p)^x}{1 - (q/p)^n}\right] \text{ for } n \leq x \leq 0.$$

b) Argue that $\mathbb{E}_x H = \lim_{n \to -\infty} \mathbb{E}_x[H_n \wedge H]$ and use this and part a) to prove Eq. (6.31).

Remark 6.23 (More on the boundary conditions). If we were to use Theorem 6.6 directly to derive Eq. (6.28) in the case that $u(x) := \mathbb{E}_x\left[H_{\{n,0\}}\right] < \infty$ we for all $0 \leq x \leq n$. we would find, for $x \neq 0$, that

$$u(x) = \sum_{y \notin \{n,0\}} q(x,y)\, u(y) + 1$$

which implies that $u(x)$ satisfies Eq. (6.28) 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. (6.28) 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 $w(0) = w(n) = 0$ as we have used above.

6.4 Wald's Equation and Gambler's Ruin

Example 6.24. Here are some example of random times some of which are stopping times and some of which are not. In these examples we will always use the convention that the minimum of the empty set is $+\infty$.

1. The random time, $\tau = \min\{k : |X_k| \geq 5\}$ (the first time, k, such that $|X_k| \geq 5$) is a stopping time since

$$\{\tau = k\} = \{|X_1| < 5, \ldots, |X_{k-1}| < 5, |X_k| \geq 5\}.$$

2. Let $W_k := X_1 + \cdots + X_k$, then the random time,
$$\tau = \min\{k : W_k \geq \pi\}$$
is a stopping time since,
$$\{\tau = k\} = \left\{ \begin{array}{c} W_j = X_1 + \cdots + X_j < \pi \text{ for } j = 1, 2, \ldots, k-1, \\ \& \ X_1 + \cdots + X_{k-1} + X_k \geq \pi \end{array} \right\}.$$
3. For $t \geq 0$, let $N(t) = \#\{k : W_k \leq t\}$. Then
$$\{N(t) = k\} = \{X_1 + \cdots + X_k \leq t, \ X_1 + \cdots + X_{k+1} > t\}$$
which shows that $N(t)$ is **not** a stopping time. On the other hand, since
$$\begin{aligned} \{N(t) + 1 = k\} &= \{N(t) = k - 1\} \\ &= \{X_1 + \cdots + X_{k-1} \leq t, \ X_1 + \cdots + X_k > t\}, \end{aligned}$$
we see that $N(t) + 1$ **is** a stopping time!
4. If τ is a stopping time then so is $\tau + 1$ because,
$$1_{\{\tau+1=k\}} = 1_{\{\tau=k-1\}} = \sigma_{k-1}(X_0, \ldots, X_{k-1})$$
which is also a function of $(X_0, \ldots, X_k)$ which happens not to depend on X_k.
5. On the other hand, if τ is a stopping time it is **not** necessarily true that $\tau - 1$ is still a stopping time as seen in item 3. above.
6. One can also see that the last time, k, such that $|X_k| \geq \pi$ is typically **not** a stopping time. (Think about this.)

The following presentation of Wald's equation is taken from Ross [17, p. 59-60].

Theorem 6.25 (Wald's Equation). *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables, $f(x)$ is a non-negative function of $x \in \mathbb{R}$, and τ is a stopping time. Then*
$$\mathbb{E}\left[\sum_{n=1}^{\tau} f(X_n)\right] = \mathbb{E}f(X_1) \cdot \mathbb{E}\tau. \tag{6.32}$$
This identity also holds if $f(X_n)$ are real valued but integrable and τ is a stopping time such that $\mathbb{E}\tau < \infty$. (See Resnick for more identities along these lines.)

Proof. If $f(X_n) \geq 0$ for all n, then the the following computations need no justification,
$$\begin{aligned} \mathbb{E}\left[\sum_{n=1}^{\tau} f(X_n)\right] &= \mathbb{E}\left[\sum_{n=1}^{\infty} f(X_n) 1_{n\leq\tau}\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[f(X_n) 1_{n\leq\tau}\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}\left[f(X_n) u_n(X_1, \ldots, X_{n-1})\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}\left[f(X_n)\right] \cdot \mathbb{E}\left[u_n(X_1, \ldots, X_{n-1})\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}\left[f(X_n)\right] \cdot \mathbb{E}\left[1_{n\leq\tau}\right] = \mathbb{E}f(X_1) \sum_{n=1}^{\infty} \mathbb{E}\left[1_{n\leq\tau}\right] \\ &= \mathbb{E}f(X_1) \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} 1_{n\leq\tau}\right] = \mathbb{E}f(X_1) \cdot \mathbb{E}\tau. \end{aligned}$$
If $\mathbb{E}|f(X_n)| < \infty$ and $\mathbb{E}\tau < \infty$, the above computation with f replaced by $|f|$ shows all sums appearing above are equal $\mathbb{E}|f(X_1)| \cdot \mathbb{E}\tau < \infty$. Hence we may remove the absolute values to again arrive at Eq. (6.32). ■

Example 6.26. Let $\{X_n\}_{n=1}^\infty$ be i.i.d. such that $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$ and let
$$\tau := \min\{n : X_1 + \cdots + X_n = 10\}.$$
For example τ is the first time we have flipped 10 heads of a fair coin. By Wald's equation (valid because $X_n \geq 0$ for all n) we find
$$10 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = \frac{1}{2}\mathbb{E}\tau$$
and therefore $\mathbb{E}\tau = 20 < \infty$.

Example 6.27 (Gambler's ruin). Let $\{X_n\}_{n=1}^\infty$ be i.i.d. such that $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = 1/2$ and let
$$\tau := \min\{n : X_1 + \cdots + X_n = 1\}.$$
So τ may represent the first time that a gambler is ahead by 1. Notice that $\mathbb{E}X_1 = 0$. If $\mathbb{E}\tau < \infty$, then we would have $\tau < \infty$ a.s. and by Wald's equation would give,
$$1 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = 0 \cdot \mathbb{E}\tau$$
which can not hold. Hence it must be that
$$\mathbb{E}\tau = \mathbb{E}[\text{first time that a gambler is ahead by } 1] = \infty.$$

6.5 Some more worked examples

Example 6.28. Let $S = \{1, 2\}$ and $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with jump diagram in Figure 6.5. In this case $\mathbf{P}^{2n} = I$ while $\mathbf{P}^{2n+1} = \mathbf{P}$ and therefore $\lim_{n\to\infty} \mathbf{P}^n$ does not exist.

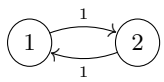


Fig. 6.5. A non-random chain.

On the other hand it is easy to see that the invariant distribution, π, for $\mathbf{P}$ is $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ and, moreover,

$$\frac{\mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^N}{N} \to \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix}.$$

Let us compute

$$\begin{bmatrix} \mathbb{E}_1 R_1 \\ \mathbb{E}_2 R_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbb{E}_1 R_2 \\ \mathbb{E}_2 R_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

so that indeed, $\pi_1 = 1/\mathbb{E}_1 R_1$ and $\pi_2 = 1/\mathbb{E}_2 R_2$. Of course $R_1 = 2$ (P_1 -a.s.) and $R_2 = 2$ (P_2 -a.s.) so that it is obvious that $\mathbb{E}_1 R_1 = \mathbb{E}_2 R_2 = 2$.

Example 6.29. Again let $S = \{1, 2\}$ and $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with jump diagram in Figure 6.6. In this case the chain is not irreducible and every $\pi = [a\ b]$ with

Fig. 6.6. A simple non-irreducible chain.

$a + b = 1$ and $a, b \geq 0$ is an invariant distribution.

Example 6.30. Suppose that $S = \{1, 2, 3\}$, and

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \hline \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

has the jump graph given by 6.7. Notice that $\mathbf{P}^2_{11} > 0$ and $\mathbf{P}^3_{11} > 0$ that $\mathbf{P}$ is

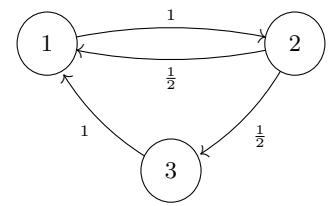


Fig. 6.7. A simple 3 state jump diagram.

"aperiodic." We now find the invariant distribution,

$$\operatorname{Nul}(\mathbf{P} - I)^{\mathrm{tr}} = \operatorname{Nul} \begin{bmatrix} -1 & \frac{1}{2} & 1 \\ 1 & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{bmatrix} = \mathbb{R} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

Therefore the invariant distribution is given by

$$\pi = \frac{1}{5} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix}.$$

Let us now observe that

$$\mathbf{P}^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{P}^3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{P}^{20} = \begin{bmatrix} \frac{409}{1024} & \frac{205}{512} & \frac{205}{1024} \\ \frac{205}{512} & \frac{409}{1024} & \frac{205}{1024} \\ \frac{205}{512} & \frac{205}{512} & \frac{51}{256} \end{bmatrix} = \begin{bmatrix} 0.399\,41 & 0.400\,39 & 0.200\,20 \\ 0.400\,39 & 0.399\,41 & 0.200\,20 \\ 0.400\,39 & 0.400\,39 & 0.199\,22 \end{bmatrix}.$$

Let us also compute $\mathbb{E}_2 R_3$ via,

$$\begin{bmatrix} \mathbb{E}_1 R_3 \\ \mathbb{E}_2 R_3 \\ \mathbb{E}_3 R_3 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$$

so that

$$\frac{1}{\mathbb{E}_3 R_3} = \frac{1}{5} = \pi_3.$$

Example 6.31. The transition matrix,

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

is represented by the jump diagram in Figure 6.8. This chain is aperiodic. We

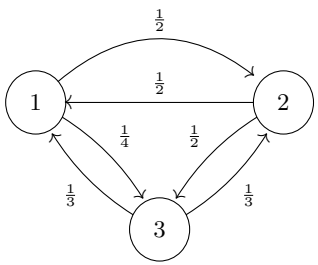


Fig. 6.8. In the above diagram there are jumps from 1 to 1 with probability 1/4 and jumps from 3 to 3 with probability 1/3 which are not explicitly shown but must be inferred by conservation of probability.

find the invariant distribution as,

$$\begin{aligned} \operatorname{Nul}(\mathbf{P} - I)^{\mathrm{tr}} &= \operatorname{Nul}\left(\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)^{\mathrm{tr}} \\ &= \operatorname{Nul}\left(\begin{bmatrix} -\frac{3}{4} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & -1 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & -\frac{2}{3} \end{bmatrix} \right) = \mathbb{R} \begin{bmatrix} 1 \\ \frac{5}{6} \\ 1 \end{bmatrix} = \mathbb{R} \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix} \end{aligned}$$

$$\pi = \frac{1}{17} \begin{bmatrix} 6 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 0.352\,94 & 0.294\,12 & 0.352\,94 \end{bmatrix}.$$

In this case

$$\mathbf{P}^{10} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}^{10} = \begin{bmatrix} 0.352\,98 & 0.294\,04 & 0.352\,98 \\ 0.352\,89 & 0.294\,23 & 0.352\,89 \\ 0.352\,95 & 0.294\,1 & 0.352\,95 \end{bmatrix}.$$

Let us also compute

$$\begin{bmatrix} \mathbb{E}_1 R_2 \\ \mathbb{E}_2 R_2 \\ \mathbb{E}_3 R_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 0 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{11}{5} \\ \frac{17}{5} \\ \frac{13}{5} \end{bmatrix}$$

so that

$$1/\mathbb{E}_2 R_2 = 5/17 = \pi_2.$$

Example 6.32. Consider the following Markov matrix,

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 0 & 3/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

with jump diagram in Figure 6.9. Since this matrix is **doubly stochastic** (i.e.

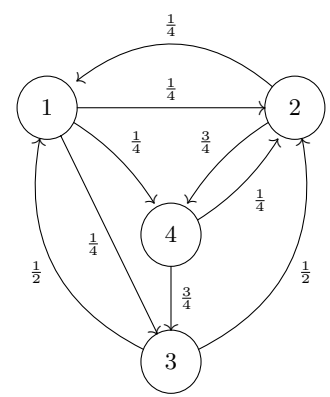


Fig. 6.9. The jump diagram for Q.

$\sum_{i=1}^4 \mathbf{P}_{ij} = 1$ for all j as well as $\sum_{j=1}^4 \mathbf{P}_{ij} = 1$ for all i), it is easy to check that $\pi = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. Let us compute $\mathbb{E}_3 R_3$ as follows

$$\begin{bmatrix}\mathbb{E}_1R_3\\ \mathbb{E}_2R_3\\ \mathbb{E}_3R_3\\ \mathbb{E}_4R_3\end{bmatrix} = \left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} - \begin{bmatrix}1/4&1/4&0&1/4\\1/4&0&0&3/4\\1/2&1/2&0&0\\0&1/4&0&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$
$$= \begin{bmatrix}\frac{50}{17}\\ \frac{52}{17}\\ 4\\ \frac{30}{17}\end{bmatrix}$$

so that $\mathbb{E}_3R_3 = 4 = 1/\pi_4$ as it should be. Similarly,

$$\begin{bmatrix}\mathbb{E}_1R_2\\ \mathbb{E}_2R_2\\ \mathbb{E}_3R_2\\ \mathbb{E}_4R_2\end{bmatrix} = \left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} - \begin{bmatrix}1/4&0&1/4&1/4\\1/4&0&0&3/4\\1/2&0&0&0\\0&0&3/4&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$
$$= \begin{bmatrix}\frac{54}{17}\\ 4\\ \frac{44}{17}\\ \frac{50}{17}\end{bmatrix}$$

and again $\mathbb{E}_2R_2 = 4 = 1/\pi_2$.

Example 6.33. Consider the following example,

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc}1&2&3\end{array}\\ \begin{bmatrix}1/2&1/2&0\\0&1/2&1/2\\1/2&1/2&0\end{bmatrix}\end{array}\begin{array}{c}1\\2\\3\end{array}$$

with jump diagram given in Figure 6.10.We have

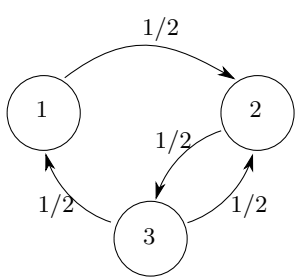


Fig. 6.10. The jump diagram associated to P.

$$\mathbf{P}^2 = \begin{bmatrix}1/2&1/2&0\\0&1/2&1/2\\1/2&1/2&0\end{bmatrix}^2 = \begin{bmatrix}\frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\end{bmatrix}$$

and

$$\mathbf{P}^3 = \begin{bmatrix}1/2&1/2&0\\0&1/2&1/2\\1/2&1/2&0\end{bmatrix}^3 = \begin{bmatrix}\frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\end{bmatrix}.$$

To have a picture what is going on here, imaging that $\pi = (\pi_1, \pi_2, \pi_3)$ represents the amount of sand at the sites, 1, 2, and 3 respectively. During each time step we move the sand on the sites around according to the following rule. The sand at site j after one step is $\sum_i \pi_i \mathbf{P}_{ij}$, namely site i contributes $\mathbf{P}_{ij}$ fraction its sand, π_i, to site j. Everyone does this to arrive at a new distribution. Hence π is an invariant distribution if each π_i remains unchanged, i.e. $\pi = \pi\mathbf{P}$. (Keep in mind the sand is still moving around it is just that the size of the piles remains unchanged.)

As a specific example, suppose $\pi = (1, 0, 0)$ so that all of the sand starts at 1. After the first step, the pile at 1 is split into two and 1/2 is sent to 2 to get $\pi_1 = (1/2, 1/2, 0)$ which is the first row of $\mathbf{P}$. At the next step the site 1 keeps 1/2 of its sand ($= 1/4$) and still receives nothing, while site 2 again receives the other 1/2 and keeps half of what it had ($= 1/4 + 1/4$) and site 3 then gets ($1/2 \cdot 1/2 = 1/4$) so that $\pi_2 = \begin{bmatrix}\frac14&\frac12&\frac14\end{bmatrix}$ which is the first row of $\mathbf{P}^2$. It turns out in this case that this is the invariant distribution. Formally,

$$\begin{bmatrix}\frac14&\frac12&\frac14\end{bmatrix}\begin{bmatrix}1/2&1/2&0\\0&1/2&1/2\\1/2&1/2&0\end{bmatrix} = \begin{bmatrix}\frac14&\frac12&\frac14\end{bmatrix}.$$

In general we expect to reach the invariant distribution only in the limit as $n \to \infty$.

Notice that if π is any stationary distribution, then $\pi\mathbf{P}^n = \pi$ for all n and in particular,

$$\pi = \pi\mathbf{P}^2 = \begin{bmatrix}\pi_1&\pi_2&\pi_3\end{bmatrix}\begin{bmatrix}\frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\\ \frac14&\frac12&\frac14\end{bmatrix} = \begin{bmatrix}\frac14&\frac12&\frac14\end{bmatrix}.$$

Hence $\begin{bmatrix}\frac14&\frac12&\frac14\end{bmatrix}$ is the unique stationary distribution for $\mathbf{P}$ in this case.

Example 6.34 (§3.2. p108 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 $X_n = (\#$ 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. Suppose there are N total number of particles in the flask, then the transition probabilities are given by,

$$\mathbf{P}_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i-1, i+1\} \\ \frac{i}{N} & \text{if} \quad j = i-1 \\ \frac{N-i}{N} & \text{if} \quad j = i+1. \end{cases}$$

For example, if $N = 2$ we have

$$(\mathbf{P}_{ij}) = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \hline \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}$$

and if $N = 3$, then we have in matrix form,

$$(\mathbf{P}_{ij}) = \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \hline \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}.$$

In the case $N = 2$,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

and when $N = 3$,

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{7}{9} & 0 & \frac{2}{9} \\ \frac{2}{9} & 0 & \frac{7}{9} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & \frac{7}{9} & 0 & \frac{2}{9} \\ \frac{7}{27} & 0 & \frac{20}{27} & 0 \\ 0 & \frac{20}{27} & 0 & \frac{7}{27} \\ \frac{2}{9} & 0 & \frac{7}{9} & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{25} \cong \begin{bmatrix} 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{26} \cong \begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix}$$

:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{100} \cong \begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix}$$

We also have

$$(\mathbf{P} - I)^{\mathrm{tr}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ \frac{1}{3} & -1 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & -1 & \frac{1}{3} \\ 0 & 0 & 1 & -1 \end{bmatrix}^{\mathrm{tr}} = \begin{bmatrix} -1 & \frac{1}{3} & 0 & 0 \\ 1 & -1 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & -1 & 1 \\ 0 & 0 & \frac{1}{3} & -1 \end{bmatrix}$$

and

$$\mathrm{Nul}\left((\mathbf{P} - I)^{\mathrm{tr}}\right) = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix}.$$

Hence if we take, $\pi = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix}$ then

$$\pi\mathbf{P} = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix} = \pi$$

is the stationary distribution. Notice that

$$\frac{1}{2}\left(\mathbf{P}^{25}+\mathbf{P}^{26}\right) \cong \frac{1}{2}\begin{bmatrix} 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix}$$
$$= \begin{bmatrix} 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \pi \\ \pi \end{bmatrix}.$$

Example 6.35. Let us consider the Markov matrix,

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}.$$

In this case we have

$$\mathbf{P}^{25} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^{25} \cong \begin{bmatrix} 0.399\,9 & 0.400\,15 & 0.199\,95 \\ 0.400\,02 & 0.399\,9 & 0.200\,07 \\ 0.400\,15 & 0.399\,9 & 0.199\,95 \end{bmatrix}$$
$$\mathbf{P}^{26} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^{26} \cong \begin{bmatrix} 0.400\,02 & 0.399\,9 & 0.200\,07 \\ 0.400\,02 & 0.400\,02 & 0.199\,95 \\ 0.399\,9 & 0.400\,15 & 0.199\,95 \end{bmatrix}$$
$$\mathbf{P}^{100} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^{100} \cong \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$$

and observe that

$$\begin{bmatrix} 0.4 & 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.4 & 0.2 \end{bmatrix}.$$

so that $\pi = \begin{bmatrix} 0.4 & 0.4 & 0.2 \end{bmatrix}$ is a stationary distribution for $\mathbf{P}$.

6.5.1 Life Time Processes

A computer component has life time T, with $\mathbb{P}(T=k)=a_k$ for $k \in \mathbb{N}$. Let X_n denote the **age of the component** in service at time n. The set up is then

$$[0,T_1] \cup (T_1, T_1+T_2] \cup (T_1+T_2, T_1+T_2+T_3] \cup \dots$$

so for example if $(T_1, T_2, T_3, \dots) = (1,3,4,\dots)$, then

$$X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2, X_4 = 0, X_5 = 1, X_5 = 2, X_5 = 3, X_6 = 0, \dots.$$

The transition probabilities are then

$$\mathbb{P}(X_{n+1} = 0 | X_n = k) = \mathbb{P}(T = k+1 | T > k) = \frac{a_{k+1}}{\sum_{m>k} a_k}$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = k+1 | X_n = k) &= \mathbb{P}(T > k+1 | T > k) \\ &= \frac{\mathbb{P}(T > k+1)}{\mathbb{P}(T > k)} = \frac{\sum_{m>k+1} a_k}{\sum_{m>k} a_k} \\ &= \frac{\sum_{m>k} a_k - a_{k+1}}{\sum_{m>k} a_k} = 1 - \frac{a_{k+1}}{\sum_{m>k} a_k}. \end{aligned}$$

See Exercise IV.2.E6 of Karlin and Taylor for a concrete example involving a chain of this form.

There is another way to look at this same situation, namely let Y_n denote the **remaining life** of the part in service at time n. So if $(T_1, T_2, T_3, \dots) = (1,3,4,\dots)$, then

$$Y_0 = 1, Y_1 = 3, Y_2 = 2, Y_3 = 1, Y_4 = 4, Y_5 = 3, Y_5 = 2, Y_5 = 1, \dots.$$

and the corresponding transition matrix is determined by

$$\mathbb{P}(Y_{n+1} = k-1 | Y_n = k) = 1 \text{ if } k \geq 2$$

while

$$\mathbb{P}(Y_{n+1} = k | Y_n = 1) = \mathbb{P}(T = k).$$

Example 6.36 (Exercise IV.2.E6 revisited). Let Y_n denote the **remaining life** of the part in service at time n. So if $(T_1, T_2, T_3, \dots) = (1,3,4,\dots)$, then

$$Y_0 = 1, Y_1 = 3, Y_2 = 2, Y_3 = 1, Y_4 = 4, Y_5 = 3, Y_5 = 2, Y_5 = 1, \dots.$$

If

k	0	1	2	3	4
$\mathbb{P}(T=k)$	0	0.1	0.2	0.3	0.4

the transition matrix is now given by

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \frac{1}{10} & \frac{1}{5} & \frac{3}{10} & \frac{2}{5} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}$$

whose invariant distribution is given by

$$\pi = \frac{1}{30}\left[10\ 9\ 7\ 4\right] = \left[\tfrac{1}{3}\ \tfrac{3}{10}\ \tfrac{7}{30}\ \tfrac{2}{15}\right].$$

The failure of a part is indicated by Y_n being 1 and so again the failure frequency is $\frac{1}{3}$ of the time as found before. Observe that expected life time of a part is;

$$\mathbb{E}[T] = 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.3 + 4 \cdot 0.4 = 3.$$

Thus we see that $\pi_1 = 1/\mathbb{E}T$ which is what we should have expected. To go a little further notice that from the jump diagram in Figure 6.11,

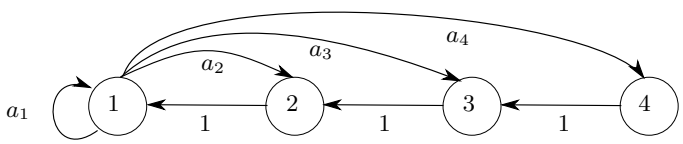


Fig. 6.11. The jump diagram for this "renewal" chain.

one see that

k	1	2	3	4
$\mathbb{P}_1(R_1 = k)$	$a_1 = 0.1$	$a_2 = 0.2$	$a_3 = 0.3$	$a_4 = 0.4$

and therefore,

$$\mathbb{E}_1 R_1 = 1a_1 + 2a_2 + 3a_3 + 4a_4 = \mathbb{E}T$$

and hence $\pi_1 = 1/\mathbb{E}_1 R_1 = 1/\mathbb{E}T$ in general for this type of chain.

6.5.2 Sampling Plans

This section summarizes the results of Section IV.2.3 of Karlin and Taylor. There one is considering at production line where each item manufactured has probability $0 \le p \le 1$ of being defective. Let i and r be two integers and sample the output of the line as follows. We begin by sampling every item until we have found i – in a row which are good. Then we sample each of then next items with probability $\frac{1}{r}$ determined randomly at end of production of each item. (If $r = 6$ we might throw a die each time a product comes off the line and sample that product when we roll a 6 say.) If we find a bad part we start the process over again. We now describe this as a Markov chain with states $\{E_k\}_{k=0}^{i}$ where E_k denotes that we have seen k – good parts in a row for $0 \le k < i$ and E_i indicates we are in stage II where we are randomly choosing to sample an item with probability $\frac{1}{r}$. The transition probabilities for this chain are given by

$$\begin{aligned}
\mathbb{P}(X_{n+1} = E_{k+1} | X_n = E_k) &= q := 1 - p \text{ for } k = 0, 1, 2 \ldots, i-1, \\
\mathbb{P}(X_{n+1} = E_0 | X_n = E_k) &= p \text{ if } 0 \le k \le i-1, \\
\mathbb{P}(X_{n+1} = E_0 | X_n = E_i) &= \frac{p}{r} \text{ and } \mathbb{P}(X_{n+1} = E_i | X_n = E_i) = 1 - \frac{p}{r},
\end{aligned}$$

with all other transitions being zero. The stationary distribution for this chain satisfies the equations,

$$\pi_k = \sum_{l=0}^{k} \mathbb{P}(X_{n+1} = E_k | X_n = E_l)\, \pi_l$$

so that

$$\begin{aligned}
\pi_0 &= \sum_{k=0}^{i-1} p\pi_i + \frac{p}{r}\pi_0, \\
\pi_1 &= q\pi_0,\ \pi_2 = q\pi_1, \ldots, \pi_{i-1} = q\pi_{i-2}, \\
\pi_i &= q\pi_{i-1} + \left(1 - \frac{p}{r}\right)\pi_i.
\end{aligned}$$

These equations may be solved (see Section IV.2.3 of Karlin and Taylor) to find in particular that

$$\begin{aligned}
\pi_k &= \frac{p(1-p)^k}{1 + (r-1)(1-p)^i} \text{ for } 1 \le k < i \text{ and} \\
\pi_i &= \frac{r(1-p)^k}{1 + (r-1)(1-p)^i}.
\end{aligned}$$

See Karlin and Taylor for more comments on this solution.

6.5.3 Extra Homework Problems

Exercises 6.12 – 6.15 refer to the following Markov matrix:

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 3/4 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \tag{6.33}$$

We will let $\{X_n\}_{n=0}^{\infty}$ denote the Markov chain associated to $\mathbf{P}$.

Exercise 6.12. Make a jump diagram for this matrix and identify the recurrent and transient classes. Also find the invariant destitutions for the chain restricted to each of the recurrent classes.

Exercise 6.13. Find all of the invariant distributions for $\mathbf{P}$.

Exercise 6.14. Compute the hitting probabilities, $h_5 = \mathbb{P}_5(X_n \text{ hits } \{3,4\})$ and $h_6 = \mathbb{P}_6(X_n \text{ hits } \{3,4\})$.

Exercise 6.15. Find $\lim_{n\to\infty} \mathbb{P}_6(X_n = j)$ for $j = 1,2,3,4,5,6$.

6.6 *Computations avoiding the first step analysis

You may safely skip the rest of this section!!

Theorem 6.37. *Let n denote a non-negative integer. If $h : B \to \mathbb{R}$ is measurable and either bounded or non-negative, then*

$$\mathbb{E}_x[h(X_n) : H_B = n] = \left(Q_A^{n-1} Q[1_B h]\right)(x)$$

and

$$\mathbb{E}_x[h(X_{H_B}) : H_B < \infty] = \left(\sum_{n=0}^{\infty} Q_A^n Q[1_B h]\right)(x). \tag{6.34}$$

If $g : A \to \mathbb{R}_+$ is a measurable function, then for all $x \in A$ and $n \in \mathbb{N}_0$,

$$\mathbb{E}_x[g(X_n) 1_{n<H_B}] = (Q_A^n g)(x).$$

In particular we have

$$\mathbb{E}_x\left[\sum_{n<H_B} g(X_n)\right] = \sum_{n=0}^{\infty} (Q_A^n g)(x) =: u(x), \tag{6.35}$$

where by convention, $\sum_{n<H_B} g(X_n) = 0$ when $H_B = 0$.

Proof. Let $x \in A$. In computing each of these quantities we will use;

$$\{H_B > n\} = \{X_i \in A \text{ for } 0 \le i \le n\} \text{ and}$$
$$\{H_B = n\} = \{X_i \in A \text{ for } 0 \le i \le n-1\} \cap \{X_n \in B\}.$$

From the second identity above it follows that for

$$\begin{aligned}
\mathbb{E}_x[h(X_n) : H_B = n] &= \mathbb{E}_x\left[h(X_n) : (X_1,\dots,X_{n-1}) \in A^{n-1},\ X_n \in B\right] \\
&= \sum_{n=1}^{\infty} \int_{A^{n-1}\times B} \prod_{j=1}^{n} Q(x_{j-1}, dx_j)\, h(x_n) \\
&= \left(Q_A^{n-1} Q[1_B h]\right)(x)
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathbb{E}_x[h(X_{H_B}) : H_B < \infty] &= \sum_{n=1}^{\infty} \mathbb{E}_x[h(X_n) : H_B = n] \\
&= \sum_{n=1}^{\infty} Q_A^{n-1} Q[1_B h] = \sum_{n=0}^{\infty} Q_A^n Q[1_B h].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}_x[g(X_n) 1_{n<H_B}] &= \int_{A^n} Q(x, dx_1)\, Q(x_1, dx_2) \dots Q(x_{n-1}, dx_n)\, g(x_n) \\
&= (Q_A^n g)(x)
\end{aligned}$$

and therefore,

$$\begin{aligned}
\mathbb{E}_x\left[\sum_{n=0}^{\infty} g(X_n) 1_{n<H_B}\right] &= \sum_{n=0}^{\infty} \mathbb{E}_x[g(X_n) 1_{n<H_B}] \\
&= \sum_{n=0}^{\infty} (Q_A^n g)(x).
\end{aligned}$$

■

In practice it is not so easy to sum the series in Eqs. (6.34) and (6.35). Thus we would like to have another way to compute these quantities. Since $\sum_{n=0}^{\infty} Q_A^n$ is a geometric series, we expect that

$$\sum_{n=0}^{\infty} Q_A^n = (I - Q_A)^{-1}$$

which is basically correct at least when $(I - Q_A)$ is invertible. This suggests that if $u(x) = \mathbb{E}_x[h(X_{H_B}) : H_B < \infty]$, then (see Eq. (6.34))

$$u = Q_A u + Q[1_B h] \text{ on } A, \tag{6.36}$$

and if $u(x) = \mathbb{E}_x\left[\sum_{n<H_B} g(X_n)\right]$, then (see Eq. (6.35))

$$u = Q_A u + g \text{ on } A. \tag{6.37}$$

That these equations are valid was the content of Corollary 6.43 below and Theorem 6.6 above. below which we will prove using the "first step" analysis in the next theorem. We will give another direct proof in Theorem 6.42 below as well.

Lemma 6.38. *Keeping the notation above we have*

$$\mathbb{E}_x T = \sum_{n=0}^{\infty} \sum_{y \in A} Q^n(x,y) \text{ for all } x \in A, \tag{6.38}$$

where $\mathbb{E}_x T = \infty$ is possible.

Proof. By definition of T we have for $x \in A$ and $n \in \mathbb{N}_0$ that,

$$\begin{aligned}\mathbb{P}_x(T > n) &= \mathbb{P}_x(X_1, \dots, X_n \in A) \\ &= \sum_{x_1,\dots,x_n \in A} p(x,x_1)\, p(x_1,x_2) \dots p(x_{n-1},x_n) \\ &= \sum_{y \in A} Q^n(x,y). \end{aligned} \tag{6.39}$$

Therefore Eq. (6.38) now follows from Lemma 5.18 and Eq. (6.39). ■

Proposition 6.39. *Let us continue the notation above and let us further assume that A is a finite set and*

$$\mathbb{P}_x(T < \infty) = \mathbb{P}(X_n \in B \text{ for some } n) > 0 \ \forall\, x \in A. \tag{6.40}$$

Under these assumptions, $\mathbb{E}_x T < \infty$ for all $x \in A$ and in particular $\mathbb{P}_x(T < \infty) = 1$ for all $x \in A$. In this case we may may write Eq. (6.38) as

$$(\mathbb{E}_x T)_{x \in A} = (I - Q)^{-1} \mathbf{1} \tag{6.41}$$

where $\mathbf{1}(x) = 1$ for all $x \in A$.

Proof. Since $\{T > n\} \downarrow \{T = \infty\}$ and $\mathbb{P}_x(T = \infty) < 1$ for all $x \in A$ it follows that there exists an $m \in \mathbb{N}$ and $0 \le \alpha < 1$ such that $\mathbb{P}_x(T > m) \le \alpha$ for all $x \in A$. Since $\mathbb{P}_x(T > m) = \sum_{y \in A} Q^m(x,y)$ it follows that the row sums of Q^m are all less than $\alpha < 1$. Further observe that

$$\begin{aligned}\sum_{y \in A} Q^{2m}(x,y) &= \sum_{y,z \in A} Q^m(x,z)\, Q^m(z,y) = \sum_{z \in A} Q^m(x,z) \sum_{y \in A} Q^m(z,y) \\ &\le \sum_{z \in A} Q^m(x,z)\, \alpha \le \alpha^2.\end{aligned}$$

Similarly one may show that $\sum_{y \in A} Q^{km}(x,y) \le \alpha^k$ for all $k \in \mathbb{N}$. Therefore from Eq. (6.39) with m replaced by km, we learn that $\mathbb{P}_x(T > km) \le \alpha^k$ for all $k \in \mathbb{N}$ which then implies that

$$\sum_{y \in A} Q^n(x,y) = \mathbb{P}_x(T > n) \le \alpha^{\lfloor \frac{n}{k} \rfloor} \text{ for all } n \in \mathbb{N},$$

where $\lfloor t \rfloor = m \in \mathbb{N}_0$ if $m \le t < m+1$, i.e. $\lfloor t \rfloor$ is the nearest integer to t which is smaller than t. Therefore, we have

$$\mathbb{E}_x T = \sum_{n=0}^{\infty} \sum_{y \in A} Q^n(x,y) \le \sum_{n=0}^{\infty} \alpha^{\lfloor \frac{n}{m} \rfloor} \le m \cdot \sum_{l=0}^{\infty} \alpha^l = m \frac{1}{1-\alpha} < \infty.$$

So it only remains to prove Eq. (6.41). From the above computations we see that $\sum_{n=0}^{\infty} Q^n$ is convergent. Moreover,

$$(I - Q) \sum_{n=0}^{\infty} Q^n = \sum_{n=0}^{\infty} Q^n - \sum_{n=0}^{\infty} Q^{n+1} = I$$

and therefore $(I - Q)$ is invertible and $\sum_{n=0}^{\infty} Q^n = (I - Q)^{-1}$. Finally,

$$(I - Q)^{-1} \mathbf{1} = \sum_{n=0}^{\infty} Q^n \mathbf{1} = \left(\sum_{n=0}^{\infty} \sum_{y \in A} Q^n(x,y) \right)_{x \in A} = (\mathbb{E}_x T)_{x \in A}$$

as claimed. ■

Remark 6.40. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $\{0, 1, 2, \dots\}$ with 0 being an absorbing state. Let $T = T_{\{0\}}$, i.e. $B = \{0\}$ so that $A = \mathbb{N}$ is now an infinite set. From Remark 6.21, we learn that $\mathbb{E}_i T = \infty$ for all $i > 0$. This shows that we can not in general drop the assumption that A ($A = \{1, 2, \dots\}$ is a finite set the statement of Proposition 6.39.

6.6.1 General facts about sub-probability kernels

Definition 6.41. *Suppose $(A, \mathcal{A})$ is a measurable space. A **sub-probability kernel** on $(A, \mathcal{A})$ is a function $\rho : A \times \mathcal{A} \to [0,1]$ such that $\rho(\cdot, C)$ is $\mathcal{A}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $C \in \mathcal{A}$ and $\rho(x, \cdot) : \mathcal{A} \to [0,1]$ is a measure for all $x \in A$.*

As with probability kernels we will identify ρ with the linear map, $\rho : \mathcal{A}_b \to \mathcal{A}_b$ given by

$$(\rho f)(x) = \rho(x, f) = \int_A f(y)\, \rho(x, dy).$$

Of course we have in mind that $\mathcal{A} = \mathcal{S}_A$ and $\rho = Q_A$. In the following lemma let $\|g\|_\infty := \sup_{x \in A} |g(x)|$ for all $g \in \mathcal{A}_b$.

Theorem 6.42. *Let ρ be a sub-probability kernel on a measurable space $(A, \mathcal{A})$ and define $u_n(x) := (\rho^n 1)(x)$ for all $x \in A$ and $n \in \mathbb{N}_0$. Then;*

1. *u_n is a decreasing sequence so that $u := \lim_{n\to\infty} u_n$ exists and is in $\mathcal{A}_b$. (When $\rho = Q_A$, $u_n(x) = \mathbb{P}_x(H_B > n) \downarrow u(x) = \mathbb{P}(H_B = \infty)$ as $n \to \infty$.)*
2. *The function u satisfies $\rho u = u$.*
3. *If $w \in \mathcal{A}_b$ and $\rho w = w$ then $|w| \le \|w\|_\infty u$. In particular the equation, $\rho w = w$, has a non-zero solution $w \in \mathcal{A}_b$ iff $u \neq 0$.*
4. *If $u = 0$ and $g \in \mathcal{A}_b$, then there is at most one $w \in \mathcal{A}_b$ such that $w = \rho w + g$.*
5. *Let*
$$U := \sum_{n=0}^\infty u_n = \sum_{n=0}^\infty \rho^n 1 : A \to [0,\infty] \tag{6.42}$$
and suppose that $U(x) < \infty$ for all $x \in A$. Then for each $g \in \mathcal{S}_b$,
$$w = \sum_{n=0}^\infty \rho^n g \tag{6.43}$$
is absolutely convergent,
$$|w| \le \|g\|_\infty U, \tag{6.44}$$
$\rho(x, |w|) < \infty$ for all $x \in A$, and w solves $w = \rho w + g$. Moreover if v also solves $v = \rho v + g$ and $|v| \le CU$ for some $C < \infty$ then $v = w$.
Observe that when $\rho = Q_A$,
$$U(x) = \sum_{n=0}^\infty \mathbb{P}_x(H_B > n) = \sum_{n=0}^\infty \mathbb{E}_x(1_{H_B>n}) = \mathbb{E}_x\left(\sum_{n=0}^\infty 1_{H_B>n}\right) = \mathbb{E}_x[H_B].$$
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 $w \le v$. Thus w is the minimal non-negative solution to $v = \rho v + g$.
7. *If there exists $\alpha < 1$ such that $u \le \alpha$ on A then $u = 0$. (When $\rho = Q_A$, this states that $\mathbb{P}_x(H_B = \infty) \le \alpha$ for all $x \in A$ implies $\mathbb{P}_x(T_A = \infty) = 0$ for all $x \in A$.)*
8. *If there exists an $\alpha < 1$ and an $n \in \mathbb{N}$ such that $u_n = \rho^n 1 \le \alpha$ on A, then there exists $C < \infty$ such that*
$$u_k(x) = \left(\rho^k 1\right)(x) \le C\beta^k \text{ for all } x \in A \text{ and } k \in \mathbb{N}_0$$
where $\beta := \alpha^{1/n} < 1$. In particular, $U \le C(1-\beta)^{-1}$ and $u = 0$ under this assumption.
(When $\rho = Q_A$ this assertion states; if $\mathbb{P}_x(H_B > n) \le \alpha$ for all $\alpha \in A$, then $\mathbb{P}_x(H_B > k) \le C\beta^k$ and $\mathbb{E}_x H_B \le C(1-\beta)^{-1}$ for all $k \in \mathbb{N}_0$.)

Proof. We will prove each item in turn.

1. First observe that $u_1(x) = \rho(x, A) \le 1 = u_0(x)$ and therefore,
$$u_{n+1} = \rho^{n+1} 1 = \rho^n u_1 \le \rho^n 1 = u_n.$$
We now let $u := \lim_{n\to\infty} u_n$ so that $u : A \to [0,1]$.
2. Using DCT we may let $n \to \infty$ in the identity, $\rho u_n = u_{n+1}$ in order to show $\rho u = u$.
3. If $w \in \mathcal{A}_b$ with $\rho w = w$, then
$$|w| = |\rho^n w| \le \rho^n |w| \le \|w\|_\infty \rho^n 1 = \|w\|_\infty \cdot u_n.$$
Letting $n \to \infty$ shows that $|w| \le \|w\|_\infty u$.
4. If $w_i \in \mathcal{A}_b$ solves $w_i = \rho w_i + g$ for $i = 1, 2$ then $w := w_2 - w_1$ satisfies $w = \rho w$ and therefore $|w| \le Cu = 0$.
5. Let $U := \sum_{n=0}^\infty u_n = \sum_{n=0}^\infty \rho^n 1 : A \to [0,\infty]$ and suppose $U(x) < \infty$ for all $x \in A$. Then $u_n(x) \to 0$ as $n \to \infty$ and so bounded solutions to $\rho u = u$ are necessarily zero. Moreover we have, for all $k \in \mathbb{N}_0$, that
$$\rho^k U = \sum_{n=0}^\infty \rho^k u_n = \sum_{n=0}^\infty u_{n+k} = \sum_{n=k}^\infty u_n \le U. \tag{6.45}$$
Since the tails of convergent series tend to zero it follows that $\lim_{k\to\infty} \rho^k U = 0$.
Now if $g \in \mathcal{S}_b$, we have
$$\sum_{n=0}^\infty |\rho^n g| \le \sum_{n=0}^\infty \rho^n |g| \le \sum_{n=0}^\infty \rho^n \|g\|_\infty = \|g\|_\infty \cdot U < \infty \tag{6.46}$$
and therefore $\sum_{n=0}^\infty \rho^n g$ is absolutely convergent. Making use of Eqs. (6.45) and (6.46) we see that
$$\sum_{n=1}^\infty \rho |\rho^n g| \le \|g\|_\infty \cdot \rho U \le \|g\|_\infty U < \infty$$
and therefore (using DCT),

$$w = \sum_{n=0}^{\infty} \rho^n g = g + \sum_{n=1}^{\infty} \rho^n g$$
$$= g + \rho \sum_{n=1}^{\infty} \rho^{n-1} g = g + \rho w,$$

i.e. w solves $w = g + \rho w$.
If $v : A \to \mathbb{R}$ is measurable such that $|v| \le CU$ and $v = g + \rho v$, then $y := w - v$ solves $y = \rho y$ with $|y| \le (C + \|g\|_\infty) U$. It follows that

$$|y| = |\rho^n y| \le (C + \|g\|_\infty) \rho^n U \to 0 \text{ as } n \to \infty,$$

i.e. $0 = y = w - v$.

6. If $g \ge 0$ we may always define w by Eq. (6.43) allowing for $w(x) = \infty$ for some or even all $x \in A$. As in the proof of the previous item (with DCT being replaced by MCT), it follows that $w = \rho w + g$. If $v \ge 0$ also solves $v = g + \rho v$, then

$$v = g + \rho (g + \rho v) = g + \rho g + \rho^2 v$$

and more generally by induction we have

$$v = \sum_{k=0}^{n} \rho^k g + \rho^{n+1} v \ge \sum_{k=0}^{n} \rho^k g.$$

Letting $n \to \infty$ in this last equation shows that $v \ge w$.

7. If $u \le \alpha < 1$ on A, then by item 3. with $w = u$ we find that

$$u \le \|u\|_\infty \cdot u \le \alpha u$$

which clearly implies $u = 0$.

8. If $u_n \le \alpha < 1$, then for any $m \in \mathbb{N}$ we have,

$$u_{n+m} = \rho^m u_n \le \alpha \rho^m 1 = \alpha u_m.$$

Taking $m = kn$ in this inequality shows, $u_{(k+1)n} \le \alpha u_{kn}$. Thus a simple induction argument shows $u_{kn} \le \alpha^k$ for all $k \in \mathbb{N}_0$. For general $l \in \mathbb{N}_0$ we write $l = kn + r$ with $0 \le r < n$. We then have,

$$u_l = u_{kn+r} \le u_{kn} \le \alpha^k = \alpha^{\frac{l-r}{n}} = C\alpha^{l/n}$$

where $C = \alpha^{-\frac{n-1}{n}}$.

■

Corollary 6.43. *If $h : B \to [0, \infty]$ is measurable, then $u(x) := \mathbb{E}_x [h(X_{H_B}) : H_B < \infty]$ is the unique minimal non-negative solution to Eq. (6.36) while if $g : A \to [0, \infty]$ is measurable, then $u(x) = \mathbb{E}_x \left[\sum_{n<H_B} g(X_n)\right]$ is the unique minimal non-negative solution to Eq. (6.37).*

Exercise 6.16. Keeping the notation of Exercise 6.9 and 6.11. Use Corollary 6.43 to show again that $\mathbb{P}_x(H_B < \infty) = (q/p)^x$ for all $x > 0$ and $\mathbb{E}_x T_0 = x/(q-p)$ for $x < 0$. You should do so without making use of the extraneous hitting times, T_n for $n \ne 0$.

Corollary 6.44. *If $\mathbb{P}_x(H_B = \infty) = 0$ for all $x \in A$ and $h : B \to \mathbb{R}$ is a bounded measurable function, then $u(x) := \mathbb{E}_x [h(X_{H_B})]$ is the **unique** solution to Eq. (6.36).*

Corollary 6.45. *Suppose now that $A = B^c$ is a finite subset of S such that $\mathbb{P}_x(H_B = \infty) < 1$ for all $x \in A$. Then there exists $C < \infty$ and $\beta \in (0,1)$ such that $\mathbb{P}_x(H_B > n) \le C\beta^n$ and in particular $\mathbb{E}_x H_B < \infty$ for all $x \in A$.*

Proof. Let $\alpha_0 = \max_{x \in A} \mathbb{P}_x(H_B = \infty) < 1$. We know that

$$\lim_{n\to\infty} \mathbb{P}_x(H_B > n) = \mathbb{P}_x(H_B = \infty) \le \alpha_0 \text{ for all } x \in A.$$

Therefore if $\alpha \in (\alpha_0, 1)$, using the fact that A is a finite set, there exists an n sufficiently large such that $\mathbb{P}_x(H_B > n) \le \alpha$ for all $x \in A$. The result now follows from item 8. of Theorem 6.42. ■

7

Markov Conditioning

We assume the $\{X_n\}_{n=0}^\infty$ is a Markov chain with values in S and transition kernel $\mathbf{P}$ and $\pi : S \to [0,1]$ is a probability on S. As usual we write $\mathbb{P}_\pi$ for the unique probability satisfying Eq. (5.2) and we will often write $p(x,y)$ for $\mathbf{P}_{xy}$. The most important result of this section is the strong Markov property stated in Theorem 7.8 below.

7.1 The Markov Property

Notation 7.1 *Let $\left\{\tilde{f}_n\right\}_{n=0}^\infty$ be an independent copy of the $\{f_n\}_{n=0}^\infty$ and for each $x \in S$ let $\{Y_n^x\}_{n=0}^\infty$ be the Markov chain defined inductively by,*

$$Y_0^x = x \text{ and } Y_{n+1}^x = \tilde{f}_n\left(Y_n^x\right) \text{ for } n \in \mathbb{N}_0.$$

It should be clear to the reader that

$$\mathbb{E}\left[F\left(Y_0^x, Y_1^x, \dots\right)\right] = \mathbb{E}_x\left[F\left(X_0, X_1, \dots\right)\right]$$

for all bounded functions $(x_0, x_1, \dots) \to F(x_0, x_1, \dots)$. The next two theorems generalize these statements in far reaching ways.

Theorem 7.2 (Markov Property). *Given any $m \in \mathbb{N}$, then*

$$\left(X_0, X_1, \dots, X_m, X_{m+1}, \dots\right) \overset{d}{=} \left(X_0, X_1, \dots, X_m, Y_1^{X_m}, Y_2^{X_m}, \dots\right).$$

Proof. Let

$$\tilde{X}_k := \begin{cases} X_k & \text{if } k \le m \\ Y_{k-m}^{X_m} & \text{if } k \ge m \end{cases},$$

$N \in \mathbb{N}$ such that $m \le N$, and $x_0, x_1, \dots, x_N \in S$. Then

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_0 = x_0, \tilde{X}_1 = x_1, \dots \tilde{X}_N = x_N\right) \\
&= \mathbb{P}\left(X_0 = x_0, \dots, X_m = x_m, Y_1^{x_m} = x_{m+1}, \dots, Y_{N-m}^{x_m} = x_N\right) \\
&= \mathbb{P}\left(X_0 = x_0, \dots, X_m = x_m\right) \mathbb{P}\left(Y_1^{x_m} = x_{m+1}, \dots, Y_{N-m}^{x_m} = x_N\right) \qquad (7.1)
\end{aligned}$$

wherein we have used the independence of $\left(Y_1^{x_m}, \dots, Y_{N-m}^{x_m}\right)$ from $\mathbb{X}$. Furthermore, as in the proof of Theorem 5.11,

$$\begin{aligned}
&\mathbb{P}\left(Y_1^{x_m} = x_{m+1}, \dots, Y_{N-m}^{x_m} = x_N\right) \\
&= \mathbb{P}\left(\tilde{f}_0\left(x_m\right) = x_{m+1}, \dots, \tilde{f}_{N-m-1}\left(x_{N-1}\right) = x_N\right) \\
&= p\left(x_m, x_{m+1}\right) \dots p\left(x_{N-1}, x_N\right)
\end{aligned}$$

which combined with Eq. (7.1) and Theorem 5.11 shows,

$$\mathbb{P}\left(\tilde{X}_0 = x_0, \tilde{X}_1 = x_1, \dots \tilde{X}_N = x_N\right) = \mathbb{P}\left(X_0 = x_0, X_1 = x_1, \dots X_N = x_N\right).$$

■

Remark 7.3. Here is the intuition behind the proof of Theorem 7.2. The trajectory $(X_0, X_1, \dots, X_m, X_{m+1}, X_{m+2}, \dots)$ is constructed using the original spinors that determine the $\{f_n\}$ while the trajectory $\left(X_0, X_1, \dots, X_m, Y_1^{X_m}, Y_2^{X_m}, \dots\right)$ is determined by using identical but independent spinors (i.e. the $\left\{\tilde{f}_n\right\}$) at time m and beyond. It should be intuitively clear that the statistics of these two trajectories should be the same and this is the content of Theorem 7.2.

We end this section with some consequences and variants of Theorem 7.2. For the first application let us give another proof of Proposition 5.13.

Corollary 7.4. *If $g : S \to \mathbb{R}$ is a bounded function, then*

$$\mathbb{E}\left[g\left(X_{m+1}\right) | \left(X_0, \dots, X_m\right)\right] = \left(\mathbf{P}g\right)\left(X_m\right) \qquad (7.2)$$

and

$$\mathbb{E}\left[g\left(X_{m+1}\right) | X_m\right] = \left(\mathbf{P}g\right)\left(X_m\right).$$

Proof. Let $u : S^{m+1} \to \mathbb{R}$ be another bounded function, then

$$\begin{aligned}
\mathbb{E}\left[g\left(X_{m+1}\right)u\left(X_0,\dots,X_m\right)\right] &= \mathbb{E}\left[g\left(Y_1^{X_m}\right)u\left(X_0,\dots,X_m\right)\right] \\
&= \sum_{x\in S}\mathbb{E}\left[g\left(Y_1^x\right)u\left(X_0,\dots,X_{m-1},x\right)1_{X_m=x}\right] \\
&= \sum_{x\in S}\mathbb{E}\left[g\left(Y_1^x\right)\right]\cdot\mathbb{E}\left[u\left(X_0,\dots,X_{m-1},x\right)1_{X_m=x}\right] \\
&= \sum_{x\in S}\left(\mathbf{P}g\right)(x)\cdot\mathbb{E}\left[u\left(X_0,\dots,X_{m-1},x\right)1_{X_m=x}\right] \\
&= \sum_{x\in S}\mathbb{E}\left[\left(\mathbf{P}g\right)(x)\,u\left(X_0,\dots,X_{m-1},x\right)1_{X_m=x}\right] \\
&= \mathbb{E}\left[\left(\mathbf{P}g\right)\left(X_m\right)u\left(X_0,\dots,X_{m-1},X_m\right)\right]
\end{aligned}$$

from which Eq. (7.2) follows. The second equation follows from Eq. (7.2) and the tower property of conditional expectation;

$$\begin{aligned}
\left(\mathbf{P}g\right)\left(X_m\right) &= \mathbb{E}\left[\left(\mathbf{P}g\right)\left(X_m\right)|X_m\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g\left(X_{m+1}\right)|\left(X_0,\dots,X_m\right)\right]|X_m\right] = \mathbb{E}\left[g\left(X_{m+1}\right)|X_m\right].
\end{aligned}$$

■

Theorem 7.5 (Markov conditioning). *Let $X=\{X_n\}_{n=0}^\infty$ be a Markov chain with transition kernel p and for each $x\in S$ let $\{Y_n\}_{n=0}^\infty$ be another Markov chain with transition kernel p starting at $Y_0=x$ which is independent of X. Then relative to $\mathbb{P}(\cdot|X_m=x)$ we have,*

$$X=\left(X_0,X_1,\dots\right)\overset{d}{=}\left(X_0,\dots,X_{m-1},x,Y_1,Y_2,\dots\right).$$

In more detail we are asserting,

$$\mathbb{E}\left[F\left(X\right)|X_m=x\right]=\mathbb{E}\left[F\left(X_0,\dots,X_{m-1},x,Y_1,Y_2,\dots\right)|X_m=x\right]$$

for all $F(X)=F(X_0,X_1,\dots)$, where F is either bounded or non-negative.

Proof. First Proof. By Theorem 7.2,

$$\begin{aligned}
\mathbb{E}\left[F\left(X\right):X_m=x\right] &= \mathbb{E}\left[F\left(X_0,X_1,\dots,X_m,Y_1^{X_m},Y_2^{X_m},\dots\right):X_m=x\right] \\
&= \mathbb{E}\left[F\left(X_0,X_1,\dots,X_{m-1},x,Y_1^x,Y_2^x,\dots\right):X_m=x\right]
\end{aligned}$$

which upon dividing by $\mathbb{P}(X_m=x)$ completes the proof.

Second Proof. In this proof we use the **fact**[1] that it suffices to verify the assertions of the Theorem for $F(x)=F(x_0,\dots,x_N)$, i.e. $F(x)$ depends on only finitely many coordinates. For any $k\in\mathbb{N}$ let

$$\rho_k\left(x_0,\dots,x_k\right):=\prod_{l=1}^k p\left(x_{l-1},x_l\right)$$

and observe that

$$\rho_m\left(x_0,\dots,x_{m-1},x_m\right)\rho_n\left(x_m,y_1,\dots,y_n\right)=\rho_{m+n}\left(x_0,\dots,x_{m-1},x_m,y_1,\dots,y_n\right).$$

Now let $N=m+n$, $\mathbf{x}:=(x_0,\dots,x_{m-1})\in S^m$, and $\mathbf{y}=(y_1,\dots,y_n)\in S^n$. We then have,

$$\begin{aligned}
&\mathbb{E}\left[F\left(X_0,\dots,X_{m-1},x,Y_1,Y_2,\dots,Y_n\right):X_m=x\right] \\
&= \sum_{\mathbf{x}}\sum_{\mathbf{y}}F\left(\mathbf{x},x,\mathbf{y}\right)\pi\left(x_0\right)\rho_m\left(\mathbf{x},x\right)\rho_n\left(x,\mathbf{y}\right) \\
&= \sum_{(\mathbf{x},\mathbf{y})}F\left(\mathbf{x},x,\mathbf{y}\right)\pi\left(x_0\right)\rho_{m+n}\left(\mathbf{x},x,\mathbf{y}\right) \\
&= \sum_{(x_0,\dots,x_{m+n})}F\left(x_0,\dots,x_{m+n}\right)1_{x_m=x}\pi\left(x_0\right)\rho_{m+n}\left(x_0,\dots\dots,x_{m+n}\right) \\
&= \mathbb{E}\left[F\left(X_0,\dots,X_N\right):X_m=x\right].
\end{aligned}$$

■

Corollary 7.6. *Again let $Y=(Y_0,Y_1,\dots)$ be a Markov chain independent of $X=(X_0,X_1,\dots)$ but having the same transition matrix, $\mathbf{P}$. If $m\in\mathbb{N}$ we have*

$$\mathbb{E}\left[F\left(X\right)|\left(X_0,\dots,X_m\right)\right]=\mathbb{E}_{X_m}^{(Y)}F\left(X_0,\dots,X_{m-1},Y_0,Y_1,Y_2,\dots\right),$$

where $\mathbb{E}_{X_m}^{(Y)}$ indicates the expectation in the Y – process only keeping the X – process "frozen." Put another way,

$$\mathbb{E}\left[F\left(X\right)|\left(X_0,\dots,X_m\right)\right]=h\left(X_0,\dots,X_m\right)$$

where

$$h\left(x_0,\dots,x_m\right)=\mathbb{E}_{x_m}\left[F\left(x_0,\dots,x_{m-1},X_0,X_1,\dots\right)\right]. \tag{7.3}$$

Proof. By Theorem 7.2,

$$\begin{aligned}
&\mathbb{E}\left[F\left(X\right):\left(X_0,\dots,X_m\right)=\left(x_0,\dots,x_m\right)\right] \\
&= \mathbb{E}\left[F\left(x_0,\dots,x_m,X_{m+1},X_{m+2},\dots\right):\left(X_0,\dots,X_m\right)=\left(x_0,\dots,x_m\right)\right] \\
&= \mathbb{E}\left[F\left(x_0,\dots,x_m,Y_1^{x_m},Y_2^{x_m},\dots\right):\left(X_0,\dots,X_m\right)=\left(x_0,\dots,x_m\right)\right] \\
&= \mathbb{E}\left[F\left(x_0,\dots,x_m,Y_1^{x_m},Y_2^{x_m},\dots\right)\right]\cdot\mathbb{P}\left(\left(X_0,\dots,X_m\right)=\left(x_0,\dots,x_m\right)\right) \\
&= \mathbb{E}_{x_m}\left[F\left(x_0,\dots,x_{m-1},X_0=x_m,X_1,\dots\right)\right]\cdot\mathbb{P}\left(\left(X_0,\dots,X_m\right)=\left(x_0,\dots,x_m\right)\right).
\end{aligned}$$

[1] We are using a standard measure theoretic result here which goes by the name of Dynkin's multiplicative system theorem. One could also use the so called $\pi-\lambda$ theorem or the monotone class theorem as well. See Math 280 for these sorts of arguments.

Dividing this equation by then shows that h defined in Eq. (7.3) is also given by

$$h(x_0,\dots,x_m) = \mathbb{E}\left[F(X) \,|\, (X_0,\dots,X_m) = (x_0,\dots,x_m)\right]$$

which along with Proposition 3.10 completes the proof. ■

Corollary 7.7. *For each $m \in \mathbb{N}$ and $x \in S$ such that $\mathbb{P}(X_m = x) > 0$, $(X_0,\dots,X_{m-1})$ and $(X_{m+1},X_{m+2},\dots)$ are independent relative to $\mathbb{P}(\cdot|X_m = x)$. Stated more poetically, the past is independent of the future given the present.*

Proof. Let $F(x_0,\dots,x_{m-1})$ and $G(x_{m+1},x_{m+2},\dots)$ be given bounded functions, then

$$\begin{aligned}
&\mathbb{E}\left[F(X_0,\dots,X_{m-1})\,G(X_{m+1},X_{m+2},\dots) : X_m = x\right] \\
&\quad = \mathbb{E}\left[F(X_0,\dots,X_{m-1})\,G(Y_1^x,Y_2^x,\dots) : X_m = x\right] \\
&\quad = \mathbb{E}\left[F(X_0,\dots,X_{m-1}) : X_m = x\right]\cdot \mathbb{E}G(Y_1^x,Y_2^x,\dots)
\end{aligned}$$

and so dividing this equation by $\mathbb{P}(X_m = x)$ shows,

$$\begin{aligned}
&\mathbb{E}\left[F(X_0,\dots,X_{m-1})\,G(X_{m+1},X_{m+2},\dots)\,|X_m = x\right] \\
&\quad = \mathbb{E}\left[F(X_0,\dots,X_{m-1})\,|X_m = x\right]\cdot \mathbb{E}G(Y_1^x,Y_2^x,\dots).
\end{aligned}$$

Taking $F \equiv 1$ in this equation implies

$$\mathbb{E}\left[G(X_{m+1},X_{m+2},\dots)\,|X_m = x\right] = \mathbb{E}G(Y_1^x,Y_2^x,\dots)$$

and hence,

$$\begin{aligned}
&\mathbb{E}\left[F(X_0,\dots,X_{m-1})\,G(X_{m+1},X_{m+2},\dots)\,|X_m = x\right] \\
&\quad = \mathbb{E}\left[F(X_0,\dots,X_{m-1})\,|X_m = x\right]\cdot \mathbb{E}\left[G(X_{m+1},X_{m+2},\dots)\,|X_m = x\right]
\end{aligned}$$

which is the desired independence statement. ■

7.2 The Strong Markov Property

The results in Section 7.1 have a far reaching extension to the case where $m \in \mathbb{N}_0$ is replaced by a stopping time τ. The strong Markov property will play an important role in Chapter 8 on the limiting behaviors of Markov chains.

Theorem 7.8 (Strong Markov Property). *Again let $\{Y_n^x\}_{n=0}^\infty$ be the Markov chain (independent of X) described in Notation 7.1. If τ is a $\mathcal{F}^{\mathbb{X}}$ - stopping time and*

$$\tilde{X}_n := \begin{cases} X_n & \text{if } n < \tau \\ Y_{n-\tau}^{X_\tau} & \text{if } n \ge \tau \end{cases} \quad \text{for } n \in \mathbb{N}_0,$$

then

$$\left(\tilde{X}_0,\tilde{X}_1,\dots\right) \overset{d}{=} (X_0,X_1,\dots).$$

Alternatively this may be stated,

$$\left(X_0,X_1,\dots,X_\tau,Y_1^{X_\tau},Y_2^{X_\tau},\dots\right) \overset{d}{=} (X_0,X_1,\dots,X_\tau,X_{\tau+1},\dots) \text{ on } \{\tau < \infty\}.$$

Proof. The intuitive proof of this theorem is essentially the same as the one described in Remark 7.3. For the formal proof, let $N \in \mathbb{N}$, $n \le N$, and then find $A \subset S^{n+1}$ such that

$$\{\tau = n\} = \{(X_0,\dots,X_N) \in A\}.$$

The for any $x_0,x_1,\dots,x_N \in S$, we have

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_0 = x_0,\tilde{X}_1 = x_1,\dots\tilde{X}_N = x_N,\tau = n\right) \\
&\quad = \mathbb{P}\left(X_0 = x_0,\dots,X_n = x_n,Y_1^{x_n} = x_{n+1},\dots,Y_{N-n}^{x_n} = x_N,\tau = n\right) \\
&\quad = 1_A(x_0,x_1,\dots,x_n)\cdot\mathbb{P}\left(X_0 = x_0,\dots X_n = x_n,Y_1^{x_n} = x_{n+1},\dots,Y_{N-n}^{x_n} = x_N\right) \\
&\quad = 1_A(x_0,x_1,\dots,x_n)\cdot\mathbb{P}\left(X_0 = x_0,\dots X_n = x_n,X_{n+1} = x_{n+1},\dots,X_N = x_N\right) \\
&\quad = \mathbb{P}\left(X_0 = x_0,\dots,X_n = x_n,X_{n+1} = x_{n+1},\dots,X_N = x_N,\tau = n\right)
\end{aligned}$$

wherein the second to last inequality we used Theorem 7.2. Summing the previous equation on $n \le N$ shows,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_0 = x_0,\tilde{X}_1 = x_1,\dots,\tilde{X}_N = x_N,\tau \le N\right) \\
&\quad = \mathbb{P}\left(X_0 = x_0,\dots X_n = x_n,X_{n+1} = x_{n+1},\dots,X_N = x_N,\tau \le N\right).
\end{aligned} \tag{7.4}$$

Finally since

$$(X_0,\dots,X_N) = \left(\tilde{X}_0,\dots,\tilde{X}_N\right) \text{ on } \{\tau > N\}$$

we conclude that

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_0 = x_0,\tilde{X}_1 = x_1,\dots,\tilde{X}_N = x_N,\tau > N\right) \\
&\quad = \mathbb{P}\left(X_0 = x_0,\dots,X_N = x_N,\tau > N\right).
\end{aligned} \tag{7.5}$$

Summing Eqs. (7.4) and (7.5) gives

$$\mathbb{P}\left(\tilde{X}_0 = x_0,\dots,\tilde{X}_N = x_N\right) = \mathbb{P}\left(X_0 = x_0,\dots,X_N = x_N\right)$$

which completes the proof.

■

Notation 7.9 *Let $\{Y_n\}_{n=0}^\infty$ be a Markov chain independent of $\{X_n\}$ but having the same transition probabilities, $\mathbf{P}$. Further let $\mathbb{E}_y^{(Y)}$ denote the expectation relative to the chain $\mathbb{Y}$ started at $y \in S$*

Corollary 7.10. *If τ is an $\mathcal{F}^{\mathbb{X}}$ – stopping time and $(x_0, x_1, \dots) \to F(x_0, x_1, \dots)$ is a bounded (or non-negative) function, then*

$$\mathbb{E}[F(X) : \tau < \infty] = \mathbb{E}\left[\mathbb{E}_{X_\tau}^{(Y)}[F(X_0, \dots, X_\tau, Y_1, Y_2, \dots)] : \tau < \infty\right].$$

Alternatively if for each $n \in \mathbb{N}_0$, we let

$$\begin{aligned} F_n(x_0, x_1, \dots, x_n) &:= \mathbb{E}_{x_n}^{(Y)} F(x_0, \dots, x_n, Y_1, Y_2, \dots) \\ &= \mathbb{E}_{x_n} F(x_0, \dots, x_n, X_1, X_2, \dots), \end{aligned}$$

then

$$\mathbb{E}[F(X) : \tau < \infty] = \mathbb{E}[F_\tau(X_0, \dots, X_\tau) : \tau < \infty].$$

Proof. By Theorem 7.8,

$$\begin{aligned} \mathbb{E}[F(X) : \tau < \infty] &= \mathbb{E}\left[F\left(\tilde{X}\right) : \tau < \infty\right] \\ &= \mathbb{E}\left[F\left(X_0, X_1, \dots, X_\tau, Y_1^{X_\tau}, Y_2^{X_\tau}, \dots\right) : \tau < \infty\right] \\ &= \mathbb{E}\left[\mathbb{E}_{X_\tau}^{(Y)}[F(X_0, \dots, X_\tau, Y_1, Y_2, \dots)] : \tau < \infty\right]. \end{aligned}$$

■

Let us end this section with one more variant of the strong Markov property.

Theorem 7.11 (Strong Markov Property). *Let $x \in X$ and $\{Y_n\}_{n=0}^\infty$ be as in Notation 7.9 and τ be a $\mathcal{F}^{\mathbb{X}}$ – stopping time. Then given $\{\tau < \infty, X_\tau = x\}$ we have*

$$(X_0, X_1, \dots) \overset{d}{=} (X_0, X_1, \dots, X_{\tau-1}, X_\tau, Y_1, Y_2, \dots), \tag{7.6}$$

i.e.

$$(X_0, X_1, \dots) \overset{\mathbb{P}(\cdot|\tau<\infty, X_\tau = x)}{=} (X_0, X_1, \dots, X_{\tau-1}, X_\tau, Y_1, Y_2, \dots)$$

We also have,

$$\begin{aligned} &\mathbb{E}[F(X_0, X_1, \dots) : \tau < \infty] \\ &= \mathbb{E}\left[\mathbb{E}_{X_\tau}^{(Y)}[F(X_0, X_1, \dots, X_{\tau-1}, Y_0, Y_1, Y_2, \dots)] : \tau < \infty\right] \\ &= \mathbb{E}\left[\mathbb{E}_{X_\tau}^{(Y)}[F(X_0, X_1, \dots, X_{\tau-1}, X_\tau, Y_1, Y_2, \dots)] : \tau < \infty\right]. \end{aligned} \tag{7.7}$$

Proof. Using

$$1_{\tau=n} = f(X_0, X_1, \dots, X_n),$$

it follows from Theorem 7.5 that

$$\begin{aligned} &\mathbb{E}_\nu[F(X_0, X_1, \dots) : \tau = n, X_n = x] \\ &\quad = \mathbb{E}_\nu[F(X_0, X_1, \dots, X_{n-1}, x, Y_1, Y_2, \dots) : \tau = n, X_n = x]. \end{aligned}$$

Summing this equation on n and using $\{\tau = n, X_n = x\} = \{\tau = n, X_\tau = x\}$, we learn that

$$\begin{aligned} &\mathbb{E}_\nu[F(X_0, X_1, \dots) : \tau < \infty, X_\tau = x] \\ &\quad = \sum_{n=0}^\infty \mathbb{E}_\nu[F(X_0, X_1, \dots) : \tau = n, X_n = x] \\ &\quad = \sum_{n=0}^\infty \mathbb{E}_\nu[F(X_0, X_1, \dots, X_{n-1}, x, Y_1, Y_2, \dots) : \tau = n, X_n = x] \\ &\quad = \sum_{n=0}^\infty \mathbb{E}_\nu[F(X_0, X_1, \dots, X_{\tau-1}, X_\tau, Y_1, Y_2, \dots) : \tau = n, X_\tau = x] \\ &\quad = \mathbb{E}_\nu[F(X_0, X_1, \dots, X_{\tau-1}, X_\tau, Y_1, Y_2, \dots) : \tau < \infty, X_\tau = x] \end{aligned} \tag{7.8}$$

which suffices to prove Eq. (7.6).

We may rewrite Eq. (7.8) as,

$$\begin{aligned} &\mathbb{E}_\nu[F(X_0, X_1, \dots) : \tau < \infty, X_\tau = x] \\ &\quad = \mathbb{E}_\nu\left[\mathbb{E}_x^{(Y)}[F(X_0, X_1, \dots, X_{\tau-1}, Y_0, Y_1, Y_2, \dots)] : \tau < \infty, X_\tau = x\right] \\ &\quad = \mathbb{E}_\nu\left[\mathbb{E}_{X_\tau}^{(Y)}[F(X_0, X_1, \dots, X_{\tau-1}, Y_0, Y_1, Y_2, \dots)] : \tau < \infty, X_\tau = x\right]. \end{aligned}$$

Summing this last equation on $x \in S$ then gives Eq. (7.7). ■

Here is a special case of Theorem 7.11.

Theorem 7.12 (Strong Markov Property). *Let $\left(\{X_n\}_{n=0}^\infty, \{\mathbb{P}_x\}_{x\in S}, p\right)$ be Markov chain as above and $\tau : \Omega \to [0, \infty]$ be a stopping time. Then*

$$\begin{aligned} &\mathbb{E}_\pi[f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}] \\ &\quad = \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau}^{(Y)} f(Y_0, Y_1, \dots)\right] g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right]. \end{aligned} \tag{7.9}$$

for all $f, g = \{g_n\} \geq 0$ or f and g bounded. In other words,

$$\mathbb{E}_\pi[f(X_\tau, X_{\tau+1}, \dots) 1_{\tau<\infty} | \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}^{(Y)}[f(Y_0, Y_1, \dots)].$$

Proof. Apply Eq. (7.7) of Theorem 7.11 with $F(X_0, X_1, \dots) = f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}$.

■

7.3 Strong Markov Proof of Hitting Times Estimates*

In this optional subsection we reprove the results in Section 5.3 by making use of the strong Markov property.

Corollary 7.13. *Let $B \subset S$ and H_B be as above, then for $n, m \in \mathbb{N}$ we have*

$$\mathbb{P}_\pi (H_B > m + n) = \mathbb{E}_\pi \left[1_{H_B > m} \mathbb{P}_{X_m} [H_B > n]\right]. \tag{7.10}$$

Proof. First proof. By the Markov property in Theorem 7.2, for each $y \in S$, we have

$$(X_0, \dots, X_m, Y_1^y, Y_2^y, \dots) \overset{\mathbb{P}_\pi(\cdot|X_m=y)}{=} (X_0, X_1, \dots).$$

This observation along with the identity,

$$\{H_B(X) > m + n\} = \{H_B(X) > m, H_B(X_m, X_{m+1}, \dots) > n\}$$

leads to

$$\begin{aligned} h(y) &= \mathbb{P}_\pi (H_B(X) > m + n | X_m = y) \\ &= \mathbb{P}_\pi (H_B(X) > m, H_B(X_m, X_{m+1}, \dots) > n | X_m = y) \\ &= \mathbb{P}_\pi (H_B(X) > m, H_B(Y_0^y, Y_1^y, Y_2^y, \dots) > n | X_m = y) \\ &= \mathbb{P}_\pi (H_B(X) > m | X_m = y) \mathbb{P}(H_B(Y_0^y, Y_1^y, Y_2^y, \dots) > n) \\ &= \mathbb{P}_\pi (H_B > m | X_m = y) \mathbb{P}_y (H_B > m) \\ &= \mathbb{E}_\pi (1_{H_B > m} | X_m = y) \mathbb{P}_y (H_B > m). \end{aligned}$$

Therefore it follows that

$$\begin{aligned} \mathbb{P}_\pi (H_B > m + n) &= \mathbb{E}_\pi \mathbb{E}_\pi \left[1_{H_B > m+n} | X_m\right] = \mathbb{E}_\pi \left[h(X_m)\right] \\ &= \mathbb{E}_\pi \left[\mathbb{E}_\pi (1_{H_B > m} | X_m) \cdot \mathbb{P}_{X_m} (H_B > m)\right]. \end{aligned}$$

Second proof. Using Theorem 7.11,

$$\begin{aligned} \mathbb{P}_\pi (H_B > m + n) &= \mathbb{E}_\pi \left[1_{H_B(X) > m+n}\right] \\ &= \mathbb{E}_\pi \left[\mathbb{E}_{X_m}^{(Y)} \left[1_{H_B(X_0, \dots, X_{m-1}, Y_0, Y_1, \dots) > m+n}\right]\right] \\ &= \mathbb{E}_\pi \left[\mathbb{E}_{X_m}^{(Y)} \left[1_{H_B(X) > m} \cdot 1_{H_B(Y) > n}\right]\right] \\ &= \mathbb{E}_\pi \left[1_{H_B(X) > m} \mathbb{E}_{X_m}^{(Y)} \left[1_{H_B(Y) > n}\right]\right] = \mathbb{E}_\pi \left[1_{H_B > m} \mathbb{P}_{X_m} [H_B > n]\right]. \end{aligned}$$

■

Corollary 7.14. *Suppose that $B \subset S$ is non-empty proper subset of S and $A = S \setminus B$. Further suppose there is some $\alpha < 1$ such that $\mathbb{P}_x (H_B = \infty) \le \alpha$ for all $x \in A$, then $\mathbb{P}_\pi (H_B = \infty) = 0$. [In words; if there is a "uniform" chance that X hits B starting from any site, then X will surely hit B from any point in A.]*

Proof. Since $H_B = 0$ on $\{X_0 \in B\}$ we in fact have $\mathbb{P}_x (H_B = \infty) \le \alpha$ for all $x \in S$. Letting $n \to \infty$ in Eq. (7.10) shows,

$$\mathbb{P}_\pi (H_B = \infty) = \mathbb{E}_\pi \left[1_{H_B > m} \mathbb{P}_{X_m} [H_B = \infty]\right] \le \mathbb{E}_\pi \left[1_{H_B > m} \alpha\right] = \alpha P_\pi (H_B > m).$$

Now letting $m \to \infty$ in this equation shows $\mathbb{P}_\pi (H_B = \infty) \le \alpha P_\pi (H_B = \infty)$ from which it follows that $\mathbb{P}_\pi (H_B = \infty) = 0$. ■

Corollary 7.15. *Suppose that $B \subset S$ is non-empty proper subset of S and $A = S \setminus B$. Further suppose there is some $\alpha < 1$ and $n < \infty$ such that $\mathbb{P}_x (H_B > n) \le \alpha$ for all $x \in A$, then*

$$\mathbb{E}_\pi (H_B) \le \frac{n}{1 - \alpha} < \infty$$

for all $x \in A$. [In words; if there is a "uniform" chance that X hits B starting from any site within a fixed number of steps, then the expected hitting time of B is finite and bounded independent of the starting distribution.]

Proof. Again using $H_B = 0$ on $\{X_0 \in B\}$ we may conclude that $\mathbb{P}_x (H_B > n) \le \alpha$ for all $x \in S$. Letting $m = kn$ in Eq. (7.10) shows

$$\begin{aligned} \mathbb{P}_\pi (H_B > kn + n) &= \mathbb{E}_\pi \left[1_{H_B > kn} \mathbb{P}_{X_m} [H_B > n]\right] \\ &\le \mathbb{E}_\pi \left[1_{H_B > kn} \cdot \alpha\right] = \alpha P_\pi (H_B > kn). \end{aligned}$$

Iterating this equation using the fact that $\mathbb{P}_\pi (H_B > 0) \le 1$ shows $\mathbb{P}_\pi (H_B > kn) \le \alpha^k$ for all $k \in \mathbb{N}_0$. Therefore with the aid of Lemma 5.18 and the observation,

$$\mathbb{P}(H_B > kn + m) \le \mathbb{P}(H_B > kn) \text{ for } m = 0, \dots, n - 1,$$

we find,

$$\begin{aligned} \mathbb{E}_x H_B &= \sum_{k=0}^{\infty} \mathbb{P}(H_B > k) \le \sum_{k=0}^{\infty} n P(H_B > kn) \\ &\le \sum_{k=0}^{\infty} n \alpha^k = \frac{n}{1 - \alpha} < \infty. \end{aligned}$$

■

8

Long Run Behavior of Discrete Markov Chains

In this chapter, X_n will be a Markov chain with a finite or countable state space, S. To each state $x \in S$, let

$$H_x := \min\{n \geq 0 : X_n = x\} \text{ and } R_x := \min\{n \geq 1 : X_n = x\} \tag{8.1}$$

be the **first hitting and passage (return) time of the chain to site** x respectively.

8.1 Introduction

We wish to begin with some informal facts and intuition that the reader should keep in mind while proceeding through this chapter.

1. If for some starting distribution (ν),

$$\pi(x) = \lim_{n\to\infty} \mathbb{P}_\nu(X_n = x) = \lim_{n\to\infty} [\nu \mathbf{P}^n]_x \text{ exists} \tag{8.2}$$

   then we expect

$$\pi\mathbf{P} = \lim_{n\to\infty} \nu\mathbf{P}^n\mathbf{P} = \pi$$

   so that π is necessarily an invariant distribution of $\mathbf{P}$. [This is easily justified if $\#(S) < \infty$. For infinite state spaces complications arise if the column sums of $\mathbf{P}$ are infinite.]
2. From your homework we expect that (at least if $\mathbb{E}_x R_x < \infty$ for all $x \in S$)

$$\pi(x) = \frac{1}{\mathbb{E}_x R_x}.$$

3. One further expects ("on average") that

$$\frac{1}{n}\sum_{m=0}^{n} 1_{X_m = x} = \frac{\#\text{ returns to } x}{\#\text{ steps}} \cong \frac{\#\text{ steps}/\mathbb{E}_x R_x}{\#\text{ steps}} = \frac{1}{\mathbb{E}_x R_x}$$

   i.e. that

$$\lim_{N\to\infty} \frac{1}{n}\sum_{m=0}^{n} 1_{X_m = x} = \frac{1}{\mathbb{E}_x R_x} \text{ a.s..} \tag{8.3}$$

4. Notice that if Eq. (8.3) holds then taking expectations (using DCT) implies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} \mathbb{P}_\nu(X_m = x) = \mathbb{E}\left[\lim_{N\to\infty} \frac{1}{n}\sum_{m=0}^{n} 1_{X_m = x}\right] = \frac{1}{\mathbb{E}_x R_x}$$

   which may be restated as saying,

$$\frac{1}{\mathbb{E}_x R_x} = \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} (\nu\mathbf{P}^m)_x$$

   which is a weaker form of Eq. (8.2).

Example 8.1 (A periodic chain). Let $S = \{1, 2, 3\}$ and

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

so that the following transitions occur with probability 1, $1 \to 2 \to 3 \to 1$. In this case ($\mathbf{P}^2$ corresponds to $1 \to 3 \to 2 \to 1$ and $\mathbf{P}^3$ corresponds to $1 \to 1$, $2 \to 2$ and $3 \to 3$), i.e.

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{P}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

With these observations we see that $\mathbf{P}^n$ cycles through the above three matrices and so $\lim_{n\to\infty} \mathbf{P}^n$ does **not** exists. On the other hand it is quite easy to verify;

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} \mathbf{P}^m = \frac{1}{3}\left[\mathbf{P} + \mathbf{P}^2 + \mathbf{P}^3\right] = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

where each row is $\pi = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ which is the unique invariant distribution of $\mathbf{P}$. [The eigenvalues of $\mathbf{P}$ are the three roots of unity in $\mathbb{C}$, namely $\left\{1, -\frac{1}{2} \pm \frac{1}{2}i\sqrt{3}\right\}$.]

Our goal in this chapter is to refine and make precise the ideas in the above remark being wary of the previous example.

8.2 The Main Results

See Kallenberg [10, pages 148-158] for more information along the lines of what is to follow. The first thing we need to do is to divide the state space up into different "classes".

Definition 8.2. *A state y is* ***accessible*** *from x (written $x \to y$) iff $\mathbb{P}_x(H_y < \infty) > 0$ and x and y* ***communicate*** *(written $x \longleftrightarrow y$) iff $x \to y$ and $y \to x$.*

Notice that (since $\mathbb{P}_x(H_x < \infty) = 1$) $x \longleftrightarrow x$ for all $x \in S$. We also have $x \to y$ iff there is a path, $x = x_0 \to x_1 \to \cdots \to x_n = y \in S$ such that $p(x_0, x_1)\, p(x_1, x_2) \ldots p(x_{n-1}, x_n) > 0$. Furthermore if $x \longleftrightarrow y$ and $y \longleftrightarrow k$, then $x \longleftrightarrow k$, that is notion of communicating is an equivalence relation.

Definition 8.3. *For each $x \in S$, let $C_x := \{y \in S : x \longleftrightarrow y\}$ be the* ***communicating class of*** *x.*

The state space, S, is partitioned into a disjoint union of its communicating classes, see Figure 8.1 for an example.

Definition 8.4. *A communicating class $C \subset S$ is* ***closed*** *provided the probability that X_n leaves C given that it started in C is zero. In other words $\mathbf{P}_{xy} = 0$ for all $x \in C$ and $y \notin C$.*

Notice that if $C \subset S$ is a closed communicating class, then X_n restricted to C is a Markov chain.

Definition 8.5. *A state $x \in S$ is:*

1. ***transient*** *if $\mathbb{P}_x(R_x < \infty) < 1$ $(\iff \mathbb{P}_x(R_x = \infty) > 0)$,*
2. ***recurrent*** *if $\mathbb{P}_x(R_x < \infty) = 1$ $(\iff \mathbb{P}_x(R_x = \infty) = 0)$,*
   a) ***positive recurrent*** *if $1/(\mathbb{E}_x R_x) > 0$, i.e. $\mathbb{E}_x R_x < \infty$,*
   b) ***null recurrent*** *if it is recurrent ($\mathbb{P}_x(R_x < \infty) = 1$) and $1/(\mathbb{E}_x R_x) = 0$, i.e. $\mathbb{E} R_x = \infty$.*

We let S_t, S_r, S_{pr}, and S_{nr} be the transient, recurrent, positive recurrent, and null recurrent states respectively.

The rest of this chapter is devoted to the main Markov chain limiting results along with some illustrative examples. The more technical aspects of the proofs will be covered in Chapter 9. Some of the more interesting examples of the theory given here will appear in Chapter 10, see Examples 10.7, 10.8, and 10.9 for examples of transient, null-recurrent, and positively recurrent states.

Proposition 8.6 (Class properties). *The notions of being recurrent, positive recurrent, null recurrent, or transient are all class properties. Namely if $C \subset S$ is a communicating class then either all $x \in C$ are recurrent, positive recurrent, null recurrent, or transient. Hence it makes sense to refer to C as being either recurrent, positive recurrent, null recurrent, or transient.*

Proof. See Proposition 8.19 for the assertion that being recurrent or transient is a class property. For the fact that positive and null recurrence is a class property, see Proposition 9.5 below. ■

Lemma 8.7. *Let $C \subset S$ be a communicating class. Then*

$$C \text{ not closed} \implies C \text{ is transient}$$

or equivalently put,

$$C \text{ is recurrent} \implies C \text{ is closed.}$$

Proof. If C is not closed and $x \in C$, there is a $y \notin C$ such that $x \to y$, i.e. there is a path $x = x_0, x_1, \ldots, x_n = y$ with all of the $\{x_j\}_{j=0}^n$ being distinct such that

$$\mathbb{P}_x(X_0 = x, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n = y) > 0.$$

Since $y \notin C$ we must have $y \nrightarrow C$ and therefore on the event,

$$A := \{X_0 = x, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n = y\},$$

$X_m \notin C$ for all $m \geq n$ and therefore $R_x = \infty$ on the event A which has positive probability. ■

Definition 8.8. *To each site $x \in S$, let*

$$M_x := \sum_{n=0}^{\infty} 1_{X_n = x} = \sum_{n=0}^{\infty} 1_{\{x\}}(X_n) \tag{8.4}$$

be ***number of visits (returns)*** *of $\{X_n\}_{n \geq 0}$ to site x.*

Proposition 8.9. *Suppose that $C \subset S$ is a finite communicating class and $T = \inf\{n \geq 0 : X_n \notin C\} = H_{S \setminus C}$ be the first exit time from C. If C is not closed, then not only is C transient but $\mathbb{E}_x H < \infty$ for all $x \in C$. We also have the equivalence of the following statements:*

1. *C is closed.*
2. *C is positive recurrent.*
3. *C is recurrent.*

*In particular if $\#(S) < \infty$, then the recurrent (= positively recurrent) states are precisely the union of the closed communication classes and the transient states are what is left over. [**Warning**: when $\#(S) = \infty$ or more importantly $\#(C) = \infty$, life is not so simple – see Example 10.6 below.]*

Proof. These results follow fairly easily from Corollary 7.15 and the fact that

$$T = \sum_{x \in C} M_x.$$

■

Remark 8.10. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $\{0, 1, 2, \dots\}$ with 0 being an absorbing state. The communication classes are $\{0\}$ and $\{1, 2, \dots\}$ with the latter class not being closed and hence transient. If $T = H_{\{0\}}$ is the first exit time from $\{1, 2, \dots\}$, Using Remark 6.21, it follows that $\mathbb{E}_x T = \infty$ for all $x > 0$ which shows we can not drop the assumption that $\#(C) < \infty$ in the first statement in Proposition 8.9. Similarly, using the fair random walk example, we see that it is not possible to drop the condition that $\#(C) < \infty$ for the equivalence statements as well.

Example 8.11. Let $\mathbf{P}$ be the Markov matrix with jump diagram given in Figure 8.1. In this case the communication classes are $\{\{1, 2\}, \{3, 4\}, \{5\}\}$. The latter two are closed and hence positively recurrent while $\{1, 2\}$ is transient.

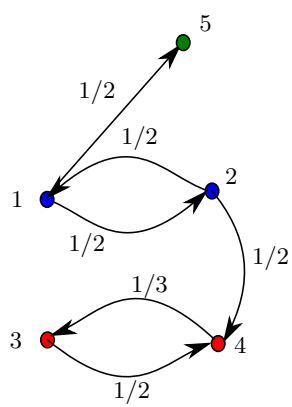


Fig. 8.1. A 5 state Markov chain with 3 communicating classes.

Example 8.12. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $S = \mathbb{Z}$, then this chain is irreducible. On the other hand if $\{X_n\}_{n=0}^{\infty}$ is the fair random walk on $\{0, 1, 2, \dots\}$ with 0 being an absorbing state, then the communication classes are $\{0\}$ (closed) and $\{1, 2, \dots\}$ (not closed).

Remark 8.13 (Warnings). There are some subtle point which may arise when $\#(S) = \infty$.

1. If $C \subset S$ is closed and $\#(C) = \infty$, C could be recurrent or it could be transient. Transient in this case means the walk goes off to "infinity."
2. Interchanging sums and limits is not always permissible. For example, consider the fair random walk on $\mathbb{Z}$ which satisfies $\mathbb{E}_x R_x = \infty$ for all $x \in \mathbb{Z}$. In this case, for any starting distribution, ν, we will see below that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = \frac{1}{\mathbb{E}_x R_x} = 0 \ \mathbb{P}_\nu \text{ - a.s.}$$

while

$$1 = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1 = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x\in\mathbb{Z}} 1_{X_m = x} \neq \sum_{x\in\mathbb{Z}} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = 0.$$

More generally, in the terminology below if C is a closed communication class which is positively recurrent or transient, then for any starting distribution ν with $\nu(C) = 1$ we will have, $\mathbb{P}_\nu$ – a.s., that

$$1 = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1 = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x\in C} 1_{X_m = x} \neq \sum_{x\in C} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = 0.$$

The following proposition is a consequence of the strong Markov property in Theorem 7.11.

Remark 8.14 (MC Segment Decomposition). The following excursion decomposition is the key idea behind most of the results in this chapter. Let $\{X_n\}_{n=0}^{\infty}$ be our Markov chain starting with some initial distribution, $\nu : S \to [0, 1]$ and let $x \in S$ be given. We then decompose the chain as follows. Let T_N be the time of the Nth – visit to site x so that $T_1 = H_x$, $T_2 = \min\{n > H_x : X_n = x\}, \dots,$ $T_{N+1} = \min\{n > T_N : X_n = x\}$ with the convention that minimum of the empty set is ∞. The chain then decomposes into "**segments**" (or **excursions**),

$$\begin{aligned} S_1 &:= (X_0, \dots, X_{T_1} = x), \\ S_2 &:= (X_{T_1} = x, X_{T_1+1} \dots, X_{T_2} = x), \\ S_3 &:= (X_{T_2} = x, X_{T_2+1} \dots, X_{T_3} = x), \\ &\vdots \end{aligned}$$

It is possible there is only one segment but what we do know is, by the strong Markov property, that given $T_N < \infty$, the segments $S_1, \dots, S_N$ are independent and $S_2, \dots, S_N$ have the same distributions. In particular we conclude, relative

to $\mathbb{P}_\nu(\cdot|T_N<\infty)$, that $\{\tau_K := T_K - T_{K-1}\}_{K=2}^N$ are i.i.d. random variables such that

$$\mathbb{P}_\nu(\tau_K = n|T_N<\infty) = \mathbb{P}_x(R_x = n) \text{ for } n\in\mathbb{N} \text{ and } 2\le K\le N$$

and in particular

$$\mathbb{E}_\nu[\tau_K|T_N<\infty] = \mathbb{E}_x R_x \text{ for } 2\le K\le N.$$

Proposition 8.15. *If $x\in S$, $N\in\mathbb{N}$, and $\nu: S\to[0,1]$ is any probability on S, then*

$$\mathbb{P}_\nu(M_x\ge N) = \mathbb{P}_\nu(H_x<\infty)\cdot\mathbb{P}_x(R_x<\infty)^{N-1}. \tag{8.5}$$

[Also see Proposition 9.1 for another proof.]

Proof. Following the ideas in Remark 8.14: intuitively $M_x\ge N$ happens iff the chain first visits x with probability $\mathbb{P}_\nu(H_x<\infty)$ and then revisits x again $N-1$ times which the probability of each revisit being $\mathbb{P}_x(R_x<\infty)$. Since Markov chains are forgetful, these probabilities are all independent and hence we arrive at Eq. (8.5). See Proposition 9.1 below for the formal proof based on the strong Markov property in Theorem 7.11.

Formal proof. Using the strong Markov property as stated in Theorem 7.8 we know that

$$\mathbb{X} \overset{d}{=} (X_0,\dots,X_{\tau_N},Y_1^x,Y_2^x,\dots) \text{ on } \{\tau_N<\infty\}$$

and therefore

$$\begin{aligned}\mathbb{P}(M_x\ge N+1) &= \mathbb{P}(\tau_{N+1}<\infty) = \mathbb{P}(\tau_N<\infty,\tau_{N+1}-\tau_N<\infty)\\ &= \mathbb{P}(\tau_N<\infty, R_x(X_{\tau_N},X_{\tau_N+1},\dots)<\infty)\\ &= \mathbb{P}(\tau_N<\infty, R_x(Y_0^x,Y_1^x,\dots)<\infty)\\ &= \mathbb{P}(\tau_N<\infty)\cdot\mathbb{P}(R_x(Y_0^x,Y_1^x,\dots)<\infty)\\ &= \mathbb{P}(\tau_N<\infty)\cdot\mathbb{P}_x(R_x<\infty) = \mathbb{P}_x(R_x<\infty)\cdot\mathbb{P}(M_x\ge N)\end{aligned}$$

and so Equation (8.5) now follows by induction on N. ■

Corollary 8.16. *If $x\in S$ and $\nu: S\to[0,1]$ is any probability on S, then*[1]

$$\mathbb{P}_\nu(M_x=\infty) = \mathbb{P}_\nu(X_n = x \text{ i.o.}) = \mathbb{P}_\nu(H_x<\infty)1_{x\in S_r}, \tag{8.6}$$

$$\mathbb{P}_x(M_x=\infty) = \mathbb{P}_x(X_n = x \text{ i.o.}) = 1_{x\in S_r}, \tag{8.7}$$

[1] See Definition 2.4 to review the meaning of $\{X_n = y \text{ i.o.}\}$. In this case, $\{X_n = y \text{ i.o.}\} = \{M_y=\infty\}$.

$$\mathbb{E}_\nu M_x = \sum_{n=0}^\infty\sum_{y\in S}\nu(y)\mathbf{P}_{yx}^n = \frac{\mathbb{P}_\nu(H_x<\infty)}{1-\mathbb{P}_x(R_x<\infty)}, \tag{8.8}$$

and (taking $\nu=\delta_y$)

$$\mathbb{E}_y M_x = \sum_{n=0}^\infty \mathbf{P}_{yx}^n = \frac{\mathbb{P}_y(H_x<\infty)}{1-\mathbb{P}_x(R_x<\infty)} \tag{8.9}$$

where the following conventions are used in interpreting the right hand side of Eqs. (8.8) and (8.9); $a/0:=\infty$ if $a>0$ while $0/0:=0$.

Proof. Since

$$\{M_x\ge N\}\downarrow\{M_x=\infty\} = \{X_n = x \text{ i.o. } n\} \text{ as } N\uparrow\infty,$$

it follows, using Eq. (8.5), that

$$\mathbb{P}_\nu(X_n = x \text{ i.o. } n) = \lim_{N\to\infty}\mathbb{P}_\nu(M_x\ge N) = \mathbb{P}_\nu(H_x<\infty)\cdot\lim_{N\to\infty}\mathbb{P}_x(R_x<\infty)^{N-1} \tag{8.10}$$

which gives Eq. (8.6). Equation (8.7) follows by taking $\nu=\delta_x$ in Eq. (8.6) and recalling that $x\in S_r$ iff $\mathbb{P}_x(R_x<\infty)=1$. Similarly Eq. (8.9) is a special case of Eq. (8.8) with $\nu=\delta_x$. We now prove Eq. (8.8).

Using the definition of M_x in Eq. (8.4),

$$\begin{aligned}\mathbb{E}_\nu M_x &= \mathbb{E}_\nu\sum_{n=0}^\infty 1_{X_n=x} = \sum_{n=0}^\infty \mathbb{E}_\nu 1_{X_n=x}\\ &= \sum_{n=0}^\infty \mathbb{P}_\nu(X_n=x) = \sum_{n=0}^\infty\sum_{y\in S}\nu(y)\mathbf{P}_{yx}^n\end{aligned}$$

which is the first equality in Eq. (8.8). For the second equality we combine Lemma 5.18 along with Eq. (8.5) to find;

$$\begin{aligned}\mathbb{E}_\nu M_x &= \sum_{N=1}^\infty \mathbb{P}_\nu(M_x\ge N)\\ &= \sum_{N=1}^\infty \mathbb{P}_\nu(H_x<\infty)\mathbb{P}_x(R_x<\infty)^{N-1}\\ &= \frac{\mathbb{P}_\nu(H_x<\infty)}{1-\mathbb{P}_x(R_x<\infty)}.\end{aligned}$$

■

Corollary 8.17. *If $\nu : S \to [0,1]$ is any starting distribution and $x \in S_t$, then*

$$\mathbb{P}_\nu (M_x < \infty) = 1 \text{ and } \lim_{n\to\infty} \mathbb{P}_\nu (X_n = x) = 0.$$

The first equality above states that, $\mathbb{P}_\nu$ – a.s., X_n visits x at most a finite number of times which further implies;

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = 0 \quad \mathbb{P}_\nu - a.s.$$

Proof. When $x \in S_t$, Eq. (8.8) implies $\mathbb{E}_\nu M_x < \infty$ which immediately implies $\mathbb{P}_\nu (M_x < \infty) = 1$. For the second assertion we simply observe that

$$\infty > \mathbb{E}_\nu M_x = \sum_{n=0}^{\infty} \mathbb{E}_\nu 1_{X_n = x} = \sum_{n=0}^{\infty} \mathbb{P}_\nu (X_n = x)$$

and therefore $\lim_{n\to\infty} \mathbb{P}_\nu (X_n = x) = 0$. ■

Theorem 8.18 (Recurrent States). *Let $y \in S$. Then the following are equivalent;*

1. *y is recurrent, i.e. $\mathbb{P}_y (R_y < \infty) = 1$,*
2. *$\mathbb{P}_y (X_n = y \text{ i.o. } n) = 1$,*
3. *$\mathbb{E}_y M_y = \sum_{n=0}^{\infty} \mathbf{P}^n_{yy} = \infty$.*

Proof. The equivalence of the first two items follows directly from Eq. (8.7) or directly from Remark 8.14 when $\nu = \delta_y$. The equivalent of items 1. and 3. follows directly from Eq. (8.9) with $x = y$. ■

Proposition 8.19. *If $x \longleftrightarrow y$, then x is recurrent iff y is recurrent, i.e. the property of being recurrent or transient is a class property.*

Proof. First proof. For any $\alpha, \beta \in n \in \mathbb{N}_0$ we have,

$$\mathbf{P}^{n+\alpha+\beta}_{xx} = \sum_{u,v\in S} \mathbf{P}^{\alpha}_{xu} \mathbf{P}^{n}_{uv} \mathbf{P}^{\beta}_{vx} \geq \mathbf{P}^{\alpha}_{xy} \mathbf{P}^{n}_{y\,y} \mathbf{P}^{\beta}_{yx}$$

and therefore,

$$\mathbb{E}_x M_x \geq \sum_{n=0}^{\infty} \mathbf{P}^{n+\alpha+\beta}_{xx} \geq \mathbf{P}^{\alpha}_{xy} \mathbf{P}^{\beta}_{yx} \sum_{n=0}^{\infty} \mathbf{P}^{n}_{y\,y} = \left(\mathbf{P}^{\alpha}_{xy} \mathbf{P}^{\beta}_{yx}\right) \mathbb{E}_y M_y.$$

Since $x \longleftrightarrow y$ we may choose $\alpha, \beta \in \mathbb{N}$ so that $c := \mathbf{P}^{\alpha}_{xy} \mathbf{P}^{\beta}_{yx} > 0$ and so the above inequality shows if $\mathbb{E}_y M_y = \infty$ then $\mathbb{E}_x M_x = \infty$ as well. By interchanging the roles of x and y we may similarly may show $\mathbb{E}_x M_x = \infty$ implies $\mathbb{E}_y M_y = \infty$. Thus using item 3. of Theorem 8.18, it follows that x is recurrent iff y is recurrent.

Second proof. See the aside in the proof of Corollary 8.20 below. ■

Corollary 8.20. *If $C \subset S_r$ is a recurrent communication class, then*

$$\mathbb{P}_x(R_y < \infty) = 1 = \mathbb{P}_x (M_y = \infty) \text{ for all } x, y \in C \tag{8.11}$$

and in fact

$$\mathbb{P}_x(\cap_{y\in C}\{X_n = y \text{ i.o. } n\}) = \mathbb{P}_x(\cap_{y\in C}\{M_y = \infty\}) = 1 \text{ for all } x \in C. \tag{8.12}$$

More generally if $\nu : S \to [0,1]$ is a probability such that $\nu(x) = 0$ for $x \notin C$, then

$$\mathbb{P}_\nu(\cap_{y\in C}\{X_n = y \text{ i.o. } n\}) = 1 \text{ for all } x \in C. \tag{8.13}$$

In words, if we start in C then every state in C is visited an infinite number of times. (Notice that $\mathbb{P}_x (R_y < \infty) = \mathbb{P}_x(\{X_n\}_{n\geq 1} \text{ hits } y)$.)

Proof. First Proof. Let C be a communicating class and suppose $x, y \in C$ with x being recurrent. (After the proof we will show directly that y is recurrent as well, thus giving another proof of Proposition 8.19.) Since x is recurrent, $\mathbb{P}_x$ – a.s. their are infinitely many independent round trip excursions from x to x in the segment decomposition described in Remark 8.14 with $\nu = \delta_x$. For $n \in \mathbb{N}$, let A_n be the event that y is visited by the chain during the nth - segment which is round trip excursion from x to x. By the strong Markov property, $\{A_n\}_{n=1}^{\infty}$ is an independent collection of sets with $p = \mathbb{P}_x (A_n)$ being independent of n. Moreover since $x \to y$, it follows that $p > 0$. Therefore by an application of the second Borel Cantelli Lemma 2.6, $\mathbb{P}_x (A_n \text{ i.o.}) = 1$, i.e. $\mathbb{P}_x (M_y = \infty) = 1$ which proves Eq. (8.11). If we now assume Proposition 8.19 we may conclude that $\mathbb{P}_x (M_y = \infty) = 1$ for all $x, y \in C$. Equation (8.12) is a consequence of Eq. (8.11) and the fact that the countable intersection of probability one events is again a probability 1 event. Equation (8.13) follows by multiplying Eq. (8.12) by $\nu(x)$ and then summing on $x \in C$.

Aside: a second proof of Proposition 8.19. Let $x \in S$ be recurrent and $y \in S$ such that $y \neq x$ and y communicates with x, i.e. $y \longleftrightarrow x$. Recall that we have just shown Eq. (8.12) holds (i.e. $\mathbb{P}_x (M_y = \infty) = 1 = \mathbb{P}_x (H_y < \infty)$) under these assumptions. Using the identity,

$$\begin{aligned} M_y (X_0, X_1, \dots) &= \sum_{n=0}^{\infty} 1_{X_n = y} = \sum_{n=0}^{\infty} 1_{X_{H_y+n} = y} \\ &= M_y \left(X_{H_y}, X_{H_y+1}, X_{H_y+2}, \dots\right), \end{aligned}$$

along with the strong Markov property implies;

$$\begin{aligned} 1 &= \mathbb{P}_x (M_y (X_0, X_1, \dots) = \infty) = \mathbb{P}_x (M_y (X_0, X_1, \dots) = \infty, H_y < \infty) \\ &= \mathbb{P}_x (H_y < \infty) \cdot \mathbb{P}_y (M_y (X_0, X_1, \dots) = \infty) \\ &= \mathbb{P}_y (M_y = \infty). \end{aligned}$$

This certainly shows y is recurrent as well.

Second proof of Corollary 8.20. Let $x, y \in C \subset S_r$ and choose $m \in \mathbb{N}$ such that $\mathbb{P}_y(X_m = x) > 0$. Then using $\mathbb{P}_y(M_y = \infty) = 1$ we learn that

$$\begin{aligned}\mathbb{P}_y(X_m = x) &= \mathbb{P}_y(M_y = \infty, X_m = x) = \mathbb{P}_y(M_y = \infty | X_m = x)\,\mathbb{P}_y(X_m = x) \\ &= \mathbb{P}_x(M_y = \infty)\,\mathbb{P}_y(X_m = x)\end{aligned}$$

which implies,

$$1 = \mathbb{P}_x(M_y = \infty) \le \mathbb{P}_x(R_y < \infty) \implies \text{Eq. (8.11)}.$$

Here we have used

$$\{M_y = \infty\} = \left\{\sum_{k=m}^{\infty} 1_{\{y\}}(X_k) = \infty\right\}$$

along with the Markov property (see Theorem 7.5) to assert that

$$\begin{aligned}\mathbb{P}_y(M_y = \infty | X_m = x) &= \mathbb{P}_y\left(\sum_{k=m}^{\infty} 1_{\{y\}}(X_k) = \infty | X_m = x\right) \\ &= \mathbb{P}_x\left(\sum_{k=0}^{\infty} 1_{\{y\}}(X_k) = \infty\right) = \mathbb{P}_x(M_y = \infty).\end{aligned}$$

Equation (8.12) is a consequence of Eq. (8.11) and the fact that the countable intersection of probability one events is again a probability 1 event. Equation (8.13) follows by multiplying Eq. (8.12) by $\nu(x)$ and then summing on $x \in C$. ■

Theorem 8.21 (Transient States). *Let $y \in S$. Then the following are equivalent;*

1. *y is transient, i.e. $\mathbb{P}_y(R_y < \infty) < 1$,*
2. *$\mathbb{P}_y(X_n = y \text{ i.o. } n) = 0$, and*
3. *$\mathbb{E}_y M_y = \sum_{n=0}^{\infty} \mathbf{P}^n_{y,y} < \infty$.*

Moreover, if $x \in S$ and $y \in S_t$, then

$$\sum_{n=0}^{\infty} \mathbf{P}^n_{xy} = \mathbb{E}_x M_y < \infty \tag{8.14}$$

$$\implies \begin{cases} \mathbb{P}_x(X_n = y \text{ i.o. } n) = 0 \text{ and} \\ \lim_{n\to\infty} \mathbb{P}_x(X_n = y) = \lim_{n\to\infty} \mathbf{P}^n_{xy} = 0. \end{cases} \tag{8.15}$$

and more generally if $\nu : S \to [0,1]$ is any probability, then

$$\sum_{n=0}^{\infty} \mathbb{P}_\nu(X_n = y) = \mathbb{E}_\nu M_y < \infty \tag{8.16}$$

$$\implies \begin{cases} \mathbb{P}_\nu(X_n = y \text{ i.o. } n) = 0 \text{ and} \\ \lim_{n\to\infty} \mathbb{P}_\nu(X_n = y) = \lim_{n\to\infty} [\nu \mathbf{P}^n]_y = 0. \end{cases} \tag{8.17}$$

Proof. The equivalence of the first two items follows directly from Eq. (8.7) and the equivalent of items 1. and 3. follows directly from Eq. (8.9) with $x = y$. The fact that $\mathbb{E}_x M_y < \infty$ and $\mathbb{E}_\nu M_y < \infty$ in Eqs. (8.14) and (8.16) for all $y \in S_t$ are consequences of Eqs. (8.9) and (8.8) respectively. The remaining implications in Eqs. (8.15) and (the more general) Eq. (8.17) are consequences of the facts; 1) the nth – term in a convergent series tends to zero as $n \to \infty$, 2) the $\{X_n = y \text{ i.o. } n\} = \{M_y = \infty\}$, and 3) $\mathbb{E}_\nu M_\nu < \infty$ implies $\mathbb{P}_\nu(M_y = \infty) < \infty$. ■

Corollary 8.22. *1) If the state space, S, is a finite set, then $S_r \neq \emptyset$. 2) Any finite and closed communicating class $C \subset S$ is a recurrent.*

Proof. First suppose that $\#(S) < \infty$ and for the sake of contradiction, suppose $S_r = \emptyset$ or equivalently that $S = S_t$. Then by Theorem 8.21, $\lim_{n\to\infty} \mathbf{P}^n_{xy} = 0$ for all $x, y \in S$. On the other hand, $\sum_{y\in S} \mathbf{P}^n_{xy} = 1$ so that

$$1 = \lim_{n\to\infty} \sum_{y\in S} \mathbf{P}^n_{xy} = \sum_{y\in S} \lim_{n\to\infty} \mathbf{P}^n_{xy} = \sum_{y\in S} 0 = 0,$$

which is a contradiction. (Notice that if S were infinite, we could not (in general) interchange the limit and the above sum without some extra conditions.)

To prove the first statement, restrict X_n to C to get a Markov chain on a finite state space C. By what we have just proved, there is a recurrent state $x \in C$. Since recurrence is a class property, it follows that all states in C are recurrent. ■

Definition 8.23. *A function, $\pi : S \to [0,1]$ is a **sub-probability** if $\sum_{y\in S} \pi(y) \le 1$. We call $\sum_{y\in S} \pi(y)$ the **mass** of π. So a probability is a sub-probability with mass one.*

Definition 8.24. *We say a sub-probability, $\pi : S \to [0,1]$, is **invariant** if $\pi\mathbf{P} = \pi$, i.e.*

$$\sum_{x\in S} \pi(x)\,\mathbf{P}_{xy} = \pi(y) \text{ for all } y \in S. \tag{8.18}$$

*An invariant probability, $\pi : S \to [0,1]$, is called an **invariant distribution.***

Example 8.25. If $\#(S) < \infty$ and $p : S \times S \to [0,1]$ is a Markov transition matrix with column sums adding up to 1 then $\pi(x) := \frac{1}{\#(S)}$ is an invariant distribution for p. In particular, if p is a symmetric Markov transition matrix ($p(x,y) = p(y,x)$ for all $x, y \in S$), then the uniform distribution π is an invariant distribution for p.

Example 8.26. If S is a finite set of nodes and G be an undirected graph on S, i.e. G is a subset of $S \times S$ such that

1. $(x,x) \notin G$ for all $x \in S$,
2. if $(x,y) \in G$, then $(y,x) \in G$ [the graph is undirected], and
3. for all $x \in G$, the set $S_x := \{y \in S : (x,y) \in G\}$ is not empty. [We are not allowing for any isolated nodes in our graph.]

Let

$$\nu(x) := \#(S_x) = \sum_{y \in S} 1_{(x,y) \in G}$$

be the **valence** of G at x. The **random walk on this graph** is then the Markov chain on S with Markov transition matrix,

$$p(x,y) := \frac{1}{\nu(x)} 1_{S_x}(y) = \frac{1}{\nu(x)} 1_{(x,y) \in G}.$$

Notice that

$$\sum_{x \in S} \nu(x) p(x,y) = \sum_{x \in S} 1_{(x,y) \in G} = \sum_{x \in S} 1_{(y,x) \in G} = \nu(y).$$

Thus if we let $Z := \sum_{x \in S} \nu(x)$ and $\pi(x) := \nu(x)/Z$, we will have π is an invariant distribution for $\mathbf{P}$.

Theorem 8.27 (Ergodic Theorem). *Suppose that $\mathbf{P} = (p_{xy})$ is an **irreducible** Markov kernel and*

$$\pi_y := \frac{1}{\mathbb{E}_y R_y} \text{ for all } y \in S. \tag{8.19}$$

Then:

1. *For all $x, y \in S$, we have*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n} 1_{X_m = y} = \pi_y \quad \mathbb{P}_x - a.s. \tag{8.20}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_x(X_m = y) = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \mathbf{P}^m_{xy} = \pi_y. \tag{8.21}$$

2. *If $\mu : S \to [0,1]$ is an invariant sub-probability, then either $\mu(x) > 0$ for all x or $\mu(x) = 0$ for all x.*
3. *$\mathbf{P}$ has at most one invariant distribution.*
4. *$\mathbf{P}$ has a (necessarily unique) invariant distribution, $\mu : S \to [0,1]$, iff $\mathbf{P}$ is positive recurrent in which case $\mu(x) = \pi(x) = \frac{1}{\mathbb{E}_x R_x} > 0$ for all $x \in S$.*

(These results may of course be applied to the restriction of a general non-irreducible Markov chain to any one of its communication classes.)

Proof. These results are the contents of Theorem 9.3 and Propositions 9.5 and 9.7 below.

■

Using this result we can give another proof of Proposition 8.9.

Theorem 8.28 (General Convergence Theorem). *Let $\nu : S \to [0,1]$ be any probability, $y \in S$, $C = C_y$ be the communicating class*[2] *containing y, $H_C = \inf\{n : X_n \in C\}$ with $H_C = \infty$ when $X_n \notin C$ for all n,*[3] *and*

$$\pi_y := \pi_y(\nu) = \frac{\mathbb{P}_\nu(H_C < \infty)}{\mathbb{E}_y R_y}, \tag{8.22}$$

where $1/\infty := 0$. Then:

1. *$\mathbb{P}_\nu$ – a.s.,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = \frac{1}{\mathbb{E}_y R_y} \cdot 1_{\{H_C < \infty\}}, \tag{8.23}$$

2.

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x \in S} \nu(x) \mathbf{P}^m_{xy} = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\nu(X_m = y) = \pi_y, \tag{8.24}$$

3. *π is an invariant sub-probability for P, and*
4. *the mass of $\pi(S) := \sum_{y \in S} \pi_y$ is given by*

$$\pi(S) = \sum_{C:\ pos.\ recurrent} \mathbb{P}_\nu(H_C < \infty) \le 1. \tag{8.25}$$

Proof. If $y \in S$ is a transient site, then according to Eq. (8.17), $\mathbb{P}_\nu(M_y < \infty) = 1$ and therefore $\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} 1_{X_m = y} = 0$ which agrees with Eq. (8.23) for $y \in S_t$.

So now suppose that $y \in S_r$. Let C be the communication class containing y and

$$H = H_C := \inf\{n \ge 0 : X_n \in C\}$$

be the first time when X_n enters C. It is clear that $\{R_y < \infty\} \subset \{H < \infty\}$. On the other hand, for any $z \in C$, it follows by the strong Markov property of Theorem 7.11 and Corollary 8.20 that, conditioned on $\{H < \infty, X_H = z\}$, $\{X_n\}$ hits y i.o. and hence $\mathbb{P}_\nu(R_y < \infty | H < \infty, X_H = z) = 1$. Equivalently put,

[2] No assumptions on the type of this class are needed here.
[3] We sometimes write the event $\{H_C < \infty\}$ as $\{X_n \text{ hits } C\}$.

$$\mathbb{P}_\nu\left(R_y<\infty, H<\infty, X_H=z\right)=\mathbb{P}_\nu\left(H<\infty, X_H=z\right) \text{ for all } z\in C.$$

Summing this last equation on $z\in C$ then shows

$$\mathbb{P}_\nu\left(R_y<\infty\right)=\mathbb{P}_\nu\left(R_y<\infty, H<\infty\right)=\mathbb{P}_\nu\left(H<\infty\right)$$

and therefore $\{R_y<\infty\}=\{H<\infty\}$ modulo an event with $\mathbb{P}_\nu$ – probability zero.

Another application of the strong Markov property of Theorem 7.11, observing that $X_{R_y}=y$ on $\{R_y<\infty\}$, allows us to conclude that the $\mathbb{P}_\nu\left(\cdot|R_y<\infty\right)=\mathbb{P}_\nu\left(\cdot|H<\infty\right)$ – law of $\left(X_{R_y}, X_{R_y+1}, X_{R_y+2}, \dots\right)$ is the same as the $\mathbb{P}_y$ – law of $\left(X_0, X_1, X_2, \dots\right)$. Therefore, we may apply Theorem 8.27 to conclude that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n 1_{X_m=y}=\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n 1_{X_{R_y+m}=y}=\frac{1}{\mathbb{E}_y R_y}\,\mathbb{P}_\nu\left(\cdot|R_y<\infty\right) \text{ – a.s.}$$

On the other hand, on the event $\{R_y=\infty\}$ we have $\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n 1_{X_m=y}=0$. Thus we have shown $\mathbb{P}_\nu$ – a.s. that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n 1_{X_m=y}=\frac{1}{\mathbb{E}_y R_y}1_{R_y<\infty}=\frac{1}{\mathbb{E}_y R_y}1_{H<\infty}=\frac{1}{\mathbb{E}_y R_y}1_{\{H_C<\infty\}}$$

which is Eq. (8.23). Taking expectations of this equation, using the dominated convergence theorem, gives Eq. (8.24).

Since $\mathbb{E}_x R_x=\infty$ unless x is a positive recurrent site, it follows that

$$\sum_{x\in S}\pi_x\mathbf{P}_{xy}=\sum_{x\in S_{\rm pr}}\pi_x\mathbf{P}_{xy}=\sum_{C:\text{ pos-rec.}}\mathbb{P}_\nu\left(H_C<\infty\right)\sum_{x\in C}\frac{1}{\mathbb{E}_x R_x}\mathbf{P}_{xy}. \tag{8.26}$$

As each positive recurrent class, C, is closed; if $x\in C$ and $y\notin C$, then $\mathbf{P}_{xy}=0$. Therefore $\sum_{x\in C}\frac{1}{\mathbb{E}_x R_x}\mathbf{P}_{xy}$ is zero unless $y\in C$. So if $y\notin S_{\rm pr}$ we have $\sum_{x\in S}\pi_x\mathbf{P}_{xy}=0=\pi_y$ and if $y\in S_{\rm pr}$, then by Theorem 8.27,

$$\sum_{x\in C}\frac{1}{\mathbb{E}_x R_x}\mathbf{P}_{xy}=1_{y\in C}\cdot\frac{1}{\mathbb{E}_y R_y}.$$

Using this result in Eq. (8.26) shows that

$$\sum_{x\in S}\pi_x\mathbf{P}_{xy}=\sum_{C:\text{ pos-rec.}}\mathbb{P}_\nu\left(H_C<\infty\right)1_{y\in C}\cdot\frac{1}{\mathbb{E}_y R_y}=\pi_y$$

so that π is an invariant distribution. Similarly, using Theorem 8.27 again,

$$\sum_{x\in S}\pi_x=\sum_{C:\text{ pos-rec.}}\mathbb{P}_\nu\left(H_C<\infty\right)\cdot\left(\sum_{x\in C}\frac{1}{\mathbb{E}_x R_x}\right)=\sum_{C:\text{ pos-rec.}}\mathbb{P}_\nu\left(H_C<\infty\right).$$

The point is that for any communicating class C of S we have,

$$\sum_{x\in C}\frac{1}{\mathbb{E}_x R_x}=\begin{cases}1 \text{ if } C \text{ is positive recurrent}\\ 0 \qquad\quad\text{ otherwise.}\end{cases}$$

■

Corollary 8.29. *If C is a closed finite communicating class then C is positive recurrent. (Recall that we already know that C is recurrent by Corollary 8.22.)*

Proof. For $x,y\in C$, let

$$\pi_y:=\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n\mathbb{P}_x\left(X_m=y\right)=\frac{1}{\mathbb{E}_y R_y} \tag{8.27}$$

where the second equality is a consequence of Theorem 8.28. Since C is closed,

$$\sum_{y\in C}\mathbb{P}_x\left(X_n=y\right)=1$$

and therefore,

$$\sum_{y\in C}\pi_y=\lim_{n\to\infty}\frac{1}{n}\sum_{y\in C}\sum_{m=1}^n\mathbb{P}_x\left(X_m=y\right)=\lim_{m\to\infty}\frac{1}{n}\sum_{m=1}^n\sum_{y\in C}\mathbb{P}_x\left(X_m=y\right)=1.$$

In particular it follows we must have $\pi_y>0$ for some $y\in C$ and hence all $y\in C$ by Theorem 8.27 with S replaced by C. So by Eq. (8.27) we conclude $\mathbb{E}_y R_y<\infty$ for all $y\in C$, i.e. every $y\in C$ is a positive recurrent state. ■

Remark 8.30 (The MCMC Seed). Suppose that S is a finite set, $\mathbf{P}=(p_{xy})$ is an **irreducible** Markov kernel, ν is a probability on S, and $f:S\to\mathbb{C}$ is a function. Then

$$\mathbb{E}_\pi f:=\sum_{x\in S}\pi\left(x\right)f\left(x\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n f\left(X_m\right)\quad\mathbb{P}_\nu \text{ – a.s.}$$

where π is the invariant distribution for $\mathbf{P}$. Indeed,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n f\left(X_m\right)&=\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n\sum_{x\in S}f\left(x\right)1_{X_m=x}\\&=\sum_{x\in S}f\left(x\right)\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n 1_{X_m=y}\\&=\sum_{x\in S}f\left(x\right)\pi\left(x\right)\quad\mathbb{P}_\nu \text{ – a.s.}\end{aligned}$$

8.3 Aperiodic chains

Definition 8.31. *A state $x \in S$ is* ***aperiodic*** *if $\mathbf{P}^n_{xx} > 0$ for all n sufficiently large.*

Lemma 8.32. *If $x \in S$ is aperiodic and $y \longleftrightarrow x$, then y is aperiodic. So being aperiodic is a class property.*

Proof. We have

$$\mathbf{P}^{n+m+k}_{yy} = \sum_{w,z\in S} \mathbf{P}^n_{y,w}\mathbf{P}^m_{w,z}\mathbf{P}^k_{z,y} \geq \mathbf{P}^n_{y,x}\mathbf{P}^m_{x,x}\mathbf{P}^k_{x,y}.$$

Since $y \longleftrightarrow x$, there exists $n, k \in \mathbb{N}$ such that $\mathbf{P}^n_{y,x} > 0$ and $\mathbf{P}^k_{x,y} > 0$. Since $\mathbf{P}^m_{x,x} > 0$ for all large m, it follows that $\mathbf{P}^{n+m+k}_{yy} > 0$ for all large m and therefore, y is aperiodic as well. ■

Lemma 8.33. *A state $x \in S$ is aperiodic iff 1 is the greatest common divisor of the set,*

$$\{n \in \mathbb{N} : \mathbb{P}_x(X_n = x) = \mathbf{P}^n_{xx} > 0\}.$$

Proof. Use the number theory Lemma 8.44 below. ■

Theorem 8.34. *If $\mathbf{P}$ is an irreducible, aperiodic, and recurrent Markov chain, then*

$$\lim_{n\to\infty} \mathbf{P}^n_{xy} = \pi_y = \frac{1}{\mathbb{E}_y R_y}. \tag{8.28}$$

More generally, if C is a communication class which is assumed to be aperiodic if it is recurrent, then

$$\lim_{n\to\infty} \mathbb{P}_\nu(X_n = y) := \lim_{n\to\infty} \sum_{x\in S} \nu(x)\mathbf{P}^n_{xy} = \mathbb{P}_\nu(H_C < \infty) \cdot \frac{1}{\mathbb{E}_y R_y} \text{ for all } y \in C. \tag{8.29}$$

Proof. We will only give the proof under the added assumption that $\mathbf{P}$ is positive recurrent or equivalently $\mathbf{P}$ has an invariant distribution.[4] The proof uses the important idea of a **coupling argument** (we follow [14, Theorem 1.8.3] or Kallenberg [10, Chapter 8]). Here is the idea. Let $\{X_n\}$ and $\{Y_n\}$ be two independent Markov chains having $\mathbf{P}$ as their transition matrix. Then the chain $(X_n, Y_n) \in S \times S$ is again a Markov chain with transition matrix $\mathbf{P} \otimes \mathbf{P}$.[5] The aperiodicity and irreducibility assumption guarantees that $\mathbf{P} \otimes \mathbf{P}$ is still irreducible and aperiodic (but the aperiodicity is not needed).[6] Moreover if π is an invariant distribution for $\mathbf{P}$ then[7] $\pi \otimes \pi$ is an invariant distribution for $\mathbf{P} \otimes \mathbf{P}$ and so $\mathbf{P} \otimes \mathbf{P}$ is positive recurrent by Theorem 8.27. We now we have a couple of choices how to proceed. We might take H to be the hitting time of the chain (X_n, Y_n) of a fixed point $(x, x) \in \Delta \subset S \times S$ or we can let H be the first hitting time of the diagonal itself. For either choice $H < \infty$ a.s. no matter the starting distributions because the chain is positive recurrent. We then let

$$\tilde{Y}_n = \begin{cases} Y_n & \text{if } n \leq H \\ X_n & \text{if } n > H. \end{cases}$$

By the strong Markov property $\tilde{Y}$ and Y have the same distribution. Thus if $f : S \to \mathbb{R}$ is a bounded function and μ, ν are two initial distributions on S we will have,

$$\begin{aligned} \mathbb{E}_\mu f(X_n) - \mathbb{E}_\nu f(X_n) &= \mathbb{E}_\mu f(X_n) - \mathbb{E}_\nu f(Y_n) = \mathbb{E}_{\mu\otimes\nu}\left[f(X_n) - f\left(\tilde{Y}_n\right)\right] \\ &= \mathbb{E}_{\mu\otimes\nu}\left[f(X_n) - f\left(\tilde{Y}_n\right) : n < H\right] \end{aligned}$$

from which it follows that

$$|\mathbb{E}_\mu f(X_n) - \mathbb{E}_\nu f(X_n)| \leq 2\|f\|_\infty \mathbb{P}(H > n) \to 0 \text{ as } n \to \infty, \tag{8.30}$$

where $\|f\|_\infty = \sup_{x\in S}|f(x)|$. This inequality shows that the initial distribution plays no role in the limiting distribution of the $\{X_n\}_{n=0}^\infty$.

Taking $\mu = \pi$ to be the invariant distribution and using $\mathbb{E}_\pi f(X_n) = \mathbb{E}_\pi f(X_0) = \pi(f)$, we find,

$$|\mathbb{E}_\nu f(X_n) - \pi(f)| \leq 2\|f\|_\infty \mathbb{P}(H > n) \to 0 \text{ as } n \to \infty.$$

which shows $\mathrm{Law}_{\mathbb{P}_\nu}(X_n) \implies \pi$ as $n \to \infty$, and this gives Eq. (8.29). QED.

Alternative derivation of Eq. (8.30). In Nate Eldredge's notes, the inequality in Eq. (8.30) is derived a bit differently. He take H to be the hitting time of $(x, x) \in \Delta$ and then notes by the strong Markov property that

[4] We still have to handle the null-recurrent case which is going to be omitted in these notes but see Kallenberg [10, Chapter 8, Theorem 8.18, p. 152.]. The key new ingredient not explained here may be found in [10, Lemma 8.21].

[5] By definition the state space is now $S \times S$ and $[\mathbf{P} \otimes \mathbf{P}]_{(x,y),(x',y')} = \mathbf{P}_{x,x'}\mathbf{P}_{y',y'}$.

[6] If $\mathbf{P}$ is periodic, then $\mathbf{P} \otimes \mathbf{P}$ may be reducible. For example if our chain transitions are $1 \to 2 \to 3 \to 1$ all with probability 1, then $\mathbf{P} \otimes \mathbf{P}$ is reducible. For example if the chain starts at $(1, 1)$ it will only visit $(2, 2)$ and $(3, 3)$ and never (i, j) for $i \neq j$.

[7] Again by definition the state space is $S \times S$ and $\pi \otimes \pi : S \times S \to [0, 1]$ is defined by $\pi \otimes \pi(x, y) = \pi(x)\pi(y)$.

$$\begin{aligned}\mathbb{E}_\mu f(X_n) &= \mathbb{E}_{\mu\otimes\nu} f(X_n) \\ &= \sum_{k=0}^{n} \mathbb{E}_{\mu\otimes\nu}[f(X_n)|H=k]\,\mathbb{P}(H=k) + \mathbb{E}_{\mu\otimes\nu}[f(X_n):H>n] \\ &= \sum_{k=0}^{n} \mathbb{E}_x[f(X_{n-k})]\,\mathbb{P}(H=k) + \mathbb{E}_{\mu\otimes\nu}[f(X_n):H>n].\end{aligned}$$

As similar calculation shows

$$\begin{aligned}\mathbb{E}_\nu f(X_n) = \mathbb{E}_{\mu\otimes\nu} f(Y_n) &= \sum_{k=0}^{n} \mathbb{E}_x[f(Y_{n-k})]\,\mathbb{P}(H=k) + \mathbb{E}_{\mu\otimes\nu}[f(Y_n):H>n] \\ &= \sum_{k=0}^{n} \mathbb{E}_x[f(X_{n-k})]\,\mathbb{P}(H=k) + \mathbb{E}_{\mu\otimes\nu}[f(Y_n):H>n].\end{aligned}$$

Then subtracting these two equations shows,

$$\begin{aligned}\mathbb{E}_\mu f(X_n) - \mathbb{E}_\nu f(X_n) &= \mathbb{E}_{\mu\otimes\nu}[f(X_n):H>n] - \mathbb{E}_{\mu\otimes\nu}[f(Y_n):H>n] \\ &= \mathbb{E}_{\mu\otimes\nu}[f(X_n)-f(Y_n):H>n] \\ &= \mathbb{E}_{\mu\otimes\nu}\left[f(X_n)-f\left(\tilde{Y}_n\right):H>n\right]\end{aligned}$$

and so again,

$$|\mathbb{E}_\mu f(X_n) - \mathbb{E}_\nu f(X_n)| \le 2\|f\|_\infty \mathbb{P}(H>n).$$

■

8.4 Some finite state space examples

Example 8.35 (Analyzing a non-irreducible Markov chain). In this example we are going to analyze the limiting behavior of the non-irreducible Markov chain determined by the Markov matrix,

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}.$$

Here are the steps to follow.

1. **Find the jump diagram for P**. In our case it is given in Figure 8.2.

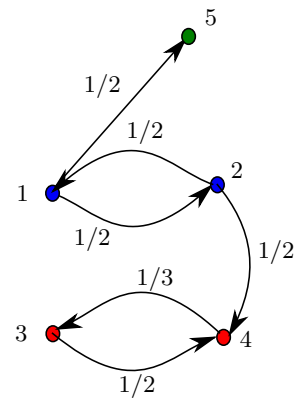


Fig. 8.2. The jump diagram for P.

2. **Identify the communication classes**. In our example they are $\{1,2\}$, $\{5\}$, and $\{3,4\}$. The first is not closed and hence transient while the second two are closed and finite sets and hence recurrent.
3. **Find the invariant distributions for the recurrent classes.** For $\{5\}$ it is simply $\pi'_{\{5\}} = [1]$ and for $\{3,4\}$ we must find the invariant distribution for the 2×2 Markov matrix,

$$Q = \begin{array}{c} \begin{array}{cc} 3 & 4 \end{array} \\ \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 3 \\ 4 \end{array}.$$

We do this in the usual way, namely

$$\operatorname{Nul}\left(I - Q^{\mathrm{tr}}\right) = \operatorname{Nul}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{2}{3} \end{bmatrix}\right) = \mathbb{R}\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

so that $\pi'_{\{3,4\}} = \frac{1}{5}\begin{bmatrix} 2 & 3 \end{bmatrix}$.

4. We can turn $\pi'_{\{3,4\}}$ and $\pi'_{\{5\}}$ into invariant distributions for $\mathbf{P}$ by padding the row vectors with zeros to get

$$\begin{aligned}\pi_{\{3,4\}} &= \begin{bmatrix} 0 & 0 & 2/5 & 3/5 & 0 \end{bmatrix} \\ \pi_{\{5\}} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}.\end{aligned}$$

The general invariant distribution may then be written as;

$$\pi = \alpha\pi_{\{5\}} + \beta\pi_{\{3,4\}} \text{ with } \alpha,\beta \ge 0 \text{ and } \alpha+\beta = 1.$$

5. We can now work out the $\lim_{n\to\infty}\mathbf{P}^n$. If we start at site x we are considering the xth – row of $\lim_{n\to\infty}\mathbf{P}^n$. If we start in the recurrent class $\{3,4\}$ we will

simply get $\pi_{\{3,4\}}$ for these rows and we start in the recurrent class $\{5\}$ we will get $\pi_{\{5\}}$. However if start in the non-closed transient class, $\{1,2\}$ we have

$$\text{first row of } \lim_{n\to\infty} \mathbf{P}^n = \mathbb{P}_1(X_n \text{ hits } 5)\,\pi_{\{5\}} + \mathbb{P}_1(X_n \text{ hits } \{3,4\})\,\pi_{\{3,4\}} \tag{8.31}$$

and

$$\text{second row of } \lim_{n\to\infty} \mathbf{P}^n = \mathbb{P}_2(X_n \text{ hits } 5)\,\pi_{\{5\}} + \mathbb{P}_2(X_n \text{ hits } \{3,4\})\,\pi_{\{3,4\}}. \tag{8.32}$$

Recall that $\{X_n \text{ hits } 5\} = \{H_5 < \infty\}$ and $\{X_n \text{ hits } \{3,4\}\} = \{H_{\{3,4\}} < \infty\}$.

6. **Compute the required hitting probabilities.** Let us now compute the required hitting probabilities by taking $B = \{3,4,5\}$ and $A = \{1,2\}$. Then we have,

$$\begin{aligned}\{\mathbb{P}_x(X_{H_B} = y)\}_{x\in A, y\in B} &= (I - \mathbf{P}_{A\times A})^{-1}\mathbf{P}_{A\times B} \\ &= \left(\begin{bmatrix}1&0\\0&1\end{bmatrix} - \begin{bmatrix}0&1/2\\1/2&0\end{bmatrix}\right)^{-1}\begin{bmatrix}0&0&1/2\\0&1/2&0\end{bmatrix} \\ &= \begin{bmatrix}\frac{4}{3}&\frac{2}{3}\\\frac{2}{3}&\frac{4}{3}\end{bmatrix}\begin{bmatrix}0&0&1/2\\0&1/2&0\end{bmatrix} = \begin{bmatrix}0&\frac{1}{3}&\frac{2}{3}\\0&\frac{2}{3}&\frac{1}{3}\end{bmatrix}.\end{aligned}$$

From this we learn

$$\mathbb{P}_1(X_n \text{ hits } 5) = \frac{2}{3},\ \mathbb{P}_2(X_n \text{ hits } 5) = \frac{1}{3},$$
$$\mathbb{P}_1(X_n \text{ hits } \{3,4\}) = \frac{1}{3} \text{ and } \mathbb{P}_2(X_n \text{ hits } \{3,4\}) = \frac{2}{3}.$$

7. Using these results in Eqs. (8.31) and (8.32) shows,

$$\begin{aligned}\text{first row of } \lim_{n\to\infty}\mathbf{P}^n &= \frac{2}{3}\pi_{\{5\}} + \frac{1}{3}\pi_{\{3,4\}} \\ &= \begin{bmatrix}0&0&\frac{2}{15}&\frac{1}{5}&2/3\end{bmatrix} \\ &= \begin{bmatrix}0.0&0.0&0.133\,33&0.2&0.666\,67\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\text{second row of } \lim_{n\to\infty}\mathbf{P}^n &= \frac{1}{3}\pi_{\{5\}} + \frac{2}{3}\pi_{\{3,4\}} \\ &= \frac{1}{3}\begin{bmatrix}0&0&0&0&1\end{bmatrix} + \frac{2}{3}\begin{bmatrix}0&0&2/5&3/5&0\end{bmatrix} \\ &= \begin{bmatrix}0&0&\frac{4}{15}&\frac{2}{5}&\frac{1}{3}\end{bmatrix} \\ &= \begin{bmatrix}0.0&0.0&0.266\,67&0.4&0.333\,33\end{bmatrix}.\end{aligned}$$

These answers already compare well with

$$\mathbf{P}^{10} = \begin{bmatrix}9.7656\times10^{-4} & 0.0 & 0.132\,76 & 0.200\,24 & 0.666\,02\\ 0.0 & 9.7656\times10^{-4} & 0.266\,26 & 0.399\,76 & 0.333\,01\\ 0.0 & 0.0 & 0.4 & 0.600\,00 & 0.0\\ 0.0 & 0.0 & 0.400\,00 & 0.6 & 0.0\\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0\end{bmatrix}.$$

8.5 Periodic Chain Considerations

Definition 8.36. *For each $x \in S$, let $d(x)$ be the greatest common divisor of $\{n \ge 1 : \mathbf{P}^n_{xx} > 0\}$ with the convention that $d(x) = 0$ if $\mathbf{P}^n_{xx} = 0$ for all $n \ge 1$. We refer to $d(x)$ as the **period** of x. We say a site x is **aperiodic** if $d(x) = 1$.*

Example 8.37. Each site of the fair random walk on $S = \mathbb{Z}$ has period 2. While for the fair random walk on $\{0,1,2,\dots\}$ with 0 being an absorbing state, each $x \ge 1$ has period 2 while 0 has period 1, i.e. 0 is aperiodic.

Theorem 8.38. *The period function is constant on each communication class of a Markov chain.*

Proof. Let $x, y \in C$ and $a = d(x)$ and $b = d(y)$. Now suppose that $\mathbf{P}^m_{xy} > 0$ and $\mathbf{P}^n_{yx} > 0$, then $\mathbf{P}^{m+n}_{x,x} \ge \mathbf{P}^m_{xy}\mathbf{P}^n_{yx} > 0$ and so $a|(m+n)$. Further suppose that $\mathbf{P}^l_{y,y} > 0$ for come $l \in \mathbb{N}$, then

$$\mathbf{P}^{m+n+l}_{x,x} \ge \mathbf{P}^m_{xy}\mathbf{P}^l_{y,y}\mathbf{P}^n_{yx} > 0$$

and therefore $a|(m+n+l)$ which coupled with $a|(m+n)$ implies $a|l$. We may therefore conclude that $a \le b$ (in fact $a|b$) as $b = \gcd\left(\left\{l \in \mathbb{N} : \mathbf{P}^l_{y,y} > 0\right\}\right)$. Similarly we show that $b \le a$ and therefore $b = a$. ∎

Lemma 8.39. *If $d(x)$ is the period of site x, then*

1. *if $m \in \mathbb{N}$ and $\mathbf{P}^m_{x,x} > 0$ then $d(x)$ divides m and*
2. *$\mathbf{P}^{nd(x)}_{x,x} > 0$ for all $n \in \mathbb{N}$ sufficiently large.*
3. *If x is aperiodic iff $\mathbf{P}^n_{x,x} > 0$ for all $n \in \mathbb{N}$ sufficiently large.*

In summary, $A_x := \left\{m \in \mathbb{N} : \mathbf{P}^m_{x,x} > 0\right\} \subset d(x)\mathbb{N}$ and $d(x)n \in A_x$ for all $n \in \mathbb{N}$ sufficiently large.

Proof. Choose $n_1, \dots, n_k \in \{n \ge 1 : \mathbf{P}^n_{xx} > 0\}$ such that $d(x) = \gcd(n_1, \dots, n_k)$. For part 1. we also know that $d(x) = \gcd(n_1, \dots, n_k, m)$ and therefore $d(x)$ divides m. For part 2., if $m_x \in \mathbb{N}$ we have,

$$\left(\mathbf{P}^{\sum_{l=1}^{k} m_l n_l}\right)_{x,x} \geq \prod_{l=1}^{k} \left[\mathbf{P}^{n_l}_{x,x}\right]^{m_l} > 0.$$

This observation along with the number theoretic Lemma 8.44 below is enough to show $\mathbf{P}^{nd(x)}_{x,x} > 0$ for all $n \in \mathbb{N}$ sufficiently large. The third item is a special case of item 2. ■

Example 8.40. Suppose that $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then $\mathbf{P}^m = \mathbf{P}$ if m is odd and $\mathbf{P}^m = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ if m is even. Therefore $d(x) = 2$ for $x = 1, 2$ and in this case $\mathbf{P}^{2n}_{x,x} = 1 > 0$ for all $n \in \mathbb{N}$. However observe that $\mathbf{P}^2$ is no longer irreducible – there are now two communication classes.

Example 8.41. Consider the Markov chain with jump diagram given in Figure 8.3.In this example, $d(x) = 2$ for all x and all states for $\mathbf{P}^2$ are aperiodic.

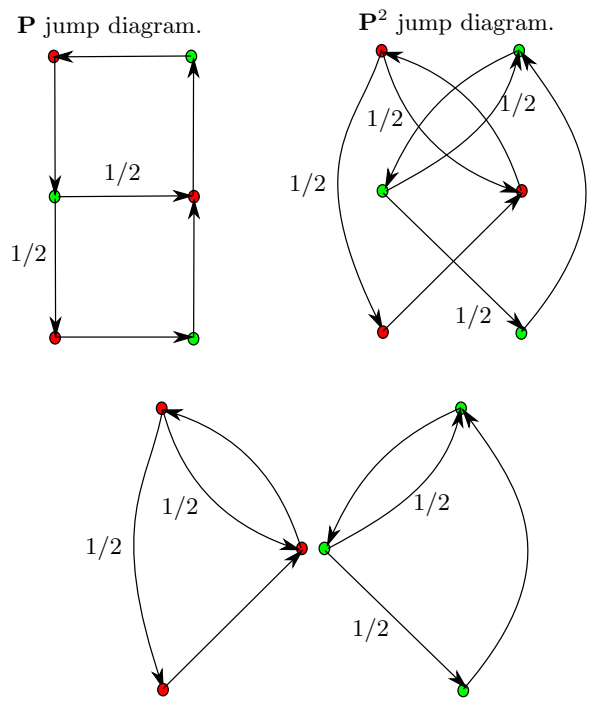


Fig. 8.3. All arrows are assumed to have weight 1 unless otherwise specified. Notice that each state has period $d = 2$ and that $\mathbf{P}^2$ is the transition matrix having two aperiodic communication classes.

However $\mathbf{P}^2$ is no longer irreducible. This is an indication of the what happens in general. In terms of matrices,

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \text{ and } \mathbf{P}^2 = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}.$$

Example 8.42. Consider the Markov chain with jump diagram given in Figure 8.4.Assume there are no implied jumps from a site back to itself, i.e. $\mathbf{P}_{x,x} = 0$

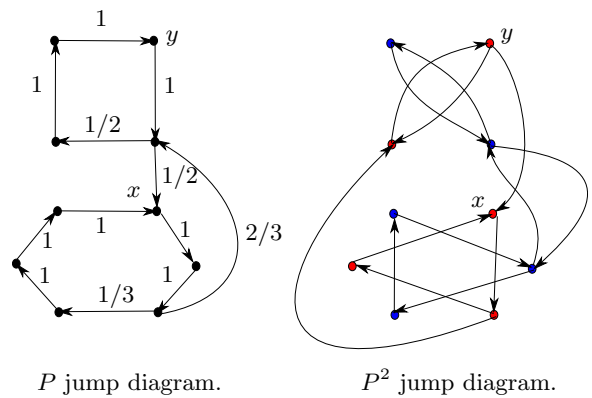


Fig. 8.4. Assume there are no implied jumps from a site back to itself, i.e. $\mathbf{P}_{i,i} = 0$ for all i. This chain is then irreducible and has period 2.

for all x. This chain is then irreducible and has period 2. To calculate the period notice that starting at y there is an obvious loop of length 4 and starting at x there is one of length 6. Therefore the period must divide both 4 and 6 and so must be either 2 or 1. The period is not 1 as one can only return to a site with an even number of jumps in this picture. If on the other hand there was any one vertex, x, such that $\mathbf{P}_{x,x} = 1$, then the period of the chain would have been one, i.e. the chain would have been aperiodic. Further notice that the jump diagram for $\mathbf{P}^2$ is no longer irreducible. The red vertices and the blue vertices split apart. This has to happen as is a consequence of Proposition 8.43 below.

Proposition 8.43. *If* $\mathbf{P}$ *is the Markov matrix for a finite state irreducible aperiodic chain, then there exists* $n_0 \in \mathbb{N}$ *such that* $\mathbf{P}^n_{xy} > 0$ *for all* $x, y \in S$ *and* $n \geq n_0$.

Proof. Let $x, y \in S$. By Lemma 8.39 with $d(x) = 1$ we know that $\mathbf{P}^m_{x,x} > 0$ for all m large. As $\mathbf{P}$ is irreducible there exists $a \in \mathbb{N}$ such that $\mathbf{P}^a_{xy} > 0$ and therefore $\mathbf{P}^{a+m}_{x,y} \geq \mathbf{P}^m_{x,x}\mathbf{P}^a_{x,y} > 0$ for all m sufficiently large. This shows for all $x, y \in S$ there exists $n_{x,y} \in \mathbb{N}$ such that $\mathbf{P}^n_{xy} > 0$ for all $n \geq n_{x,y}$. Since there are only finitely many steps we may now take $n_0 := \max\{n_{x,y} : x, y \in S\} < \infty$. ■

8.5.1 A number theoretic lemma

Lemma 8.44 (A number theory lemma). *Suppose that* 1 *is the greatest common denominator of a set of positive integers,* $\Gamma := \{n_1, \ldots, n_k\}$. *Then there exists* $N \in \mathbb{N}$ *such that the set,*

$$A = \{m_1 n_1 + \cdots + m_k n_k : m_i \geq 0 \text{ for all } i\},$$

contains all $n \in \mathbb{N}$ *with* $n \geq N$. *More generally if* $q = \gcd(\Gamma)$ *(perhaps not* 1*), then* $A \subset q\mathbb{N}$ *and contains all points* qn *for* n *sufficiently large.*

Proof. First proof. The set $I := \{m_1 n_1 + \cdots + m_k n_k : m_i \in \mathbb{N} \text{ for all } i\}$ is an ideal in $\mathbb{Z}$ and as $\mathbb{Z}$ is a principle ideal domain there is a $q \in I$ with $q > 0$ such that $I = q\mathbb{Z} = \{qm : m \in \mathbb{Z}\}$. In fact $q = \min(I \cap \mathbb{N})$. Since $q \in I$ we know that $q = m_1 n_1 + \cdots + m_k n_k$ for some $m_i \in \mathbb{N}$ and so if l is a common divisor of $n_1, \ldots, n_k$ then l divides q. Moreover as $I = q\mathbb{Z}$ and $n_i \in I$ for all i, we know that $q | n_i$ as well. This shows that $q = \gcd(n_1, n_2, \ldots, n_k)$.

Now suppose that $n \gg n_1 + \cdots + n_k$ is given and large (to be explained shortly). Then write $n = l(n_1 + \cdots + n_k) + r$ with $l \in \mathbb{N}$ and $0 \leq r < n_1 + \cdots + n_k$ and therefore,

$$\begin{aligned} nq &= ql(n_1 + \cdots + n_k) + rq \\ &= ql(n_1 + \cdots + n_k) + r(m_1 n_1 + \cdots + m_k n_k) \\ &= (ql + rm_1) n_1 + \cdots + (ql + rm_k) n_k \end{aligned}$$

where

$$ql + rm_i \geq ql - (n_1 + \cdots + n_k) m_i$$

which is greater than 0 for l and hence n sufficiently large.

Second proof. (The following proof is from Durrett [6].) We first will show that A contains two consecutive positive integers, a and $a + 1$. To prove this let,

$$k := \min\{|b - a| : a, b \in A \text{ with } a \neq b\}$$

and choose $a, b \in A$ with $b = a + k$. If $k > 1$, there exists $n \in \Gamma \subset A$ such that k does not divide n. Let us write $n = mk + r$ with $m \geq 0$ and $1 \leq r < k$. It then follows that $(m+1)b$ and $(m+1)a + n$ are in A,

$$(m+1)b = (m+1)(a+k) > (m+1)a + mk + r = (m+1)a + n,$$

and

$$(m+1)b - (m+1)a + n = k - r < k.$$

This contradicts the definition of k and therefore, $k = 1$.

Let $N = a^2$. If $n \geq N$, then $n - a^2 = ma + r$ for some $m \geq 0$ and $0 \leq r < a$. Therefore,

$$n = a^2 + ma + r = (a + m)a + r = (a + m - r)a + r(a + 1) \in A.$$

■

*Proofs of Long Run Results

In proving the results above, we are going to make essential use of a strong form of the Markov property of Theorem 7.11.

9.1 Strong Markov Property Consequences

Let

$$f_{xx}^{(n)} = \mathbb{P}_x(R_x = n) = \mathbb{P}_x(X_1 \neq x, \ldots, X_{n-1} \neq x, X_n = x)$$

and $m_{xy} := \mathbb{E}_x(M_y)$ – the expected number of visits to y after $n = 0$.

Proposition 9.1. *Let $x \in S$ and $n \geq 1$. Then $\mathbf{P}_{xx}^n$ satisfies the "renewal equation,"*

$$\mathbf{P}_{xx}^n = \sum_{m=1}^n \mathbb{P}(R_x = m)\mathbf{P}_{xx}^{n-m}. \tag{9.1}$$

Also if $y \in S$, $K \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then Eq. (8.5) holds, i.e.

$$\mathbb{P}_\nu\left(M_y \geq K\right) = \mathbb{P}_\nu\left(H_y < \infty\right) \cdot \mathbb{P}_y\left(R_y < \infty\right)^{K-1}. \tag{9.2}$$

[Equation (9.2) was also proved in Proposition 8.15.]

Proof. To prove Eq. (9.1) we first observe for $n \geq 1$ that $\{X_n = x\}$ is the disjoint union of $\{X_n = x, R_x = m\}$ for $1 \leq m \leq n$ and therefore[1],

$$\begin{aligned}
\mathbf{P}_{xx}^n &= \mathbb{P}_x(X_n = x) = \sum_{m=1}^n \mathbb{P}_x(R_x = m, X_n = x) \\
&= \sum_{m=1}^n \mathbb{P}_x(X_1 \neq x, \ldots, X_{m-1} \neq x, X_m = x, X_n = x) \\
&= \sum_{m=1}^n \mathbb{P}_x(X_1 \neq x, \ldots, X_{m-1} \neq x, X_m = x)\mathbf{P}_{xx}^{n-m} \\
&= \sum_{m=1}^n \mathbf{P}_{xx}^{n-m}\mathbb{P}(R_x = m).
\end{aligned}$$

For Eq. (9.2) we have $\{M_y \geq 1\} = \{R_y < \infty\}$ so that $\mathbb{P}_x\left(M_y \geq 1\right) = \mathbb{P}_x\left(R_y < \infty\right)$. For $K \geq 2$, since $R_y < \infty$ if $M_y \geq 1$, we have

$$\mathbb{P}_x\left(M_y \geq K\right) = \mathbb{P}_x\left(M_y \geq K | R_y < \infty\right)\mathbb{P}_x\left(R_y < \infty\right).$$

Since, on $R_y < \infty$, $X_{R_y} = y$, it follows by the strong Markov property of Theorem 7.11 that;

$$\begin{aligned}
\mathbb{P}_x\left(M_y \geq K | R_y < \infty\right) &= \mathbb{P}_x\left(M_y \geq K | R_y < \infty,\ X_{R_y} = y\right) \\
&= \mathbb{P}_x\left(1 + \sum_{n \geq 1} 1_{X_{R_y+n} = y} \geq K | R_y < \infty,\ X_{R_y} = y\right) \\
&= \mathbb{P}_y\left(1 + \sum_{n \geq 1} 1_{X_n = y} \geq K\right) = \mathbb{P}_y\left(M_y \geq K - 1\right).
\end{aligned}$$

By the last two displayed equations,

$$\mathbb{P}_x\left(M_y \geq K\right) = \mathbb{P}_y\left(M_y \geq K - 1\right)\mathbb{P}_x\left(R_y < \infty\right) \tag{9.3}$$

Taking $x = y$ in this equation shows,

$$\mathbb{P}_y\left(M_y \geq K\right) = \mathbb{P}_y\left(M_y \geq K - 1\right)\mathbb{P}_y\left(R_y < \infty\right)$$

and so by induction,

$$\mathbb{P}_y\left(M_y \geq K\right) = \mathbb{P}_y\left(R_y < \infty\right)^K. \tag{9.4}$$

Equation (9.2) now follows from Eqs. (9.3) and (9.4). ■

[1] Alternatively, we could use the Markov property to show,

$$\begin{aligned}
P_{ii}^n = P_i(X_n = i) &= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot 1_{X_n = i}) = \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot \mathbb{E}_i 1_{X_{n-k} = i}) \\
&= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k})\mathbb{E}_i\left(1_{X_{n-k} = i}\right) = \sum_{k=1}^n P_i(R_i = k)P_i(X_{n-k} = i) \\
&= \sum_{k=1}^n P_{ii}^{n-k}P(R_i = k).
\end{aligned}$$

9.2 Irreducible Recurrent Chains

For this section we are going to assume that X_n is an *irreducible and recurrent* Markov chain. Let us now fix a state, $y \in S$ and define,

$$\begin{aligned}\tau_1 &= H_y = \min\{n \geq 0 : X_n = y\},\\ \tau_2 &= \min\{n \geq 1 : X_{n+\tau_1} = y\},\\ &\vdots\\ \tau_N &= \min\{n \geq 1 : X_{n+\tau_{N-1}} = y\},\end{aligned}$$

so that τ_N is the time it takes for the chain to visit y after the $(N-1)$'st visit to y. By Corollary 8.20 we know that $\mathbb{P}_\nu(\tau_N < \infty) = 1$ for all $N \in \mathbb{N}$ and any starting distribution, $\nu : S \to [0,1]$. We will use strong Markov property to prove the following key lemma in our development.

Lemma 9.2. *We continue to use the notation above and in particular assume that $\{X_n\}_{n=0}^\infty$ is an irreducible recurrent Markov chain. If $\nu : S \to [0,1]$ is any starting distribution, the random times, $\{\tau_N\}_{N=1}^\infty$, above are independent and moreover $\{\tau_N\}_{N=2}^\infty$ are identically distributed with $\mathbb{P}_\nu(\tau_N = n) = \mathbb{P}_y(R_y = n)$ for all $n \in \mathbb{N}$ and $N \geq 2$.*

Proof. This lemma follows rather immediately from Remark 8.14. Here is a slightly longer argument for those wishing for more detail.

Let $T_0 = 0$ and then define T_N inductively by, $T_{N+1} = \inf\{n > T_N : X_n = y\}$ so that T_N is the time of the N'th visit of $\{X_n\}_{n=1}^\infty$ to site y. Observe that $T_1 = \tau_1$,

$$\tau_{N+1}(X_0, X_1, \dots) = \tau_1\left(X_{T_N}, X_{T_N+1}, X_{T_{N+2}}, \dots\right),$$

and $(\tau_1, \dots, \tau_N)$ is a function of $(X_0, \dots, X_{T_N})$. Since $\mathbb{P}_\nu(T_N < \infty) = 1$ (Corollary 8.20) and $X_{T_N} = y$, we may apply the strong Markov property in the form of Theorem 7.11 to learn:

1. τ_{N+1} is independent of $(X_0, \dots, X_{T_N})$ and hence τ_{N+1} is independent of $(\tau_1, \dots, \tau_N)$, and
2. the distribution of τ_{N+1} under $\mathbb{P}_\nu$ is the same as the distribution of R_y under $\mathbb{P}_y$.

The result now follows from these two observations and induction. ■

Theorem 9.3. *Suppose that X_n is a irreducible recurrent Markov chain, and let $y \in S$ be a fixed state. Define*

$$\pi_y = \pi(y) := \frac{1}{\mathbb{E}_y R_y}, \tag{9.5}$$

with the understanding that $\pi_y = 0$ if $\mathbb{E}_y R_y = \infty$. If $\nu : S \to [0,1]$ is any probability distribution on S, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^n 1_{X_m = y} = \pi_y \quad \mathbb{P}_\nu - a.s. \tag{9.6}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^n \mathbb{P}_\nu(X_m = y) = \pi_y \tag{9.7}$$

and in particular (take $\nu = \delta_x$)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^n \mathbf{P}^m_{xy} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^n \mathbb{P}_x(X_m = y) = \pi_y. \tag{9.8}$$

for all $x \in S$.

Proof. Let us first note that Eq. (9.7) follows by taking expectations of Eq. (9.6). So it only remains to prove Eq. (9.6).

By Corollary 8.20 we already know that $1 = \mathbb{P}_\nu(\tau_1 < \infty) = \mathbb{P}_\nu(H_y < \infty)$ and so that $\mathbb{P}_\nu$ - a.s. the chain will hit y. By Remark 8.14 (or more formally by Lemma 9.2(, the sequence $\{\tau_N\}_{N\geq 2}$ is i.i.d. relative to $\mathbb{P}_\nu$ and $\mathbb{E}_\nu \tau_N = \mathbb{E}_y R_y$. We may now use the strong law of large numbers to conclude that

$$\lim_{N\to\infty} \frac{\tau_1 + \tau_2 + \cdots + \tau_N}{N} = \mathbb{E}_\nu \tau_2 = \mathbb{E}_y R_y \qquad (\mathbb{P}_\nu\text{– a.s.}). \tag{9.9}$$

This may be expressed as follows, let $T_N = \tau_1 + \tau_2 + \cdots + \tau_N$, be the time when the chain first visits y for the Nth occurrence, then

$$\lim_{N\to\infty} \frac{T_N}{N} = \mathbb{E}_y R_y \qquad (\mathbb{P}_\nu\text{– a.s.}) \tag{9.10}$$

Given a time $n \in \mathbb{N}$, let

$$N_n = \sum_{m=0}^n 1_{X_m = y} = (\#\text{ visits of } y \text{ up to time } n).$$

We will make use of the following two simple observations.

- Since y is visited infinitely often, $N_n \to \infty$ as $n \to \infty$ and therefore, $\lim_{n\to\infty} \frac{N_n+1}{N_n} = 1$.
- A little thought shows that the following arrangements of times must hold;

  $$0 - - - - - - T_{N_n} - - n - - - - - T_{(N_n+1)} - - - > \text{time}.$$

  In more detail, since there were N_n visits to y in the first n steps, the time of the ${N_n}^{\text{th}}$ visit to y must be less than or equal to n, i.e. $T_{N_n} \leq n$. Similarly, the time, T_{N_n+1}, of the $(N_n+1)^{\text{st}}$ visit to y must be larger than n, i.e. $n < T_{N_n+1}$.

Using the last bullet point we have

$$\frac{T_{N_n}}{N_n} \le \frac{n}{N_n} < \frac{T_{N_n+1}}{N_n} = \frac{T_{N_n+1}}{N_n+1} \cdot \frac{N_n+1}{N_n}.$$

Letting $n \to \infty$ in these inequalities, making use of the "sandwich" theorem, the first bullet point, and Eq. (9.10) shows;

$$\lim_{n\to\infty} \frac{n}{N_n} = \mathbb{E}_y R_y \qquad \mathbb{P}_\nu - a.s.$$

Taking reciprocals of the last equation gives Eq. (9.6).

■

Remark 9.4 (Rates of Convergence). According to Proposition 2.18 above[2] or Corollary 13.59 below, roughly speaking we know that $\frac{T_N}{N} = \mathbb{E}_x R_x + O\left(\frac{1}{\sqrt{N}}\right)$ and therefore

$$\begin{aligned}\frac{N_n}{n} &\sim \left(\frac{T_{N_n}}{N_n}\right)^{-1} = \frac{1}{\mathbb{E}_x R_x + O\left(\frac{1}{\sqrt{N_n}}\right)} \\ &= \frac{1}{\mathbb{E}_x R_x}\left(1 + O\left(\frac{1}{\sqrt{N_n}}\right)\right) = \frac{1}{\mathbb{E}_x R_x} + O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

What we could formally prove is that for any $0 < \alpha < 1/2$ we have

$$\lim_{n\to\infty} n^\alpha \left|\frac{N_n}{n} - \frac{1}{\mathbb{E}_x R_x}\right| = 0 \ \mathbb{P}_\nu - \text{a.s.}$$

We will discuss rates of convergence more in Chapter 11.

Proposition 9.5. *Suppose that X_n is a irreducible, recurrent Markov chain and let $\pi_x = 1/(\mathbb{E}_x R_x)$ for all $x \in S$ as in Eq. (9.5). Then either $\pi_x = 0$ for all $x \in S$ (in which case X_n is null recurrent) or $\pi_x > 0$ for all $x \in S$ (in which case X_n is positive recurrent). Moreover if $\pi_x > 0$ then*

$$\sum_{x\in S} \pi_x = 1 \text{ and} \tag{9.11}$$

$$[\pi \mathbf{P}]_y := \sum_{x\in S} \pi_x \mathbf{P}_{xy} = \pi_y \text{ for all } y \in S. \tag{9.12}$$

That is $\pi = (\pi_x)_{x\in S}$ is the unique stationary distribution for $\mathbf{P}$.

[2] Here we also use the results in 5.24 along with the first step analysis to conclude that $\mathbb{E}_x R_x^N < \infty$ for all $N \in \mathbb{N}$.

Proof. Suppose that $\mathbb{E}_{x_0} R_{x_0} < \infty$ for some x_0 i.e. $\pi(x_0) > 0$ for some $x_0 \in S$ and let

$$\alpha := \sum_{x\in S} \pi(x).$$

For any starting distribution, $\nu : S \to [0,1]$, we have

$$\sum_{x\in S} \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_m = x) = \frac{1}{n} \sum_{m=1}^n \sum_{x\in S} \mathbb{P}_\nu(X_m = x) = \frac{1}{n} \sum_{m=1}^n 1 = 1.$$

This equality along Eq. (9.7) and Fatou's lemma gives;

$$\begin{aligned}0 < \pi(x_0) \le \alpha &= \sum_{x\in S} \pi(x) = \sum_{x\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_m = x) \\ &\le \liminf_{n\to\infty} \sum_{x\in S} \frac{1}{n} \sum_{n=1}^n \mathbb{P}_\nu(X_m = x) = 1.\end{aligned}$$

Using

$$\sum_{x\in S} \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} = \sum_{x\in S} [\nu \mathbf{P}^m]_x \mathbf{P}_{xy} = \left[\nu \mathbf{P}^{m+1}\right]_y = \mathbb{P}_\nu(X_{m+1} = y),$$

it is easy to conclude that

$$\sum_{x\in S} \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} = \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_{m+1} = y) \to \pi(y) \text{ as } n \to \infty.$$

So by another application of Fatou's lemma along Eq. (9.7),

$$\begin{aligned}\sum_{x\in S} \pi(x)\mathbf{P}_{xy} &= \sum_{x\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} \\ &\le \liminf_{n\to\infty} \frac{1}{n} \sum_{m=1}^n \mathbb{P}_\nu(X_{m+1} = y) = \pi(y).\end{aligned}$$

If $\sum_{x\in S} \pi(x)\mathbf{P}_{x,y_0} < \pi(y_0)$ for some $y_0 \in S$, we would find,

$$\alpha = \sum_{x\in S} \pi(x) = \sum_{x\in S} \pi(x) \sum_{y\in S} \mathbf{P}_{xy} = \sum_{y\in S} \sum_{x\in S} \pi(x)\mathbf{P}_{xy} < \sum_{y\in S} \pi(y) = \alpha$$

which is absurd and so we must conclude that in fact;

$$\sum_{x\in S} \pi(x)\mathbf{P}_{x,y} = \pi(y) \text{ for all } y \in S,$$

i.e. $\pi\mathbf{P} = \pi$.

Taking $\nu := \alpha^{-1}\pi$ (so ν is an invariant distribution on S) in Eq. (9.7) allows us to conclude that, for all $y \in S$,

$$\pi(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_{\alpha^{-1}\pi}(X_m = y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \alpha^{-1}\pi(y) = \alpha^{-1}\pi(y).$$

Using this equation with $y = x_0$ (where $\pi(x_0) > 0$) forces $\alpha = 1$ and so π is indeed an invariant distribution for $\mathbf{P}$.

The assertion that $\mathbb{E}_x R_x < \infty$ (i.e. $\pi(x) > 0$) for all $x \in S$ follows directly from Proposition 8.19. Alternatively, since x communicates with x_0 there exists $m \in \mathbb{N}$ so that $\mathbf{P}^m_{x_0,x} > 0$ and hence

$$\pi(x) = [\pi\mathbf{P}^m]_x = \sum_{y\in S} \pi(y)\mathbf{P}^m_{y,x} \geq \pi(x_0)\mathbf{P}^m_{x_0,x} > 0.$$

Lastly if ν is another invariant distribution for $\mathbf{P}$ then form Eq. (9.7) we will have

$$\pi(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} \mathbb{P}_\nu(X_m = y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} \nu(y) = \lim_{n\to\infty} \frac{n+1}{n}\nu(y) = \nu(y)$$

which shows that the invariant distribution is unique.

■

Remark 9.6. If we further assume that $\#(S) < \infty$ in Proposition 9.5, then the previous proof becomes much simpler. For in this case we know by our hitting time estimates that $\mathbb{E}_x R_x < \infty$ for all $x \in S$ and

$$\begin{aligned}\sum_{x\in S} \pi(x) &= \sum_{x\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\nu(X_m = x) = \lim_{n\to\infty} \frac{1}{n} \sum_{x\in S} \sum_{m=1}^{n} \mathbb{P}_\nu(X_m = x) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x\in S} \mathbb{P}_\nu(X_m = x) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1 = 1.\end{aligned}$$

Similarly, in the previous proof we have directly that

$$\begin{aligned}\sum_{x\in S} \pi(x)\mathbf{P}_{xy} &= \sum_{x\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} \\ &= \lim_{n\to\infty} \sum_{x\in S} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{x\in S} \mathbb{P}_\nu(X_m = x)\mathbf{P}_{xy} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\nu(X_{m+1} = y) = \pi(y)\end{aligned}$$

showing π is the invariant distribution for $\mathbf{P}$.

Proposition 9.7. *Suppose that $\mathbf{P}$ is an irreducible Markov kernel which admits a stationary distribution μ. Then $\mathbf{P}$ is positive recurrent and $\mu_y = \pi_y = 1/(\mathbb{E}_y R_y)$ for all $y \in S$. In particular, an irreducible Markov kernel has at most one invariant distribution and it has exactly one iff $\mathbf{P}$ is positive recurrent.*

Proof. Taking $\nu = \mu$ in Eq. (9.7) shows

$$\pi(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{P}_\mu(X_m = y) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mu(y) = \mu(y)$$

and so $\mu(y) = \pi(y)$ is an invariant distribution and so $\mathbb{E}_x R_x = \frac{1}{\pi(x)} < \infty$ for some $x \in S$. The fact that $\mathbf{P}$ is positive recurrent now follows from Proposition 9.5.

■

Transience and Recurrence Examples

Let us begin by recalling some of the basic definitions and theorems involving recurrence and transience.

Definition 10.1. *A state $x \in S$ is:*

1. ***transient*** *if* $\mathbb{P}_x(R_x < \infty) < 1 \ \left(\iff \mathbb{P}_x(R_x = \infty) > 0 \right)$,
2. ***recurrent*** *if* $\mathbb{P}_x(R_x < \infty) = 1 \ \left(\iff \mathbb{P}_x(R_x = \infty) = 0 \right)$,
   a) ***positive recurrent*** *if* $1/\left(\mathbb{E}_x R_x\right) > 0$, *i.e.* $\mathbb{E}_x R_x < \infty$,
   b) ***null recurrent*** *if it is recurrent* ($\mathbb{P}_x(R_x < \infty) = 1$) *and* $1/\left(\mathbb{E}_x R_x\right) = 0$, *i.e.* $\mathbb{E} R_x = \infty$.

We let S_t, S_r, S_{pr}, and S_{nr} be the transient, recurrent, positive recurrent, and null recurrent states respectively.

The following theorem summarizes Theorem 8.18 and Corollary 8.20.

Theorem 10.2 (Recurrent States). *Let $y \in S$. Then the following are equivalent;*

1. *y is recurrent, i.e.* $\mathbb{P}_y\left(R_y < \infty\right) = 1$,
2. $\mathbb{P}_y\left(X_n = y \text{ i.o. } n\right) = 1$,
3. $\mathbb{E}_y M_y = \sum_{n=0}^{\infty} \mathbf{P}_{yy}^n = \infty$.

Moreover if $C \subset S$ is a recurrent communication class, then $\mathbb{P}_x(\cap_{y \in C}\{X_n = y \text{ i.o. } n\}) = 1$ for all $x \in C$.In words, if we start in C then every state in C is visited an infinite number of times.

The next result summarizes Propositions 9.5 and 9.7.

Proposition 10.3. *Suppose that $\mathbf{P}$ is an irreducible Markov kernel, $\{X_n\}$ is the corresponding Markov Chain, and $\pi_y = 1/\left(\mathbb{E}_y R_y\right)$ for all $y \in S$.*

1. *If $\{X_n\}$ is recurrent, then either $\pi_x = 0$ for all $x \in S$ (in which case X_n is null recurrent) or $\pi_x > 0$ for all $x \in S$ (in which case X_n is positive recurrent). Moreover if $\pi_x > 0$ then $\pi = (\pi_x)_{x \in S}$ is the unique stationary distribution for $\mathbf{P}$.*
2. *If $\{X_n\}$ admits a stationary distribution μ, then $\{X_n\}$ (or equivalently $\mathbf{P}$) is positive recurrent and $\mu_y = \pi_y = 1/\left(\mathbb{E}_y R_y\right)$ for all $y \in S$.*

In particular, an irreducible Markov kernel, $\mathbf{P}$, has at most one invariant distribution and it has exactly one iff $\mathbf{P}$ is positive recurrent.

The next theorem summarizes Theorem 8.21.

Theorem 10.4 (Transient States). *Let $y \in S$. Then the following are equivalent;*

1. *y is transient, i.e.* $\mathbb{P}_y\left(R_y < \infty\right) < 1$,
2. $\mathbb{P}_y\left(X_n = y \text{ i.o. } n\right) = 0$, *and*
3. $\mathbb{E}_y M_y = \sum_{n=0}^{\infty} \mathbf{P}_{yy}^n < \infty$.

More generally if $\nu : S \to [0,1]$ is any probability and $y \in S$ is transient, then

$$\mathbb{E}_\nu M_y = \sum_{n=0}^{\infty} \mathbb{P}_\nu\left(X_n = y\right) < \infty \implies \begin{cases} \lim_{n\to\infty} \mathbb{P}_\nu\left(X_n = y\right) = 0 \\ \mathbb{P}_\nu\left(X_n = y \text{ i.o. } n\right) = 0. \end{cases} \tag{10.1}$$

For the reader's convenience here is a combination of Theorems 8.18 and 8.21.

Theorem 10.5 (Recurrent/Transient States). *Let $y \in S$. Then the following are equivalent;*

1. *y is recurrent (transient), i.e.* $\mathbb{P}_y\left(R_y < \infty\right) = 1 \ (< 1)$
2. $\mathbb{P}_y\left(X_n = y \text{ i.o. } n\right) = 1 \ (= 0)$,
3. $\mathbb{E}_y M_y = \sum_{n=0}^{\infty} \mathbf{P}_{yy}^n = \infty \ (< \infty)$.

Moreover if y is either transient or null-recurrent, then

$$\lim_{n\to\infty} \mathbb{P}_\nu\left(X_n = y\right) = \lim_{n\to\infty} \left[\nu \mathbf{P}^n\right]_y = 0.$$

10.1 Examples

Example 10.6. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $\{0, 1, 2, \dots\}$ with 0 being an absorbing state and $H = H_0$ be the first hitting time of 0. The communication classes are now $\{0\}$ and $\{1, 2, \dots\}$ with the latter class not being

closed and hence is transient. Using Exercise 6.7 or Exercise 6.8, it follows that $\mathbb{E}_x H_0 = \infty$ for all $x > 0$ which shows we can not drop the assumption that $\#(C) < \infty$ in the first statement in Proposition 8.9. Similarly, using the fair random walk example, we see that it is not possible to drop the condition that $\#(C) < \infty$ for the equivalence statements as well.

The next examples show that if $C \subset S$ is closed and $\#(C) = \infty$, then C could be recurrent or it could be transient. Transient in this case means the chain goes off to "infinity," i.e. eventually leaves every finite subset of C never to return again. In Examples 10.7 – 10.9 below, let $X = \{X_n\}$ be a (possibly biased) random walk on $S = \mathbb{Z}$, $R_y = \inf\{n > 0 : X_n = y\}$ and $H_y = \inf\{n \geq 0 : X_n = y\}$.

Example 10.7. Let $S = \mathbb{Z}$ and $X = \{X_n\}$ be the standard fair random walk on $\mathbb{Z}$, i.e. $\mathbb{P}(X_{n+1} = x \pm 1 | X_n = x) = \frac{1}{2}$. Then S itself is a closed class and every element of S is (null) recurrent. Indeed, using Exercise 6.5 or Exercise 6.6 and the first step analysis we know that

$$\begin{aligned}\mathbb{P}_0[R_0 = \infty] &= \frac{1}{2}\left(\mathbb{P}_0[R_0 = \infty | X_1 = 1] + \mathbb{P}_0[R_0 = \infty | X_1 = -1]\right) \\ &= \frac{1}{2}\left(\mathbb{P}_1[H_0 = \infty] + \mathbb{P}_{-1}[H_0 = \infty]\right) = \frac{1}{2}(0 + 0) = 0.\end{aligned}$$

This shows 0 is recurrent. Similarly using Exercise 6.7 or Exercise 6.8 and the first step analysis we find,

$$\begin{aligned}\mathbb{E}_0[R_0] &= \frac{1}{2}\left(\mathbb{E}_0[R_0 | X_1 = 1] + \mathbb{E}_0[R_0 | X_1 = -1]\right) \\ &= \frac{1}{2}\left(1 + \mathbb{E}_1[H_0] + 1 + \mathbb{E}_{-1}[H_0]\right) = \frac{1}{2}(\infty + \infty) = \infty\end{aligned}$$

and so 0 is null recurrent. As this chain is invariant under translation it follows that every $x \in \mathbb{Z}$ is a null recurrent site.

Example 10.8. Let $S = \mathbb{Z}$ and $X = \{X_n\}$ be a biassed random walk on $\mathbb{Z}$, i.e. $\mathbb{P}(X_{n+1} = x + 1 | X_n = x) = p$ and $\mathbb{P}(X_{n+1} = x - 1 | X_n = x) = q := 1 - p$ with $p > \frac{1}{2}$. Then every site of is now transient. Recall from Exercises 6.9 and 6.10 (see Eq. (6.27)) that

$$\mathbb{P}_x(H_0 < \infty) = \begin{cases} (q/p)^x & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases}. \tag{10.2}$$

Using these result and the first step analysis implies,

$$\begin{aligned}\mathbb{P}_0[R_0 = \infty] &= pP_0[R_0 = \infty | X_1 = 1] + qP_0[R_0 = \infty | X_1 = -1] \\ &= pP_1[H_0 = \infty] + qP_{-1}[H_0 = \infty] \\ &= p\left[1 - (q/p)^1\right] + q(1 - 1) \\ &= p - q = 2p - 1 > 0.\end{aligned}$$

Example 10.9. Again let $S = \mathbb{Z}$ and $p \in \left(\frac{1}{2}, 1\right)$ and suppose that $\{X_n\}$ is the random walk on $\mathbb{Z}$ described the jump diagram in Figure 10.1. In this case using

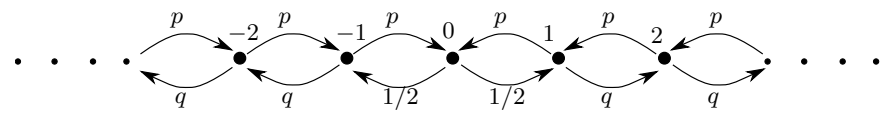


Fig. 10.1. A positively recurrent Markov chain.

the results of Exercise 6.11 we learn that

$$\begin{aligned}\mathbb{E}_0[R_0] &= \frac{1}{2}\left(\mathbb{E}_0[R_0 | X_1 = 1] + \mathbb{E}_0[R_0 | X_1 = -1]\right) \\ &= \frac{1}{2}\left(1 + \mathbb{E}_1[H_0] + 1 + \mathbb{E}_{-1}[H_0]\right) \\ &= 1 + \frac{1}{2}\left(\frac{1}{p-q} + \frac{1}{p-q}\right) = 1 + \frac{1}{p-q} = \frac{2p}{2p-1} < \infty.\end{aligned}$$

This shows the site 0 is positively recurrent. Thus according to Proposition 8.6, every site in $\mathbb{Z}$ is positively recurrent. (Notice that $\mathbb{E}_0[R_0] \to \infty$ as $p \downarrow \frac{1}{2}$, i.e. as the chain becomes closer to the unbiased random walk of Example 10.7.

Example 10.10. Let us revisit the fair random walk on $\mathbb{Z}$ describe before Exercise 6.5. In this case $\mathbb{P}_0(X_n = 0) = 0$ if n is odd and

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2}\left(\frac{1}{2}\right)^{2n}. \tag{10.3}$$

Making use of Stirling's formula, $n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$, we find,

$$\left(\frac{1}{2}\right)^{2n}\frac{(2n)!}{(n!)^2} \sim \left(\frac{1}{2}\right)^{2n}\frac{\sqrt{2\pi}(2n)^{2n+\frac{1}{2}}e^{-2n}}{2\pi n^{2n+1}e^{-2n}} = \sqrt{\frac{1}{\pi}}\frac{1}{\sqrt{n}} \tag{10.4}$$

and therefore,

$$\sum_{n=0}^{\infty}\mathbb{P}_0(X_n = 0) = \sum_{n=0}^{\infty}\mathbb{P}_0(X_{2n} = 0) \sim 1 + \sum_{n=1}^{\infty}\sqrt{\frac{1}{\pi}}\frac{1}{\sqrt{n}} = \infty$$

which shows again that this walk is recurrent. To now verify that this walk is null-recurrent it suffices to show there is not invariant probability distribution, π, for this walk. Such an invariant measure must satisfy,

$$\pi(x) = \frac{1}{2}\left[\pi(x+1) + \pi(x-1)\right]$$

which has general solution given by $\pi(x) = A + Bx$. In order for $\pi(x) \geq 0$ for all x we must take $B = 0$, i.e. π is constant. As $\sum_{x\in\mathbb{Z}} \pi(x) = A \cdot \infty$, there is no way to normalize π to become a probability distribution and hence $\{X_n\}_{n=0}^{\infty}$ is null-recurrent.

Fact 10.11 *Simple Random walk in $\mathbb{Z}^d$ is recurrent if $d = 1$ or 2 and is transient if $d \geq 3$. [For a informal proof of this fact, see page 49 of Lawler. For a formal proof see Section 10.2 below or Todd Kemp's lecture notes.]*

Example 10.12. The above method may easily be modified to show that the biased random walk on $\mathbb{Z}$ (see Exercise 6.9) is transient. In this case $\frac{1}{2} < p < 1$ and

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n = \binom{2n}{n} [p(1-p)]^n .$$

Since $p(1-p)$ has a maximum at $p = \frac{1}{2}$ of $\frac{1}{4}$ we have $\rho_p := 4p(1-p) < 1$ for $\frac{1}{2} < p < 1$. Therefore,

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} \left[\rho_p \frac{1}{4}\right]^n = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \rho_p^n \sim \sqrt{\frac{1}{\pi}} \frac{1}{\sqrt{n}} \rho_p^n .$$

Hence

$$\sum_{n=0}^{\infty} \mathbb{P}_0(X_n = 0) = \sum_{n=0}^{\infty} \mathbb{P}_0(X_{2n} = 0) \sim 1 + \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \rho_p^n \leq 1 + \frac{1}{1-\rho_p} < \infty$$

which again shows the biased random walk is transient.

Exercise 10.1. Let $\{X_n\}_{n=0}^{\infty}$ be the fair random walk on $\mathbb{Z}$ (as in Exercise 6.5) starting at 0 and let

$$A_N := \mathbb{E}\left[\sum_{k=0}^{2N} 1_{X_k=0}\right]$$

denote the expected number of visits to 0. Using Sterling's formula and integral approximations for $\sum_{n=1}^{N} \frac{1}{\sqrt{n}}$ to argue that $A_N \sim c\sqrt{N}$ for some constant $c > 0$.

10.2 *Transience and Recurrence for R.W.s by Fourier Series Methods

In the next result we will give another way to compute (or at least estimate) $\mathbb{E}_x M_y$ for random walks and thereby determine if the walk is transient or recurrent.

Theorem 10.13. *Let $\{\xi_i\}_{i=1}^{\infty}$ be i.i.d. random vectors with values in $\mathbb{Z}^d$ and for $x \in \mathbb{Z}^d$ let $X_n := x + \xi_1 + \cdots + \xi_n$ for all $n \geq 1$ with $X_0 = x$. As usual let $M_z := \sum_{n=0}^{\infty} 1_{X_{n=z}}$ denote the number of visits to $z \in \mathbb{Z}^d$. Then*

$$\mathbb{E}M_z = \lim_{\alpha\uparrow 1} \left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \frac{e^{i(x-z)\cdot\theta}}{1 - \alpha\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right]} \mathbf{d}\theta$$

where $\mathbf{d}\theta = d\theta_1 \ldots d\theta_d$ and in particular,

$$\mathbb{E}M_x = \lim_{\alpha\uparrow 1} \left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \frac{1}{1 - \alpha\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right]} \mathbf{d}\theta. \tag{10.5}$$

Proof. For $0 < \alpha \leq 1$ let

$$M_y^{(\alpha)} := \sum_{n=0}^{\infty} \alpha^n 1_{X_{n=y}}$$

so that $M_y^{(1)} = M_y$. Given any $\theta \in [-\pi,\pi]^d$ we have,

$$\begin{aligned}
\sum_{y\in\mathbb{Z}^d} \mathbb{E}M_y^{(\alpha)} e^{iy\cdot\theta} &= \mathbb{E}\sum_{y\in\mathbb{Z}^d} M_y^{(\alpha)} e^{iy\cdot\theta} = \mathbb{E}\sum_{y\in\mathbb{Z}^d}\sum_{n=0}^{\infty} \alpha^n 1_{X_{n=y}} e^{iy\cdot\theta} \\
&= \mathbb{E}\sum_{n=0}^{\infty} \alpha^n \sum_{y\in\mathbb{Z}^d} 1_{X_{n=y}} e^{iy\cdot\theta} = \mathbb{E}\sum_{n=0}^{\infty} \alpha^n e^{i\theta\cdot X_n} \\
&= \sum_{n=0}^{\infty} \alpha^n \mathbb{E}\left[e^{i\theta\cdot X_n}\right]
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta\cdot X_n}\right] &= e^{i\theta\cdot x}\mathbb{E}\left[e^{i\theta\cdot(\xi_1+\cdots+\xi_n)}\right] = e^{i\theta\cdot x}\mathbb{E}\left[\prod_{j=1}^{n} e^{i\theta\cdot\xi_j}\right] \\
&= e^{i\theta\cdot x}\left[\prod_{j=1}^{n} \mathbb{E}e^{i\theta\cdot\xi_j}\right] = e^{i\theta\cdot x}\cdot\left(\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right]\right)^n .
\end{aligned}$$

Combining the last two equations shows,

$$\sum_{y\in\mathbb{Z}^d} \mathbb{E}M_y^{(\alpha)} e^{iy\cdot\theta} = \frac{e^{i\theta\cdot x}}{1 - \alpha\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right]}.$$

Multiplying this equation by $e^{-iz\cdot\theta}$ for some $z \in \mathbb{Z}^d$ we find, using the orthogonality of $\left\{e^{iy\cdot\theta}\right\}_{y\in\mathbb{Z}^d}$, that

$$\begin{aligned}\mathbb{E}M_z^{(\alpha)} &= \left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \mathbf{d}\theta \left(\sum_{y\in\mathbb{Z}^d} \mathbb{E}M_y^{(\alpha)} e^{iy\cdot\theta} e^{-iz\cdot\theta}\right) \\ &= \left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \frac{e^{i\theta\cdot(x-z)}}{1-\alpha\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right]}\mathbf{d}\theta.\end{aligned}$$

Since $M_z^{(\alpha)} \uparrow M_z$ as $\alpha \uparrow 1$ the result is now a consequence of the monotone convergence theorem. ■

Example 10.14. Suppose that $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = -1) = q := 1-p$ and $x = 0$. Then

$$\begin{aligned}\mathbb{E}\left[e^{i\xi_1\cdot\theta}\right] &= pe^{i\theta} + qe^{-i\theta} = p(\cos\theta + i\sin\theta) + q(\cos\theta - i\sin\theta) \\ &= \cos\theta + i(p-q)\sin\theta.\end{aligned}$$

Therefore according to Eq. (10.5) we have,

$$\mathbb{E}_0 M_0 = \lim_{\alpha\uparrow 1} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{1-\alpha(\cos\theta + i(p-q)\sin\theta)} d\theta. \tag{10.6}$$

(We could compute the integral in Eq. (10.6) exactly using complex contour integral methods but I will not do this here.)

The integrand in Eq. (10.6) may be written as,

$$\frac{1-\alpha\cos\theta + i\alpha(p-q)\sin\theta}{(1-\alpha\cos\theta)^2 + \alpha^2(p-q)^2\sin^2\theta}.$$

As $\sin\theta$ is odd while the denominator is now even we find,

$$\mathbb{E}_0 M_0 = \lim_{\alpha\uparrow 1} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-\alpha\cos\theta}{(1-\alpha\cos\theta)^2 + \alpha^2(p-q)^2\sin^2\theta} d\theta. \tag{10.7}$$

a Let first suppose that $p = \frac{1}{2} = q$ in which case the above equation reduces to

$$\mathbb{E}_0 M_0 = \lim_{\alpha\uparrow 1} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{1-\alpha\cos\theta} d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{1-\cos\theta} d\theta$$

wherein we have used the MCT (for $\theta \sim 0$) and DCT (for θ away from 0) to justify passing the limit inside of the integral. Since $1-\cos\theta \approx \theta^2/2$ for θ near zero and $\int_{-\varepsilon}^{\varepsilon} \frac{1}{\theta^2} d\theta = \infty$ it follows that $\mathbb{E}_0 M_0 = \infty$ and the fair random walk on $\mathbb{Z}$ is recurrent.

b Now suppose that $p \neq \frac{1}{2}$ and let us write $\alpha := 1-\varepsilon$ for some ε which we will eventually let tend down to zero. With this notation the integrand $(f_\alpha(\theta))$ in Eq. (10.7) satisfies,

$$\begin{aligned}f_\alpha(\theta) &= \frac{1-\cos\theta + \varepsilon\cos\theta}{(1-\cos\theta+\varepsilon\cos\theta)^2 + (1-\varepsilon)^2(p-q)^2\sin^2\theta} \\ &= \frac{1-\cos\theta}{(1-\cos\theta+\varepsilon\cos\theta)^2 + (1-\varepsilon)^2(p-q)^2\sin^2\theta} \\ &+ \frac{\varepsilon\cos\theta}{(1-\cos\theta+\varepsilon\cos\theta)^2 + (1-\varepsilon)^2(p-q)^2\sin^2\theta} \\ &\le \frac{1-\cos\theta}{(1-\varepsilon)^2(p-q)^2\sin^2\theta} + \frac{\varepsilon\cos\theta}{\varepsilon^2\cos^2\theta + (1-\varepsilon)^2(p-q)^2\sin^2\theta}.\end{aligned}$$

The first term is bounded in θ and ε because

$$\lim_{\theta\downarrow 0} \frac{1-\cos\theta}{\sin^2\theta} = \frac{1}{2}$$

and therefore only makes a finite contribution to the integral. Integrating the second term near zero and making the change of variables $u = \sin\theta$ (so $du = \cos\theta d\theta$) shows,

$$\begin{aligned}&\int_{-\delta}^{\delta} \frac{\varepsilon\cos\theta}{\varepsilon^2\cos^2\theta + (1-\varepsilon)^2(p-q)^2\sin^2\theta} d\theta \\ &= \int_{-\sin(\delta)}^{\sin(\delta)} \frac{\varepsilon}{\varepsilon^2(1-u^2) + (1-\varepsilon)^2(p-q)^2u^2} du \\ &\le \int_{-\sin(\delta)}^{\sin(\delta)} \frac{\varepsilon}{\varepsilon^2\frac{1}{2} + \frac{1}{2}(p-q)^2u^2} du \\ &= 2\int_{-\sin(\delta)}^{\sin(\delta)} \frac{\varepsilon}{\varepsilon^2 + (p-q)^2u^2} du\end{aligned}$$

provided δ is sufficiently small but fixed and ε is small. Lastly we make the change of variables $u = \varepsilon x/|p-q|$ in order to find

$$\begin{aligned}&\int_{-\delta}^{\delta} \frac{\varepsilon\cos\theta}{\varepsilon^2\cos^2\theta + (1-\varepsilon)^2(p-q)^2\sin^2\theta} d\theta \\ &\le \frac{4}{|p-q|}\int_0^{\sin(\delta)|p-q|/\varepsilon} \frac{1}{1+x^2} du \uparrow \frac{2\pi}{|p-q|} < \infty \text{ as } \varepsilon \downarrow 0.\end{aligned}$$

Combining these estimates shows the limit in Eq. (10.7) is now finite so the random walk is transient when $p \neq \frac{1}{2}$.

Example 10.15 (Unbiased R.W. in $\mathbb{Z}^d$). Now suppose that $\mathbb{P}(\xi_i = \pm e_j) = \frac{1}{2d}$ for $j = 1, 2, \ldots, d$ and $X_n = \xi_1 + \cdots + \xi_n$. In this case,

$$\mathbb{E}\left[e^{i\theta\cdot\xi_j}\right] = \frac{1}{d}\left[\cos(\theta_1) + \cdots + \cos(\theta_d)\right]$$

and so according to Eq. (10.5) we find (as before)

$$\begin{aligned}\mathbb{E}M_0 &= \lim_{\alpha\uparrow 1}\left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \frac{1}{1-\alpha\frac{1}{d}\left[\cos(\theta_1)+\cdots+\cos(\theta_d)\right]}\mathbf{d}\theta \\ &= \left(\frac{1}{2\pi}\right)^d \int_{[-\pi,\pi]^d} \frac{1}{1-\frac{1}{d}\left[\cos(\theta_1)+\cdots+\cos(\theta_d)\right]}\mathbf{d}\theta \qquad \text{(by MCT and DCT).}\end{aligned}$$

Again the integrand is singular near $\theta = 0$ where

$$1-\frac{1}{d}\left[\cos(\theta_1)+\cdots+\cos(\theta_d)\right] \cong 1-\frac{1}{d}\left[d-\frac{1}{2}\|\theta\|^2\right] = \frac{1}{2}\|\theta\|^2 .$$

Hence it follows that $\mathbb{E}M_0 < \infty$ iff $\int_{\|\theta\|\le R}\frac{1}{\|\theta\|^2}\mathbf{d}\theta < \infty$ for $R < \infty$. The last integral is well know to be finite iff $d \ge 3$ as can be seen by computing in polar coordinates. For example when $d = 2$ we have

$$\int_{\|\theta\|\le R}\frac{1}{\|\theta\|^2}\mathbf{d}\theta = 2\pi\int_0^R \frac{1}{r^2}rdr = 2\pi\ln r|_0^R = 2\pi(\ln R - \ln 0) = \infty$$

while when $d = 3$,

$$\int_{\|\theta\|\le R}\frac{1}{\|\theta\|^2}\mathbf{d}\theta = 4\pi\int_0^R \frac{1}{r^2}r^2dr = 4\pi R < \infty.$$

In this way we have shown that the unbiased random walk in $\mathbb{Z}$ and $\mathbb{Z}^2$ is recurrent while it is transient in $\mathbb{Z}^d$ for $d \ge 3$.

Detail Balance and MCMC

In this chapter, let $\pi : S \to (0,1)$ be a positive probability on a (*large and complicated*) finite set S. One important application of Markov chains is sampling (approximately) from such probability distributions even when π itself is computationally intractable. In order to "sample" from π we first try to find an irreducible Markov matrix $\mathbf{P}$ on S such that $\pi\mathbf{P} = \pi$. Then from Theorem 8.27 in the form Remark 8.30, if $\{X_n\}_{n=0}^{\infty}$ is the Markov chain associated to $\mathbf{P}$ we have,

$$\pi(f) := \sum_{x\in S} f(x)\,\pi(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} f(X_m) \quad \mathbb{P}_\nu - \text{a.s.},$$

Thus we expect that if we simulate, $(\{X_n\}_{n=0}^{\infty})$, we will have,

$$\pi(f) \cong \frac{1}{n}\sum_{m=1}^{n} f(X_m) \text{ for large } n,$$

[Using Remark 9.4, the error in the above approximation is roughly $C\frac{1}{\sqrt{n}}$ where $C = C(\omega)$ is a random constant which is likely to be impossible to estimate. We will come back to error estimates in the L^2 – sense below in Corollary 11.38.] The key to implementing this method is to construct $\mathbf{P}$ in such a way that the associated chain is relatively easy to simulate.

Remark 11.1 (A useless choice for $\mathbf{P}$*).* Suppose $\pi : S \to (0,1)$ is a distribution we would like to sample as the stationary distribution of a Markov chain. One way to construct such a chain would be to make the Markov matrix $\mathbf{P}$ with all rows given by π. In this case we will have $\nu\mathbf{P} = \pi$ and so $\mathbb{P}_\nu(X_1 = x) = \pi(x)$ for all $x \in S$. Thus for this chain we would reach the invariant distribution in one step. **However**, to simulate this chain is equivalent to sampling from π which is what we were trying to figure out how to do in the first place!! We need a simpler Markov matrix, $\mathbf{P}$, with $\pi\mathbf{P} = \pi$.

11.1 Detail Balance

The next Lemma gives a useful criteria on a Markov matrix $\mathbf{P}$ so that π is an invariant (=stationary) distribution of $\mathbf{P}$.

Lemma 11.2 (Detail balance). *In general, if we can find a distribution, π, satisfying the* ***detail balance equation,***

$$\pi_i \mathbf{P}_{ij} = \pi_j \mathbf{P}_{ji} \text{ for all } i \neq j, \tag{11.1}$$

then π is a stationary (=invariant) distribution, i.e. $\pi\mathbf{P} = \pi$.

Proof. First proof. Intuitively, Eq. (11.1) states that sites i and j are always exchanging sand back and forth at equal rates. Hence if all sites are doing this the size of the piles of sand at each site must remain unchanged.

Second Proof. Summing Eq. (11.1) on i making use of the fact that $\sum_i \mathbf{P}_{ji} = \mathbf{1}$ for all j implies, $\sum_i \pi_i \mathbf{P}_{ij} = \pi_j$. ■

Example 11.3 (Simple Random Walks on Graphs). If S is a finite set of nodes and G be an undirected graph on S, i.e. G is a subset of $S \times S$ such that

1. $(x,x) \notin G$ for all $x \in S$,
2. if $(x,y) \in G$, then $(y,x) \in G$ [the graph is undirected], and
3. for all $x \in G$, the set $S_x := \{y \in S : (x,y) \in G\}$ is not empty. [We are not allowing for any isolated nodes in our graph.]

Let us write $x \sim y$ to mean $(x,y) \in G$ and let

$$\nu(x) := \#(S_x) = \sum_{y\in S} 1_{(x,y)\in G}$$

be the **valence** of G at x. The **random walk on this graph** is then the Markov chain on S with Markov transition matrix,

$$p(x,y) := \frac{1}{\nu(x)} 1_{S_x}(y) = \frac{1}{\nu(x)} 1_{(x,y)\in G}.$$

Notice that

$$\nu(x)\,p(x,y) = 1_{x\sim y} = \nu(y)\,p(y,x)$$

so that $[\nu, \mathbf{P}]$ satisfy the detail balance equation. Therefore if

$$Z := \sum_{x\in S} \nu(x) \text{ and } \pi(x) := \frac{1}{Z}\nu(x),$$

then π is an invariant distribution for $\mathbf{P}$.

Remark 11.4. If $(\mathbf{P},\pi)$ are in detail balance and $\pi : S \to (0,1)$ is a strictly positive, then $\mathbf{P}_{xy} \neq 0$ iff $\mathbf{P}_{yx} \neq 0$. Thus,

$$G := \{(x,y) \in S \times S : x \neq y \text{ and } \mathbf{P}_{xy} \neq 0\}$$

defines an undirected graph on S. Moreover it is easy to check that $\mathbf{P}$ is irreducible iff G is connected.

Example 11.5. Capitalizing on Example 11.3, consider the following problem. *A Knight starts at a corner of a standard* 8×8 *chessboard, and on each step moves at random. How long, on average, does it take to return to its starting position?*

$$\begin{bmatrix} \cdot & * & \cdot & * & \cdot & * & \cdot & * \\ * & \cdot & * & \cdot & * & \cdot & * & \cdot \\ \cdot & * & \cdot & * & \cdot & * & \cdot & * \\ * & \cdot & * & \cdot & * & \cdot & * & \cdot \\ \cdot & * & \cdot & * & \cdot & * & \cdot & * \\ * & \cdot & * & \cdot & * & \cdot & * & \cdot \\ \cdot & * & \cdot & * & \cdot & * & \cdot & * \\ * & \cdot & * & \cdot & * & \cdot & * & \cdot \end{bmatrix}$$

If i is a corner, the question is what is $\mathbb{E}_i[R_i]$, where the Markov process in question is the one described in the problem: it is a simple random walk (SRW) on a graph whose vertices are the 64 squares of the chessboard, and two squares are connected in the graph if a Knight can move from one to the other. (Note: if the Knight can move from i to j, it can do so in exactly two ways: 1 up/down and 2 left/right, or 2 left/right and 1 up/down.[1] So choosing uniformly among *moves* or among *positions* amounts to the same thing.) This graph is connected (as a little thought will show), and so the SRW is irreducible; therefore there is a unique stationary distribution. By Example 11.3, the stationary distribution π is given by $\pi(i) = v_i / \sum_j v_j$, and so by the Ergodic Theorem 8.27,

$$\mathbb{E}_i[R_i] = \frac{1}{\pi(i)} = \frac{1}{\nu_i} \sum_j v_j .$$

A Knight moves 2 units in one direction and 1 unit in the other direction. Starting at a corner $(1,1)$, the Knight can only more to the positions $(3,2)$ and $(2,3)$, so $v_i = 2$. To solve the problem, we need to calculate v_j for all starting positions j on the board. By square symmetry, we need only calculate the numbers for the upper 4×4 grid. An easy calculation then gives these valences as

$$\begin{bmatrix} 2 & 3 & 4 & 4 \\ 3 & 4 & 6 & 6 \\ 4 & 6 & 8 & 8 \\ 4 & 6 & 8 & 8 \end{bmatrix}$$

The sum of all the valences of the 4×4 grid is 84, and so the sum over the full chess board is $4 \cdot 84 = 336$. Thus, the number of expected steps for the Knight to return to the corner is

$$\mathbb{E}_i[R_i] = \frac{1}{\nu_i} \sum_j v_j = \frac{1}{2} 336 = 168.$$

11.2 Some uniform measure MCMC examples

Let us continue the notation in Example 11.3, i.e. suppose S is a finite set and G is a connected unoriented graph on S with no isolated points. The simple random walk on G would allow us to sample from the distribution, $\pi(x) = v_x / \sum_{y \in S} v_y$. This is already useful as finding the $Z := \sum_{y \in S} v_y$ may be intractable and moreover for large S, Z will be very large causing problems of its own. In this section we are going to be interested in sampling the uniform distribution, $\pi(x) = Z^{-1}$ on S, where now $Z := \#(S)$. The difficulty is going to be that the set S is typically very large complicated so that $Z := \#(S)$ is large and perhaps not even known how to compute. The next example shows how one might sample the uniform distribution in this setting.

Example 11.6 (Uniform Samplers). Keeping the notation above, if $\nu_x = K$ is constant for all $x \in S$ then the SRW has the uniform distribution as its stationary distribution. What if the valences are not constant but we at least know that there exists $K < \infty$ such that $v_x \leq K$ for all $x \in S$. In this case we can take

$$p(x,y) := \frac{1}{K} 1_{\langle x,y \rangle \in G} + c(x) 1_{y=x}$$

where $c(x)$ is chosen so that

$$1 = \sum_{y \in S} p(x,y) = \frac{1}{K} \nu(x) + c(x) \implies c(x) := 1 - \frac{1}{K} \nu(x) \geq 0.$$

Then clearly $p(x,y)$ is a symmetric irreducible Markov matrix and therefore the invariant distribution of $\mathbf{P}$ is $\pi(x) = \frac{1}{\#(S)}$. Thus if $\{X_n\}_{n=0}^{\infty}$ is the Markov chain associated to $\mathbf{P}$, we will have

$$\frac{1}{|S|} \sum_{x \in S} f(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) \qquad \mathbb{P}_\nu \text{ -a.s.}$$

for any starting distribution ν.

[1] A little experimentation will show that this graph is connected so the random walk is irreducible.

Example 11.7. Suppose that S is the set of $D \times D$ - matrices with entries in $\{0,1\}$ so that $\#(S) = 2^{25\times 25} \cong 1.4\times 10^{188}$! We say that $\langle A, B\rangle \in G$ if $A, B \in S$ and A and B differ at exactly one location, i.e. $(A+B) \bmod 2$ is a matrix in S with exactly one non-zero entry. In this case $\nu_A = D^2$ and the simple random-walk on S will sample from the uniform distribution. The associated Markov matrix is

$$p(A,B) = \frac{1}{D^2} 1_{A\sim B}.$$

It is relatively easily to sample from this chain. To describe how to do this, for $i,j \in \{1,\dots,D\}$, let $E(i,j) \in S$ denote the matrix with 1 at the $(i,j)^{\text{th}}$ - position and zeros at all other positions. To simulate the associated Markov chain, $\{X_n\}_{n=0}^{\infty}$, at each n, choose $i,j \in \{1,\dots,D\}$, uniformly at random so that i is independent of j which are independent of the previous choices made. Then if $X_n = A \in S$ we take $X_{n+1} = (A + E(i,j))$.

The next example gives a variant of this example.

Example 11.8 (HCC). Let us continue the notation in Example 11.7 and further let S_{HC} denote those $A \in S$ such there are no adjacent **1**'s in the matrix A. To be more precise we are requiring for all $i,j \in \{1,\dots,D\}$ that

$$A_{ij}\cdot(A_{i-1,j} + A_{i+1,j} + A_{i,j-1} + A_{i,j+1}) = 0$$

where by convention $A_{i,j} = 0$ if either i or j is in $\{0, D+1\}$. We refer to S_{HC} as the **Hard Core Configuration space.** This is a simple model of the configuration of an ideal gas (the 1s represent gas molecules, the 0s empty space). We would like to pick a random hard-core configuration space – meaning sample from the uniform distribution on S_{HC}, i.e. each $A \in S_{HC}$ is to be chosen with weight $1/|S_{HC}|$. The problem is that $|S_{HC}|$ is unknown, even to exponential order, for large D. A simple upper-bound is $2^{D^2} = \#(S)$. It is conjectured that $|S_{HC}| \sim \beta^{D^2}$ for some $\beta \in (1,2)$, but no one even has a good guess as to what β is. Nevertheless, we can use the Markov chain associated to Markov matrix,

$$p_{HC}(A,B) = \frac{1}{D^2} 1_{A\sim B} + c(A)\, 1_{A=B}$$

in order to sample from this distribution. Proposition 11.10 below will justify the following algorithm for sampling from this chain.

Sampling Algorithm. Suppose we generated $(X_0, X_1, \dots, X_n)$ and $X_n = A \in S_{HC}$. To construct X_{n+1};

1. Choose $i,j \in \{1,\dots,D\}$, uniformly at random so that i is independent of j which are independent of the previous choices made.
2. Let $B := (A + E(i,j)) \bmod 2$ and set
$$X_{n+1} = \begin{cases} B & \text{if} \quad B \in S_{HC} \\ A & \text{otherwise.} \end{cases}$$
Note that $B \in S_{HC}$ iff
$$B_{ij}\cdot(B_{i-1,j} + B_{i+1,j} + B_{i,j-1} + B_{i,j+1}) = 0$$
where again $B_{kl} = 0$ if either k or l is in $\{0, D+1\}$.
3. loop on n.

You are asked to implement the algorithm in Example 11.8 for $D = 50$ in one of your homework problems and use your simulations to estimate the probability that $A_{25,25} = 1$ when A is chosen uniformly at random from S_{HC}.

Lemma 11.9 (Conditional detail balance). *Suppose that $(\pi, \mathbf{P})$ satisfies detail balance and $\hat{S}$ is a subset of S. Let*

$$\hat{\pi}(x) = \pi\left(x|\hat{S}\right) = \frac{\pi(x)}{\sum_{y\in\hat{S}}\pi(y)} \text{ for } x \in \hat{S}$$

be the conditional distribution of π given $\hat{S}$. For $x, y \in \hat{S}$ let,

$$\hat{p}(x,y) := \begin{cases} p(x,y) & \text{if } x \neq y \\ c(x) & \text{if } x = y \end{cases}$$

where

$$c(x) := 1 - \sum_{y\in\hat{S}\setminus\{x\}} p(x,y).$$

Then $\hat{\mathbf{P}}$ is a Markov matrix on $\hat{S}$ and $\left(\hat{\pi}, \hat{\mathbf{P}}\right)$ are in detail balance. In particular $\hat{\pi}$ is an invariant distribution for $\hat{\mathbf{P}}$.

Proof. First off,

$$\sum_{y\in\hat{S}} \hat{p}(x,y) = \sum_{y\in\hat{S}\setminus\{x\}} p(x,y) + \left(1 - \sum_{y\in\hat{S}\setminus\{x\}} p(x,y)\right) = 1$$

so that $\hat{p}$ is a Markov matrix. Moreover, if we let $Z := \sum_{y\in\hat{S}}\pi(y)$, then for $x \neq y$ in $\hat{S}$ we have,

$$\hat{\pi}(x)\,\hat{p}(x,y) = \frac{1}{Z}\pi(x)\,p(x,y) = \frac{1}{Z}\pi(y)\,p(y,x) = \hat{\pi}(y)\,\hat{p}(y,x)$$

which proves the detail balance condition. ■

Proposition 11.10 (Conditional sampling). *Let us continue the notation in Lemma 11.9 and further suppose that $\{f_n : S \to S\}_{n=0}^{\infty}$ are random functions as in Theorem 5.11 so that $X_{n+1} = f_n(X_n)$ generates the Markov chain associated to $\mathbf{P}$.If we let $\hat{f}_n : \hat{S} \to \hat{S}$ be defined by*

$$\hat{f}_n(x) = \begin{cases} f_n(x) & \text{if} \quad f_n(x) \in \hat{S} \\ x & \text{otherwise} \end{cases}$$

the $\left\{\hat{f}_n\right\}_{n=0}^{\infty}$ are independent random function which generate (in the sense that $\hat{X}_{n+1} = \hat{f}_n\left(\hat{X}_n\right)$) the Markov chain, $\left\{\hat{X}_n\right\}_{n=0}^{\infty}$, associated to $\hat{\mathbf{P}}$.

Proof. We need only note that if $x, y \in \hat{S}$ with $x \neq y$, then

$$\begin{aligned} \mathbb{P}\left(\hat{f}_n(x) = y\right) &= \mathbb{P}\left(f_n(x) = y, f_n(x) \in \hat{S}\right) + \mathbb{P}\left(x = y, f_n(x) \notin \hat{S}\right) \\ &= \mathbb{P}\left(f_n(x) = y\right) + 0 = p(x, y). \end{aligned}$$

It then must follows that

$$\begin{aligned} \mathbb{P}\left(\hat{f}_n(x) = x\right) &= 1 - \mathbb{P}\left(\hat{f}_n(x) \in \hat{S} \setminus \{x\}\right) \\ &= 1 - \sum_{y \in \hat{S} \setminus \{x\}} \mathbb{P}\left(\hat{f}_n(x) = y\right) \\ &= 1 - \sum_{y \in \hat{S} \setminus \{x\}} p(x, y) = c(x). \end{aligned}$$

■

Example 11.11. Let T be a finite set and S be the set of functions, $s : T \to \{0, 1\}$ so that $\#(S) = 2^{\#(T)}$. We say $(s, s') \in G$ iff s and s' differ at one location $t \in T$, i.e. $\sum_t |s(t) - s'(t)| = 1$. (We abbreviate this condition by writing $s \sim s'$.) In this case $\nu_s = \#(T)$. To construct a random walk on T, let $\{U_n\}_{n=1}^{\infty}$ be i.i.d. $T-$valued uniformly distributed random functions. We then construct the random walk starting at $s \in S$ by $X_0 = s$, $X_{n+1} = \left(X_n + \delta_{U_{n+1}}\right) \bmod 2$ for $n \geq 0$. Note that

$$p(s, s') = \begin{cases} \frac{1}{\#(T)} & \text{if } s' \sim s \\ 0 & \text{otherwise} \end{cases}$$

Since $p(s, s') = p(s', s)$, it follows that p has the uniform distribution as its invariant measure. If we now take some subset $\hat{S} \subset S$ we can sample from the "restricted chain" $\left\{\hat{X}_n\right\}$ on $\hat{S}$ using

$$\hat{X}_{n+1} = \begin{cases} \hat{X}_n + \delta_{U_{n+1}} & \text{if } \hat{X}_n + \delta_{U_{n+1}} \in \hat{S} \\ \hat{X}_n & \text{otherwise.} \end{cases}$$

Example 11.12. Consider the problem of **graph colorings**. Let $G = (V, E)$ be a finite graph. A q**-coloring** of G (with $q \in \mathbb{N}$) is a function $f : V \to \{1, 2, \ldots, q\}$ with the property that, if $u, v \in V$ with $u \sim v$, then $f(u) \neq f(v)$. The set of q-colorings is very hard to count (especially for small q), so again we cannot directly sample a uniformly random q-coloring. Instead, define a Markov chain on the state space of all q-colorings f, where for $f \neq g$

$$p(f, g) = \begin{cases} \frac{1}{q|V|}, & \text{if } f \text{ and } g \text{ differ at exactly one vertex,} \\ 0, & \text{otherwise.} \end{cases}$$

Again: since there are at most q different q-colorings of G that differ at a given vertex, and there are $|V|$ vertices, we have $\sum_{f \neq g} p(f, g) \leq 1$, and so setting $p(f, f) = \sum_{g \neq f} p(f, g)$ yields a stochastic matrix p which is the transition kernel of a Markov chain. It is evidently symmetric, and by considerations like those in Example 11.8 the chain is irreducible and aperiodic (so long as q is large enough for any q-colorings to exist!); hence, the stationary distribution is uniform, and this Markov chain converges to the uniform distribution.

To simulate the corresponding Markov chain: given $X_n = f$, choose one of the $|V|$ vertices, v, uniformly at random and one of the q colors, k, uniformly at random. If the new function g defined by $g(v) = k$ and $g(w) = f(w)$ for $w \neq v$ is a q-coloring of the graph, set $X_{n+1} = g$; otherwise, keep $X_{n+1} = X_n = f$. This process has the transition probabilities listed above, and so running it for large n gives an approximately uniformly random q-coloring.

Exercise 11.1. Let S be a finite set and $\mathbf{P}$ be an irreducible Markov matrix with invariant distribution, π, which satisfies detail balance. Further let $A := \hat{S}$ be a subset of S, $B = S \setminus A$, and $\hat{\pi}(x) = \pi(x) / \pi\left(\hat{S}\right)$ as above. If $\{X_n\}_{n=0}^{\infty}$ is the Markov chain associated to $\mathbf{P}$ which we will start in $\hat{S}$, let $\left\{\hat{X}_k := X_{\tau_k}\right\}_{k=0}^{\infty}$ where $\{\tau_k\}_{k=0}^{\infty}$ are the stopping times defined by, $\tau_0 = 0$, and then inductively by

$$\tau_{k+1} := \min\left\{n > \tau_k : X_n \in \hat{S}\right\} \text{ for } k \in \mathbb{N}_0.$$

Show $\left\{\hat{X}_k\right\}_{k=0}^{\infty}$ is a Markov chain on $\hat{S}$ with corresponding Markov matrix,

$$\hat{\mathbf{P}} = \mathbf{P}_{A \times A} + \mathbf{P}_{A \times B}\left(I - \mathbf{P}_{B \times B}\right)^{-1} \mathbf{P}_{B \times A}$$

and $\hat{\pi}$ is in detail balance with $\hat{\mathbf{P}}$. In particular it follows that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f\left(\hat{X}_k\right) = \hat{\pi}(f) \ \ \mathbb{P}_\nu - \text{a.s.}$$

for any distribution ν on $\hat{S}$. [This algorithm will likely run very slowly, especially when $\hat{S}$ is a rather "thin" subset of S since then the chain $\{X_n\}$ will spend a lot of time outside of $\hat{S}$.]

11.3 The Metropolis-Hastings Algorithm

In Examples 11.7, 11.8, and 11.12, we constructed Markov chains with symmetric kernels, therefore having the uniform distribution as stationary, and converging to it. This is often a good way to simulate uniform samples. But how can we sample from a non-uniform distribution?

Proposition 11.13. *Let S be a finite set, and let π be a (strictly positive) distribution on S and suppose that $\mathbf{Q}_{ij} = q(i,j)$ is a symmetric ($\mathbf{Q}_{ij} = \mathbf{Q}_{ji}$ irreducible Markov matrix on S. [Note that $\mathbf{Q}$ invariant distribution is the uniform distribution on S.] Then define a new Markov matrix $\mathbf{P}_{ij} = p(i,j)$ by,*

$$p(i,j) := q(i,j)\min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} \text{ for } i \neq j, \text{ and}$$
$$p(i,i) = 1 - \sum_{j\neq i} p(i,j).$$

Then $\mathbf{P}$ is an irreducible Markov matrix with stationary distribution, π. In fact, $(\pi, \mathbf{P})$ are in detail balance.

Proof. By construction we have $p(i,j) \in [0,1]$, $p(i,j) = 0$ iff $q(i,j) = 0$, and

$$\pi(i)p(i,j) = q(i,j)\min\{\pi(i),\pi(j)\} = q(j,i)\min\{\pi(j),\pi(i)\} = \pi(j)p(j,i)$$

so π satisfies the detailed balance condition for p. Consequently, π is the stationary distribution for p. Moreover, it follows that p is irreducible (since q is). ■

Thus if $\{X_n\}_{n=0}^{\infty}$ is the Markov chain associated to $\mathbf{P}$ we will have and therefore

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n} f(X_m) = \pi(f) \ \ \mathbb{P}_\nu - \text{a.s.}$$

In order to simulate from this chain we need to explain an effective way to construct a random function $\hat{f} : S \to S$ so that $\mathbb{P}\left(\hat{f}(i) = j\right) = p(i,j)$. We assume that q is simple enough that we have to difficulty simulating a random function, $f : S \to S$ so that $\mathbb{P}(f(i) = j) = q(i,j)$.

Lemma 11.14. *If $f : S \to S$ is a random function such that $\mathbb{P}(f(i) = j) = q(i,j)$ for all $i,j \in S$, $U \in [0,1]$ is an independent uniform random variable, and*

$$\hat{f}(i) := \begin{cases} f(i) & \text{if } U \le \frac{\pi(f(i))}{\pi(i)} \\ i & \text{if } U > \frac{\pi(f(i))}{\pi(i)} \end{cases},$$

then

$$\mathbb{P}\left(\hat{f}(i) = j\right) = p(i,j) \text{ for all } i,j \in S.$$

Proof. For $j \neq i$ we have

$$\mathbb{P}\left(\hat{f}(i) = j\right) = \mathbb{P}\left(f(i) = j, U \le \frac{\pi(j)}{\pi(i)}\right) = \mathbb{P}(f(i) = j)\,\mathbb{P}\left(U \le \frac{\pi(j)}{\pi(i)}\right)$$
$$= q(i,j)\min\left(\frac{\pi(j)}{\pi(i)}, 1\right) = p(i,j).$$

and this suffices to complete the proof. ■

With the above constructions in hand we can now explain how to simulate the chain $\{X_n\}$.

Metropolis-Hastings algorithm. Given $X_n = i$ we update to X_{n+1} using the following steps.

1. Spin the $q(i,\cdot)$ – spinner (i.e. evaluate $f(i)$), if it lands at i ($f(i) = i$) stay at i,
2. if it lands at $j \neq i$ ($f(i) = j \neq i$), then flip a biased coin with probability of heads being $\min\left(\frac{\pi(j)}{\pi(i)}, 1\right)$.
   a) If the coin flip is heads then move to j, i.e. set $X_{n+1} = j$. [This will happen for sure if $\pi(j) \ge \pi(i)$.]
   b) If the coin flip is tails then stay at i, i.e. $X_{n+1} = i$.

There are many variations on the Metropolis-Hastings algorithm, all with the same basic ingredients and goal: to use a (collection of correlated) Markov chain(s), designed to have a given stationary distribution π, to sample from π approximately. They are collectively known as **Markov chain Monte Carlo** simulation methods. One place they are especially effective is in calculating high-dimensional integrals (i.e. expectations), a common problem in multivariate statistics.

11.4 The linear algebra associated to detail balance

In this section suppose that S is a finite set, $\nu : S \to (0,1]$ is a given (strictly positive) probability on S, and $p : S \times S \to [0,1]$ is a Markov matrix. We let $\ell^2(\nu)$ denote the vector space of functions, $f : S \to \mathbb{R}$ equipped with the inner product,

$$\langle f, g\rangle_\nu = \sum_{x\in S} f(x)\,g(x)\,\nu(x) \ \ \forall\ f,g \in \ell^2(\nu).$$

We further let $\mathbf{P} : \ell^2(\nu) \to \ell^2(\nu)$ be defined by

$$\mathbf{P}f(x) := \sum_{y\in S} p(x,y)\,f(y).$$

Notation 11.15 *If* $\mathbf{P}$ *is a Markov matrix we let* $\sigma(\mathbf{P})$ *denote the set of eigenvalues of* $\mathbf{P}$.

We are now going to discuss the spectral theory of $\mathbf{P}$ – mostly in the case that $\mathbf{P}$ has in invariant distribution, π, which is in detail balance with $\mathbf{P}$. We start with a general lemma not requiring detail balance.

Lemma 11.16. *If* $\mathbf{P}$ *is a Markov matrix and* π *is an invariant distribution for* $\mathbf{P}$, *then* $\mathbf{P}$ *is a contraction on* $\ell^p(\pi)$ *for all* $1 \le p \le \infty$.

Proof. The case $f \in \ell^1(\pi)$ is all we really need so let us prove this special case first. The point is that

$$\begin{aligned}\sum_{x\in S}\pi(x)\left|\mathbf{P}f(x)\right| &= \sum_{x\in S}\pi(x)\left|\sum_{y\in S}\mathbf{P}_{x,y}f(y)\right| \\ &\le \sum_{x\in S}\pi(x)\sum_{y\in S}\mathbf{P}_{x,y}\left|f(y)\right| \\ &= \sum_{y\in S}\sum_{x\in S}\pi(x)\mathbf{P}_{x,y}\left|f(y)\right| = \sum_{y\in S}\pi(y)\left|f(y)\right|\end{aligned}$$

which is precisely the statement that $\|\mathbf{P}f\|_{\ell^1(\pi)} \le \|f\|_{\ell^1(\pi)}$.

The remaining results use Hölder's inequality. Let us be content here with the case $p=2$ where we only need the Cauchy-Schwarz inequality. If $f \in \ell^2(\pi)$, then

$$\begin{aligned}\left|\mathbf{P}f(x)\right| &\le \sum_{y\in S}\mathbf{P}_{x,y}\left|f(y)\right| \le \left(\sum_{y\in S}\mathbf{P}_{x,y}\left|f(y)\right|^2\right)^{1/2}\left(\sum_{y\in S}\mathbf{P}_{x,y}1^2\right)^{1/2} \\ &\le \left(\sum_{y\in S}\mathbf{P}_{x,y}\left|f(y)\right|^2\right)^{1/2}\cdot 1.\end{aligned}$$

Squaring this inequality, multiplying by $\pi(x)$, and then summing on x shows

$$\begin{aligned}\sum_{x\in S}\pi(x)\left|\mathbf{P}f\right|^2(x) &\le \sum_{x\in S}\pi(x)\sum_{y\in S}\mathbf{P}_{x,y}\left|f(y)\right|^2 \\ &= \sum_{y\in S}\sum_{x\in S}\pi(x)\mathbf{P}_{x,y}\left|f(y)\right|^2 \\ &= \sum_{y\in S}\pi(y)\left|f(y)\right|^2\end{aligned}$$

which is precisely the statement that $\|\mathbf{P}f\|_{\ell^2(\pi)} \le \|f\|_{\ell^2(\pi)}$. ■

Corollary 11.17. *If* $\mathbf{P}$ *is an irreducible Markov matrix with invariant distribution* π *(necessarily positive) then* $1 \in \sigma(\mathbf{P})$ *and if* $\lambda \in \sigma(\mathbf{P})$, *then* $|\lambda| \le 1$.

Proof. By very definition of a Markov matrix we know that $\mathbf{P}\mathbf{1} = \mathbf{1}$ where $\mathbf{1}$ is the function 1 on S and hence $1 \in \sigma(\mathbf{P})$. If $\lambda \in \sigma(\mathbf{P})$ and $f : S \to \mathbb{C}$ is a non-zero function such that $\mathbf{P}f = \lambda f$, then

$$|\lambda|\,\|f\|_{\ell^1(\pi)} = \|\lambda f\|_{\ell^1(\pi)} = \|\mathbf{P}f\|_{\ell^1(\pi)} \le \|f\|_{\ell^1(\pi)}$$

from which we conclude that $|\lambda| \le 1$. ■

The only statement we really need from the next three results is the part of Corollary 11.20 which states; if $(\pi, \mathbf{P})$ satisfy detail balance, then $\mathbf{P}$ is symmetric operator on $\ell^2(\pi)$.

Lemma 11.18 (Adjoint of P). *The adjoint of* $\mathbf{P}$ *(relative to* $\langle\cdot,\cdot\rangle_\nu$*) is given by*

$$(\mathbf{P}^*g)(x) = \sum_{y\in S} q_\nu(x,y)\,g(y) \quad \textit{where} \tag{11.2}$$

$$q_\nu(x,y) = \frac{\nu(y)}{\nu(x)}p(y,x). \tag{11.3}$$

Proof. This lemmas is proved as a consequence of the simple identities,

$$\begin{aligned}\langle \mathbf{P}f, g\rangle_\nu &= \sum_{x\in S}\nu(x)\,g(x)\sum_{y\in S}p(x,y)\,f(y) \\ &= \sum_{x,y\in S} g(x)\,\frac{\nu(x)}{\nu(y)}p(x,y)\,f(y)\,\nu(y) \\ &= \sum_{y\in S}\left[\sum_{x\in S} q_\nu(y,x)\,g(x)\right] f(y)\,\nu(y). \\ &= \langle f, \mathbf{P}^*g\rangle_\nu.\end{aligned}$$

■

Corollary 11.19. *Keeping the notation in Lemma 11.18. The function,* $q_\nu(x,y)$, *is Markov matrix iff* ν *is an invariant measure for* $\mathbf{P}$, *i.e.* $\nu\mathbf{P} = \nu$.

Proof. The matrix, q_ν, is Markov matrix iff for all $x \in S$,

$$1 = \sum_{y\in S} q_\nu(x,y) = \sum_{y\in S}\frac{\nu(y)}{\nu(x)}p(y,x) = \frac{1}{\nu(x)}\cdot(\nu\mathbf{P})(x)$$

from which the result immediately follows. ■

Corollary 11.20. *Keeping the notation in Lemma 11.18. We have* $\mathbf{P} = \mathbf{P}^*$ *iff the* ν *and* p *satisfy the detail balance equation.*

Proof. This is an immediate consequence of Eq. (11.3). For those that have skipped directly to this corollary let us check directly that if $(\nu, \mathbf{P})$ is in detail balance, then $\mathbf{P}$ is self-adjoint on $\ell^2(\nu)$. For $f, g \in \ell^2(\nu)$, we have

$$\begin{aligned}\langle \mathbf{P}f, g\rangle_\nu &:= \sum_{x\in S}\sum_{y\in S} \mathbf{P}_{xy} f(y)\, g(x)\, \nu(x) \\ &= \sum_{y\in S}\sum_{x\in S} f(y)\, \nu(x)\, \mathbf{P}_{xy} g(x) \\ &= \sum_{y\in S}\sum_{x\in S} f(y)\, \nu(y)\, \mathbf{P}_{yx} g(x) = \langle f, \mathbf{P}g\rangle_\nu .\end{aligned}$$

Second proof. Let $\mathbf{D}f(x) = \nu(x) f(x)$ for all $x \in S$. Then $\langle f, g\rangle_\nu = f \cdot \mathbf{D}g = f^{\mathrm{tr}}\mathbf{D}g$ where

$$f \cdot g = \sum_{x\in S} f(x)\, g(x)$$

is the usual dot product. We then find,

$$\begin{aligned}&\langle \mathbf{P}f, g\rangle_\nu = \mathbf{P}f \cdot \mathbf{D}g = f \cdot \mathbf{P}^{\mathrm{tr}}\mathbf{D}g \text{ and} \\ &\langle f, \mathbf{P}g\rangle_\nu = f \cdot \mathbf{D}\mathbf{P}g\end{aligned}$$

and so $\mathbf{P}^{*\nu} = \mathbf{P}$ iff $\mathbf{P}^{\mathrm{tr}}\mathbf{D} = \mathbf{D}\mathbf{P}$, i.e.

$$\nu(x)\, \mathbf{P}_{xy} = [\mathbf{D}\mathbf{P}]_{xy} = \left[\mathbf{P}^{\mathrm{tr}}\mathbf{D}\right]_{xy} = \mathbf{P}^{\mathrm{tr}}_{xy}\mathbf{D}_y = \mathbf{P}_{yx}\nu(y)$$

which is precisely the detail balance condition. ■

Because of this corollary, we may apply to the spectral theorem from linear algebra to arrive at the following important theorem.

Theorem 11.21 (Spectral Theorem). *If* $\mathbf{P}$ *is a Markov matrix and* $\pi : S \to (0,1]$ *is a probability distribution on* S *such that* $(\pi, \mathbf{P})$ *are in detail balance, then* $\mathbf{P}f(x) = \sum_{y\in S} p(x,y) f(y)$ *has an orthonormal (relative to the* $\ell^2(\pi)$ *– inner product) basis of eigenvectors,* $\{f_n\}_{n=1}^{|S|} \subset \ell^2(\pi)$ *with corresponding eigenvalues*

$$\sigma(\mathbf{P}) := \{\lambda_n\}_{n=1}^{|S|} \subset \mathbb{R}.$$

Advise to the reader: on your first reading of the proof of Theorem 11.22, stick to the (most important) case where $\mathbf{P}$ is aperiodic.

Theorem 11.22. *If* $(\pi, \mathbf{P})$ *are in detail balance, then* $\sigma(\mathbf{P}) \subset [-1, 1]$ *and* $1 \in \sigma(\mathbf{P})$. *We further have;*

1. *If* $\mathbf{P}$ *is irreducible and aperiodic then* 1 *has multiplicity one with* $\mathbf{P1} = \mathbf{1}$, *i.e.* $\mathbf{1}$ *is the unique column eigenvector of* $\mathbf{P}$ *modulo scaling. Moreover there exists* $\alpha < 1$ *so that* $\sigma(\mathbf{P}) \subset [-\alpha, \alpha] \cup \{1\}$.
2. *If* $\mathbf{P}$ *is only assumed to irreducible, then there exists* $\alpha < 1$ *such that* $\sigma(\mathbf{P}) \subset [-\alpha, \alpha] \cup \{\pm 1\}$ *and again* 1 *has multiplicity* 1 *and as usual* $\mathbf{P1} = \mathbf{1}$.

Proof. By Corollary 11.20, $\mathbf{P}$ is self-adjoint on $\ell^2(\pi)$ and therefore $\sigma(\mathbf{P}) \subset \mathbb{R}$. If $\mathbf{P}f = \lambda f$ with $f \neq 0$ and $\lambda \in \sigma(\mathbf{P})$, then by Lemma 11.16,

$$\|f\|_\pi \geq \|\mathbf{P}f\|_\pi = \|\lambda f\|_\pi = |\lambda|\, \|f\|_\pi$$

which shows $|\lambda| \leq 1$ so that $\sigma(\mathbf{P}) \subset [-1, 1]$. As for any Markov matrix we have $\mathbf{P1} = \mathbf{1} = 1 \cdot \mathbf{1}$ from which it follows that $1 \in \sigma(\mathbf{P})$. We now prove the remaining items in turn.

1. If $\mathbf{P}$ is irreducible and aperiodic, then according to Proposition 8.43 there exists an $n \in \mathbb{N}$ so that $\mathbf{P}^n_{xy} > 0$ for all $x, y \in S$. Suppose that $f \in \ell^2(\pi)$ is a unit vector such that $\mathbf{P}f = \lambda f$ and f is perpendicular to 1, i.e.

$$0 = \langle f, 1\rangle_\pi = \pi(f) = \sum_{x\in S} f(x)\, \pi(x).$$

This last identity forces forces $f(x) < 0$ for some $x \in S$ and $f(y) > 0$ for another $y \in S$ and therefore

$$|\lambda|^n |f(x)| = |\lambda^n f(x)| = |(\mathbf{P}^n f)(x)| = \left|\sum_{y\in S} \mathbf{P}^n_{xy} f(y)\right| < \sum_{y\in S} \mathbf{P}^n_{xy} |f(y)|.$$

Multiplying this equation by $\pi(x)$ and summing the result gives

$$|\lambda|^n \pi(|f|) < \pi(|f|) \implies |\lambda|^n < 1 \implies |\lambda| < 1$$

and this completes the proof in the aperiodic case.

2. Let $N \in \mathbb{N}$ and set

$$\mathbf{Q} := \frac{1}{N}\sum_{n=1}^{N} \mathbf{P}^n$$

so that

$$\mathbf{Q}f(x) = \sum_{y\in S} q(x,y)\, f(y)$$

where

$$q(x,y) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{P}^n_{x,y}.$$

As $\mathbf{P}$ is irreducible, we know that for N sufficiently large, $q(x,y) > 0$ for all $x, y \in S$ so that $\mathbf{Q}$ is an aperiodic irreducible Markov matrix.
Now suppose that $f \in \ell^2(\pi)$ is a unit vector such that $\mathbf{P}f = \lambda f$ and f is perpendicular to 1. Then

$$\mathbf{Q}f = \frac{1}{N}\sum_{n=1}^{N}\mathbf{P}^n f = \frac{1}{N}\sum_{n=1}^{N}\lambda^n f = c_N f \tag{11.4}$$

where

$$c_N = \begin{cases} \frac{\lambda}{N}\frac{\lambda^{N+1}-1}{\lambda-1} & \text{if } \lambda \neq 1. \\ 1 & \text{if } \lambda = 1. \end{cases} \tag{11.5}$$

From the aperiodic case above applied to $\mathbf{Q}$ we may conclude that $|c_N| < 1$. On the other hand, from Eq. (11.5) we see that $|c_N| \to \infty$ if $|\lambda| > 1$ which would contradict $|c_N| < 1$ and we may conclude $|\lambda| \leq 1$. If $\lambda = 1$ we would have $c_N = 1$ (i.e. $\mathbf{Q}f = f$) which again leads to a contradiction since by item 1. we know 1 is an eigenvalue of $\mathbf{Q}$ with multiplicity 1. Thus we may conclude that either $|\lambda| < 1$ or $\lambda = -1$.

Remark. Here is another way to see in item 2. that $1 \in \sigma(\mathbf{P})$ has multiplicity 1. Let $f \in \ell^2(\pi)$ be a unit vector such that $\mathbf{P}f = f$. Then using Eq. (8.21) of the ergodic Theorem 8.27 we find,

$$\begin{aligned} f(x) &= \frac{1}{n}\sum_{m=1}^{n}[\mathbf{P}^m f]_x = \sum_{y\in S}\frac{1}{n}\sum_{m=1}^{n}\mathbf{P}^m_{xy} f(y) \\ &\to \sum_{y\in S}\pi(y) f(y) = \pi(f) \text{ as } n \to \infty, \end{aligned}$$

where π is the unique invariant distribution for $\mathbf{P}$. This shows $f(x) = \pi(f)$ for all $x \in S$, i.e. $f = \pi(f)\cdot\mathbf{1}$. ■

For the rest of this chapter we will be making the following assumption.

Assumption 11.23 *For this section we assume S is a finite set, $p : S \times S \to [0,1]$ is an irreducible Markov matrix*[2]*, $\pi : S \to [0,1]$ be the unique invariant distribution for p which we further assume satisfies the detail balance equation,*

$$\pi(x)\,p(x,y) = \pi(y)\,p(y,x).$$

As we know $\mathbf{P}1 = 1$, let us agree that we take $f_1 = 1$ and $\lambda_1 = 1$.

[2] Thus we know the associated chain is irreducible and the p has a unique strictly positive invariant distributions, π, on S.

Example 11.24. The Markov matrix,

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

is irreducible, $\pi = \left[\frac{1}{2}, \frac{1}{2}\right]$,

$$f_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } f_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

with $\lambda_1 = 1$ and $\lambda_2 = -1$. The fact that $\lambda_2 = -1$ is a reflection of the fact that $\mathbf{P}$ has period 2.

Example 11.25. The Markov matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

is irreducible, 2-periodic, has

$$\pi = \left[\tfrac{1}{6}\ \tfrac{1}{3}\ \tfrac{1}{6}\ \tfrac{1}{3}\right]$$

as its invariant distribution which satisfies detail balance. For example

$$\pi_2 P_{24} = \frac{1}{3}\cdot\frac{1}{2} = \pi_4 P_{42}.$$

In this case, $\sigma(\mathbf{P}) = \left\{-\frac{1}{2}, 1, \frac{1}{2}, -1\right\}$.

Example 11.26. The Markov matrix,

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

describes the chain which moves around the circle, $1 \to 2 \to 3 \to 1$. It is irreducible, 3 - periodic,

$$\sigma(\mathbf{P}) = \left\{-\frac{1}{2}i\sqrt{3} - \frac{1}{2}, \frac{1}{2}i\sqrt{3} - \frac{1}{2}, 1\right\}$$

which are the third roots of unity, and $\pi = [1/3, 1/3, 1/3]$ is its invariant measure. This matrix does **not** satisfy detail balance.

11.5 Convergence rates of reversible chains

In order for the Metropolis-Hastings algorithm (or other MCMC methods) to be effective, the Markov chain designed to converge to the given distribution must converge reasonably fast. The term used to measure this rate is the **mixing time** of the chain. (To make this precise, we need a precise measure of closeness to the stationary distribution; then we can ask how long before the distance to stationary is less than $\varepsilon > 0$.)

Notation 11.27 *If ν is a probability on S and $f : S \to \mathbb{R}$ is a function, let*

$$\nu(f) := \mathbb{E}_\nu f = \sum_{x \in S} f(x)\,\nu(x).$$

Theorem 11.28 ($L^2(\mathbb{P}_\nu)$ – convergence rates). *Let $\mathbf{P}$ be an irreducible and aperiodic Markov matrix, π be the invariant distribution of $\mathbf{P}$, and assume that $(\pi, \mathbf{P})$ are in detail balance. Let $\nu : S \to [0,1]$ be any starting distribution, $f : S \to \mathbb{R}$ be any function, and*

$$S_N(f) := \frac{1}{N}\sum_{m=1}^{N} f(X_m) \text{ for all } N \in \mathbb{N}$$

be the time average of f along the chain $\{X_m\}_{m=1}^N$. Then

$$\mathbb{E}_\nu |S_N(f) - \pi(f)|^2 = O\left(\frac{1}{N}\right).$$

Proof. We carry out the proof in three steps.

1. Using the spectral theorem, for any $f : S \to \mathbb{R}$ we have,

$$\begin{aligned}\mathbf{P}^n f &= \mathbf{P}^n \sum_{j=1}^{|S|} \langle f, u_j\rangle_\pi u_j = \sum_{j=1}^{|S|} \langle f, u_j\rangle_\pi \mathbf{P}^n u_j \\ &= \sum_{j=1}^{|S|} \langle f, u_j\rangle_\pi \lambda_j^n u_j = \langle f, \mathbf{1}\rangle_\pi \mathbf{1} + O(\alpha^n) \\ &= \pi(f) + O(\alpha^n).\end{aligned}$$

   where as above $\lambda_1 = 1$, $u_1 = \mathbf{1}$, and $\alpha = \max_{j>1}|\lambda_j| < 1$.
2. For $n \ge m$,

$$\begin{aligned}\mathbb{E}[f(X_m)\,g(X_n)] &= \nu\left(\mathbf{P}^m\left[f \cdot \mathbf{P}^{n-m} g\right]\right) = \nu\left(\mathbf{P}^m\left[f \cdot \left(\pi(g) + O\left(\alpha^{n-m}\right)\right)\right]\right) \\ &= \pi(g)\,\nu(\mathbf{P}^m f) + O\left(\alpha^{n-m}\right) \\ &= \pi(g)\,\nu(\pi(f) + O(\alpha^m)) + O\left(\alpha^{n-m}\right) \\ &= \pi(g)\,\pi(f) + O(\alpha^m) + O\left(\alpha^{n-m}\right).\end{aligned}$$

   Here we have used $\mathbf{P}$ is a contraction on $\ell^2(\pi)$ (see Lemma 11.16) and item 1. to conclude that $\mathbf{P}^m f = \pi(f) + O(\alpha^m)$.
3. Letting $h := f - \pi(f)$ so that $\pi(h) = 0$, we find

$$\begin{aligned}&\mathbb{E}\left[\left(S_N(f) - \pi(f)\right)^2\right] \\ &= \mathbb{E}\left[\left[S_N(f - \pi(f))\right]^2\right] = \mathbb{E}\left[\left[S_N(h)\right]^2\right] \\ &= \frac{1}{N^2}\sum_{m,n=1}^{N} \mathbb{E}[h(X_m)\,h(X_n)] \\ &= \frac{1}{N^2}\sum_{m=1}^{N} \mathbb{E}[h(X_m)\,h(X_m)] + \frac{2}{N^2}\sum_{1\le m<n\le N} \mathbb{E}[h(X_m)\,h(X_n)] \\ &\le \frac{N}{N^2}\|h\|_\infty^2 + \frac{2}{N^2}\sum_{1\le m<n\le N} \mathbb{E}[h(X_m)\,h(X_n)] \\ &\le O\left(\frac{1}{N}\right) + \frac{2}{N^2}\sum_{1\le m<n\le N}\left[O(\alpha^m) + O\left(\alpha^{n-m}\right)\right] = O\left(\frac{1}{N}\right).\end{aligned}$$

   For the last equality we have used,

$$\begin{aligned}&\sum_{1\le m<n\le N}\left[O(\alpha^m) + O\left(\alpha^{n-m}\right)\right] \\ &\qquad \le CN \cdot \sum_{1\le m<\infty} \alpha^m + C\sum_{1\le m\le N}\sum_{m<n<\infty} \alpha^{n-m} \le CN\end{aligned}$$

   owing to that fact that

$$\sum_{k=1}^{\infty} \alpha^k = \frac{\alpha}{1-\alpha} \text{ for } |\alpha| < 1.$$

■

Roughly speaking, the previous theorem indicates that the very best we can hope for in terms of pointwise convergence is

$$\pi(f) = \frac{1}{N}\sum_{n=1}^{N} f(X_n) + O\left(\frac{1}{\sqrt{N}}\right) \ \mathbb{P}_\nu - \text{a.s.}$$

11.6 Importance of the spectral gap

Let us try to be more precise on how many runs we will have to take in our simulations in order to get a good answer. As above, suppose that $\mathbf{P}$ is an irreducible Markov matrix with invariant distribution π which is in detail balance with $\mathbf{P}$. For simplicity, let us further assume that $\mathbf{P}$ is aperiodic.

Notation 11.29 *If $\nu : S \to [0,1]$ is a probability distribution and $\kappa \in \mathbb{N}_0$, let $\nu_\kappa := \nu \mathbf{P}^\kappa$ so that $\nu_\kappa(x) = \mathbb{P}_\nu(X_\kappa = x)$.*

Remark 11.30. When $\mathbf{P}$ is not aperiodic one should replace ν_κ by

$$\nu_\kappa(x) = \frac{1}{\kappa+1}\sum_{m=0}^{\kappa} \mathbb{P}_\nu(X_\kappa = x) \text{ i.e. } \nu_\kappa := \frac{1}{\kappa+1}\sum_{m=0}^{\kappa} \nu \mathbf{P}^m$$

in the arguments below.

Theorem 11.31. *Let us keep the above assumptions and notation in force and let $\{u_j\}_{j=1}^{\#(S)}$ be a $\langle\cdot,\cdot\rangle_\pi$ – orthonormal basis for $\ell^2(\pi)$ with $u_1 = \mathbf{1}$ and $\lambda_1 = 1$ and $\alpha := \max_{j>1}|\lambda_j| < 1$. Then for any $f : S \to \mathbb{R}$, starting distribution $\nu : S \to [0,1]$, and $\kappa \in \mathbb{N}_0$ we have,*

$$|\nu_\kappa(f) - \pi(f)|^2 \le C_\kappa(\nu)\,\mathrm{Var}_\pi(f) \tag{11.6}$$

where

$$C_\kappa(\nu) := \sum_{k=2}^{\#(S)} |\lambda_k|^{2\kappa}\left|\left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi\right|^2 \le \alpha^{2\kappa}\left(\nu\left(\frac{\nu}{\pi}\right) - 1\right). \tag{11.7}$$

In summary,

$$|\nu_\kappa(f) - \pi(f)|^2 \le \alpha^{2\kappa}\,\mathrm{Var}_\pi(f)\cdot\left(\nu\left(\frac{\nu}{\pi}\right) - 1\right). \tag{11.8}$$

Proof. Since $\mathbf{P}^\kappa u_j = \lambda_j^\kappa u_j$ and

$$f = \sum_{k=1}^{\#(S)} \langle f, u_k\rangle_\pi u_k = \pi(f) + \sum_{k=2}^{\#(S)} \langle f, u_k\rangle_\pi u_k$$

we find,

$$\mathbf{P}^\kappa f = \sum_{k=1}^{\#(S)} \langle f, u_k\rangle_\pi \mathbf{P}^\kappa u_k = \sum_{k=1}^{\#(S)} \lambda_k^\kappa \langle f, u_k\rangle_\pi u_k = \pi(f) + \sum_{k=2}^{\#(S)} \lambda_k^\kappa \langle f, u_k\rangle_\pi u_k$$

and hence

$$\begin{aligned}\nu_\kappa(f) = \nu\mathbf{P}^\kappa f &= \pi(f) + \sum_{k=2}^{\#(S)} \lambda_k^\kappa \langle f, u_k\rangle_\pi \nu(u_k)\\ &= \pi(f) + \sum_{k=2}^{\#(S)} \lambda_k^\kappa \langle f, u_k\rangle_\pi \left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi.\end{aligned}$$

Making use the Cauchy-Schwarz inequality it then follows that

$$\begin{aligned}|\nu\mathbf{P}^\kappa f - \pi(f)|^2 &\le \left(\sum_{k=2}^{\#(S)} |\lambda_k|^\kappa |\langle f, u_k\rangle_\pi| \left|\left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi\right|\right)^2\\ &\le \sum_{k=2}^{\#(S)} |\lambda_k|^{2\kappa}\left|\left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi\right|^2 \cdot \sum_{k=2}^{\#(S)} |\langle f, u_k\rangle_\pi|^2\\ &= C_\kappa(\nu)\cdot\sum_{k=2}^{\#(S)} |\langle f, u_k\rangle_\pi|^2.\end{aligned}$$

Since

$$\sum_{k=2}^{\#(S)} |\langle f, u_k\rangle_\pi|^2 = \|f\|_\pi^2 - \langle f, \mathbf{1}\rangle_\pi^2 = \pi\left(f^2\right) - [\pi(f)]^2 = \mathrm{Var}_\pi(f),$$

the previous two equations proves Eq. (11.6). Moreover,

$$\begin{aligned}C_\kappa &:= \sum_{k=2}^{\#(S)} |\lambda_k|^{2\kappa}\left|\left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi\right|^2\\ &\le \alpha^{2\kappa}\sum_{k=2}^{\#(S)} \left|\left\langle \frac{\nu}{\pi}, u_k\right\rangle_\pi\right|^2\\ &= \alpha^{2\kappa}\cdot\left(\left\|\frac{\nu}{\pi}\right\|_\pi^2 - \left\langle \frac{\nu}{\pi}, \mathbf{1}\right\rangle_\pi^2\right)\\ &= \alpha^{2\kappa}\cdot\left(\nu\left(\frac{\nu}{\pi}\right) - 1\right)\end{aligned}$$

which proves the inequality in Eq. (11.7). ■

Example 11.32. If π is the uniform distribution on S and $\nu = \delta_x$ for some $x \in S$, then

$$\nu\left(\frac{\nu}{\pi}\right) - 1 = |S| - 1 \le |S|.$$

If we further assume that $|f| \le K$ for some constant K, then $\mathrm{Var}(f) \le K^2$ and we have,

$$|\nu_\kappa(f) - \pi(f)|^2 \le \alpha^{2\kappa}K^2|S|.$$

If we are given $\varepsilon > 0$ as are allowed error, then we should choose κ so that

$$\alpha^{2\kappa}K^2|S| \le \varepsilon^2 \implies 2\kappa = \frac{\ln\left(\frac{\varepsilon^2}{K^2|S|}\right)}{\ln\alpha} \implies \kappa = \frac{\ln\left(\frac{\varepsilon}{K\sqrt{|S|}}\right)}{\ln\alpha}$$

i.e.

$$\kappa \geq \frac{\ln\left(K\sqrt{|S|}/\varepsilon\right)}{|\ln\alpha|} = \frac{\ln\left(K\sqrt{|S|}/\varepsilon\right)}{\ln\left(\alpha^{-1}\right)}.$$

So for example if $K = 1$, $|S| = 2^{50^2}$, $\varepsilon = 10^{-2}$, and $\alpha^{-1} = e^{1/10} = 0.904\,84$ then we require,

$$\kappa \geq 10\ln\left(100 \cdot 2^{50^2/2}\right) \cong 8710.4.$$

Thus if take independent runs, $\{X_n^m : n \in \mathbb{N}_0\}_{m=1}^n$ of our chain starting at a fixed point $x \in S$, we will have by the law of large numbers that

$$\frac{1}{n}\sum_{m=1}^n f\left(X_\kappa^m\right) \to \nu_\kappa\left(f\right) = \mathbb{E}_\nu f\left(X_\kappa\right) \text{ as } n \to \infty.$$

Moreover, using the weak law of large numbers (see Theorem 2.20), we conclude that for any $\delta > 0$ that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^n f\left(X_\kappa^m\right) - \nu_\kappa\left(f\right)\right| \geq \delta\right) \leq \frac{1}{\delta^2}\frac{1}{n}\operatorname{Var}_{\nu_\kappa}\left(f\right) \leq \frac{1}{\delta^2 n}1^2.$$

Taking $\delta = 10^{-2}$ we find,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^n f\left(X_\kappa^m\right) - \nu_\kappa\left(f\right)\right| \geq 10^{-2}\right) \leq \frac{10^2}{n}.$$

In conclusion if we take $n = 10^4$ and $\kappa = 8711$ we would find that, 99% of the time, that

$$\left|\frac{1}{10^4}\sum_{m=1}^{10^4} f\left(X_{8711}^m\right) - \pi\left(f\right)\right| \leq 2 \cdot 10^{-2}.$$

The difficult thing here is of course getting an estimate on the size of α.

To get more detailed information about this topic, the reader might start by consulting [12], [4], and [13, Chapter 17] and the references therein.

11.7 Dropping the aperiodic assumption*

Advise to the reader: ignore this section and in fact the rest of this chapter on first reading.

Notation 11.33 *Let Π_λ denote orthogonal projection onto the λ – eigenspace of $\mathbf{P}$ for all $\lambda \in \sigma\left(\mathbf{P}\right)$. We then have,*

$$\mathbf{P} = \Pi_1 - \Pi_{-1} + R \text{ where}$$
$$R = \sum_{\lambda\in\sigma(\mathbf{P})\setminus\{\pm1\}} \lambda\Pi_\lambda.$$

We further let

$$\alpha := \max\left\{|\lambda| : \lambda \in \sigma\left(\mathbf{P}\right) \setminus \{\pm1\}\right\} < 1.$$

With this notation we have for $k \in \mathbb{N}$ that

$$\mathbf{P}^k = \Pi_1 + \left(-1\right)^k \Pi_{-1} + R^k \tag{11.9}$$

where $\left\|R^k\right\| \leq \alpha^k$. Moreover if $\mathbf{P}$ is aperiodic, then $\Pi_{-1} = 0$. In all cases

$$\Pi_1 f = \left\langle f, 1\right\rangle_\pi 1 = \pi\left(f\right).$$

Corollary 11.34. *Under Assumption 11.23, if $\nu : S \to [0,1]$ is any probability on S and $f : S \to \mathbb{R}$ is any function, then*

$$\mathbb{E}_\nu\left[f\left(X_n\right)\right] = \pi\left(f\right) + \left(-1\right)^n \nu\left(\Pi_{-1}f\right) + \|f\| \cdot O_\nu\left(\alpha^n\right). \tag{11.10}$$

In particular if $\mathbf{P}$ is aperiodic, then

$$\mathbb{E}_\nu\left[f\left(X_n\right)\right] = \pi\left(f\right) + \|f\| \cdot O_\nu\left(\alpha^n\right).$$

Proof. Let $\rho\left(x\right) := \nu\left(x\right)/\pi\left(x\right)$ where π is the invariant distribution for $\mathbf{P}$. By the properties of Markov chains we have,

$$\begin{aligned}\mathbb{E}_\nu\left[f\left(X_n\right)\right] &= \sum_{x,y\in S} \nu\left(x\right)\mathbf{P}_{x,y}^n f\left(y\right) = \sum_{x\in S} \nu\left(x\right)\left(\mathbf{P}^n f\right)\left(x\right) \\ &= \sum_{x\in S}\left(\mathbf{P}^n f\right)\left(x\right)\rho\left(x\right)\pi\left(x\right) = \left\langle \mathbf{P}^n f, \rho\right\rangle_\pi = \left\langle f, \mathbf{P}^n\rho\right\rangle_\pi.\end{aligned}$$

Making use of Eq. (11.9) we find,

$$\begin{aligned}\left\langle f, \mathbf{P}^n\rho\right\rangle_\pi &= \left\langle f, \left(\Pi_1 + \left(-1\right)^n \Pi_{-1} + R^n\right)\rho\right\rangle_\pi \\ &= \pi\left(f\right) + \left(-1\right)^n\left\langle \Pi_{-1}f, \rho\right\rangle_\pi + \|f\| O\left(\alpha^n\right) \\ &= \pi\left(f\right) + \left(-1\right)^n \nu\left(\Pi_{-1}f\right) + \|f\| O\left(\alpha^n\right).\end{aligned}$$

■

Remark 11.35. The above corollary suggests that in the aperiodic case in order to estimate $\pi(f)$ a good strategy is to choose n fairly large so that the error estimate

$$|\mathbb{E}_\nu[f(X_n)] - \pi(f)| \le C_1\alpha^n$$

forces $\mathbb{E}_\nu[f(X_n)] \cong \pi(f)$. We then estimate $\mathbb{E}_\nu[f(X_n)]$ using the law of large numbers, i.e.

$$\mathbb{E}_\nu[f(X_n)] \cong \frac{1}{N}\sum_{k=1}^N f(x_n(k))$$

where $\{x_n(k)\}_{k=0}^N$ are N independent samples from X_n. We then expect to have

$$\left|\pi(f) - \frac{1}{N}\sum_{k=1}^N f(x_n(k))\right| \le C_1\alpha^n + C_2\frac{1}{\sqrt{N}}$$

The typical difficulties here are that C_1, α, and the random constant C_2 are typically not easy to estimate.

Notation 11.36 *Let $S_N(f) := \frac{1}{N}\sum_{n=1}^N f(X_n)$ be the sample averages of $f : S \to \mathbb{R}$.*

Corollary 11.37. *Under Assumption 11.23,*

$$\mathbb{E}_\nu[S_N(f)] = \frac{1}{N}\sum_{n=1}^N \mathbb{E}_\nu[f(X_n)] = \pi(f) + \|f\| O\left(\frac{1}{N}\right). \tag{11.11}$$

Proof. Summing Eq. (11.10) on n shows,

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^N \mathbb{E}_\nu[f(X_n)] &= \pi(f) + \frac{\varepsilon_N}{N}\nu(\Pi_{-1}f) + \|f\| O\left(\frac{1}{N}\sum_{n=1}^N \alpha^n\right)\\ &= \pi(f) + \frac{\varepsilon_N}{N}\nu(\Pi_{-1}f) + \|f\| O\left(\frac{\alpha}{N}\frac{1}{1-\alpha}\right)\\ &= \pi(f) + \|f\| O\left(\frac{1}{N}\right),\end{aligned}$$

where $\varepsilon_N \in \{0, \pm 1\}$. ∎

Corollary 11.38 ($L^2(\mathbb{P}_\nu)$ – convergence rates). *Given the assumptions above for any initial distribution, ν,*

$$\mathbb{E}_\nu|S_N - \pi(f)|^2 = O\left(\frac{1}{N}\right).$$

Proof. If we let $f = \pi(f) + h$ (i.e.$h = f - \pi(f)$), then

$$S_N(f) = \frac{1}{N}\sum_{n=1}^N f(X_n) = \pi(f) + \frac{1}{N}\sum_{n=1}^N h(X_n)$$

and therefore,

$$\begin{aligned}\mathbb{E}[S_N - \pi(f)]^2 &= \frac{1}{N^2}\sum_{m,n=1}^N \mathbb{E}[h(X_n)h(X_m)]\\ &= \frac{1}{N^2}\left[\sum_{m=1}^N \mathbb{E}\left[h^2(X_m)\right] + 2\sum_{1\le m<n\le N}\mathbb{E}[h(X_n)h(X_m)]\right]\\ &\le \|h\|_\infty^2\frac{1}{N} + \frac{2}{N^2}\sum_{1\le m<n\le N}\mathbb{E}[h(X_n)h(X_m)].\end{aligned}$$

To finish the proof it suffices to show

$$\begin{aligned}&\sum_{1\le m<n\le N}\mathbb{E}[h(X_n)h(X_m)]\\ &\qquad = \sum_{1\le m<n\le N}\mathbb{E}\left[h(X_m)\left(\mathbf{P}^{n-m}h\right)(X_m)\right] = O(N),\end{aligned}$$

wherein we have used the Markov property in the first equality.

Case 1. Suppose that $\mathbf{P}$ is aperiodic. Under this additional assumption, we know that $\|\mathbf{P}^{n-m}h\|_\infty \le C\|h\|_\infty\alpha^{n-m}$ and hence we find,

$$\sum_{1\le m<n\le N}\mathbb{E}[h(X_n)h(X_m)] \le C\|h\|_\infty^2\sum_{1\le m<n\le N}\alpha^{n-m}.$$

This completes the proof in this case upon observing that

$$\sum_{1\le m<n\le N} O\left(\alpha^{n-m}\right) \le C\sum_{1\le m<n\le N}\alpha^{n-m} \le \frac{\alpha}{1-\alpha}N = O(N). \tag{11.12}$$

Case 2. For the **general case,** we decompose h as

$$h = f_- + g = \Pi_{-1}f + \sum_{\lambda\in\sigma(\mathbf{P})\setminus\{\pm\}}\Pi_\lambda f$$

so that

$$\mathbf{P}^{n-m}h = (-1)^{n-m}f_- + \mathbf{P}^{n-m}g.$$

With this notation along with Eq. (11.10) we learn

$$\begin{aligned}\mathbb{E}\left[h\left(X_n\right)h\left(X_m\right)\right] &= \mathbb{E}\left[h\left(X_m\right)\left(\mathbf{P}^{n-m}h\right)\left(X_m\right)\right] \\ &= \mathbb{E}\left[h\left(X_m\right)\left[\left(-1\right)^{n-m}f_- + \mathbf{P}^{n-m}g\right]\left(X_m\right)\right] \\ &= \left(-1\right)^{n-m}\mathbb{E}\left[h\left(X_m\right)f_-\left(X_m\right)\right] + O\left(\alpha^{n-m}\right) \\ &= \left(-1\right)^{n-m}\pi\left(hf_-\right) + \left(-1\right)^m\nu\left(\Pi_{-1}\left[hf_-\right]\right) + O\left(\alpha^m\right) + O\left(\alpha^{n-m}\right).\end{aligned}$$

The last displayed equation combined with Eq. (11.12) and Lemma 11.39 below again implies

$$\sum_{1\le m<n\le N}\mathbb{E}\left[h\left(X_n\right)h\left(X_m\right)\right] = O\left(N\right).$$

■

Lemma 11.39. *The following sums are all order* $O\left(N\right)$,

$$\sum_{1\le m<n\le N}\left(-1\right)^m,\quad \sum_{1\le m<n\le N}\left(-1\right)^n,\quad \sum_{1\le m,n\le N}\left(-1\right)^{m\wedge n},$$
$$\sum_{1\le m,n\le N}\left(-1\right)^{m\vee n},\ \textit{and}\quad \sum_{1\le m,n\le N}\left(-1\right)^{|m-n|}.$$

Proof. These results are elementary and are base on the fact that for any $0\le k<p<\infty$ we have

$$\sum_{l=k}^{p}\left(-1\right)^l\in\left\{0,\pm1\right\}.$$

For example,

$$\begin{aligned}\sum_{1\le m,n\le N}\left(-1\right)^{m\wedge n} &= \sum_{m=1}^{N}\left(-1\right)^m + 2\sum_{1\le m<n\le N}\left(-1\right)^m \\ &\in \left\{0,\pm1\right\} + 2\sum_{1<n\le N}\left\{0,\pm1\right\}\end{aligned}$$

which is certainly bounded by CN. The largest term is,

$$\begin{aligned}\sum_{1\le m,n\le N}\left(-1\right)^{|m-n|} &= \sum_{m=1}^{N}1 + 2\sum_{1\le m<n\le N}\left(-1\right)^{(n-m)} \\ &= N + \sum_{1\le m<N}\left\{0,\pm1\right\}\end{aligned}$$

which is again bounded by CN. ■

11.8 *Reversible Markov Chains

Let S be a finite or countable state space, $p:S\times S\to\left[0,1\right]$ be a Markov transition kernel, and $\nu:S\to\left[0,1\right]$ be a probability on S. We now wish to find conditions on ν and p so that, for each $N\in\mathbb{N}$, $Y_n := X_{N-n}$ for $0\le n\le N$ is distributed as a time homogeneous Markov chain. Thus we are looking for a Markov transition kernel, $q:S\times S\to\left[0,1\right]$, and a probability, $\tilde\nu:S\to\left[0,1\right]$, such that

$$\mathbb{P}_\nu\left(Y_0=x_0,\dots,Y_N=x_N\right) = \tilde\nu\left(x_0\right)q\left(x_0,x_1\right)\dots q\left(x_{N-1},x_N\right)$$

for all $x_j\in S$. By assumption,

$$\begin{aligned}\mathbb{P}_\nu\left(Y_0=x_0,\dots,Y_N=x_N\right) &= \mathbb{P}_\nu\left(X_0=x_0,\dots,X_0=x_N\right) \\ &= \nu\left(x_N\right)p\left(x_N,x_{N-1}\right)\dots p\left(x_1,x_0\right).\end{aligned}$$

Comparing these last two displayed equations leads us to require, for all $N\in\mathbb{N}_0$ and $\left\{x_j\right\}\subset S$, that

$$\tilde\nu\left(x_0\right)q\left(x_0,x_1\right)\dots q\left(x_{N-1},x_N\right) = \nu\left(x_N\right)p\left(x_N,x_{N-1}\right)\dots p\left(x_1,x_0\right).$$

Taking $N=0$ forces us to take $\tilde\nu=\nu$. Taking $N=1$ we have,

$$\nu\left(x_0\right)q\left(x_0,x_1\right) = \nu\left(x_1\right)p\left(x_1,x_0\right) \tag{11.13}$$

and then summing this equation on x_1 shows $\nu\left(x_0\right) = \left(\nu\mathbf{P}\right)\left(x_0\right)$ and forces ν to be a stationary distribution for p.

From now on let us assume that p has a strictly positive stationary distribution, π, i.e. that all communication classes of p are positive recurrent. Let ν be one of these stationary distributions. Then from Eq. (11.13) we see that we must take,

$$q\left(x_0,x_1\right) = \frac{\nu\left(x_1\right)}{\nu\left(x_0\right)}p\left(x_1,x_0\right)\ge 0.$$

Observe that

$$\sum_{x_1\in S}q\left(x_0,x_1\right) = \sum_{x_1\in S}\frac{\nu\left(x_1\right)}{\nu\left(x_0\right)}p\left(x_1,x_0\right) = \frac{\left(\nu\mathbf{P}\right)\left(x_0\right)}{\nu\left(x_0\right)} = \frac{\nu\left(x_0\right)}{\nu\left(x_0\right)} = 1$$

and that

$$\sum_{x_0\in S}\nu\left(x_0\right)q\left(x_0,x_1\right) = \sum_{x_0\in S}\nu\left(x_1\right)p\left(x_1,x_0\right)\nu\left(x_1\right)$$

which shows that $q:S\times S\to\left[0,1\right]$ is a Markov transition kernel and with stationary distribution, ν. Finally we have

$$
\begin{aligned}
&\nu(x_0) q(x_0, x_1) \dots q(x_{N-1}, x_N) \\
&\quad = \nu(x_0) \frac{\nu(x_1)}{\nu(x_0)} p(x_1, x_0) \frac{\nu(x_2)}{\nu(x_1)} p(x_2, x_1) \dots \frac{\nu(x_N)}{\nu(x_{N-1})} p(x_N, x_{N-1}) \\
&\quad = p(x_N, x_{N-1}) \dots p(x_1, x_0) \nu(x_N)
\end{aligned}
$$

as desired.

Definition 11.40. *We say the chain $\{X_n\}$ associated to p is (time)* ***reversible*** *if there exists a distributions ν on S such that for any $N \in \mathbb{N}$,*

$$
(X_0, \dots, X_N) \overset{\mathbb{P}_\nu}{=} (X_N, X_{N-1}, \dots, X_0).
$$

What we have proved above is that p is reversible iff there exists $\nu : S \to (0, \infty)$ such that

$$
\nu(x) p(x, y) = \nu(y) p(y, x) \ \ \forall\ x, y \in S.
$$

Hidden Markov Models

Problem: Let $\{X_n\} \subset S$ be a stochastic process and $\{\mathcal{E}_n : S \to O\}_{n=0}^{\infty}$ is a collection of random functions. It is often the case that we can not observe $\{X_n\}$ directly but only some noisy output $\{U_n = \mathcal{E}_n(X_n)\}_{n=0}^{\infty}$ of the $\{X_n\}$. Our goal is to predict the values of $\{X_n\}_{n=0}^{N}$ given that we have observe $\mathbf{u} := \{u_n\}_{n=0}^{N} \subset O$. One way to make such a prediction is based on finding the most probable trajectory, i.e. we are looking for $(x_0^*, \dots, x_N^*) \in S^{N+1}$ which is a maximizer of the function,

$$\begin{aligned}\hat{F}_N(x_0, \dots, x_N) &= P\left(\{X_n = x_n\}_{n=0}^{N} \,|\, \{U_n = u_n\}_{n=0}^{N}\right) \\ &= \frac{P\left(\{X_n = x_n\}_{n=0}^{N}, \{U_n = u_n\}_{n=0}^{N}\right)}{P\left(\{U_n = u_n\}_{n=0}^{N}\right)}.\end{aligned}$$

Equivalently we may define $(x_0^*, \dots, x_N^*)$ as,

$$(x_0^*, \dots, x_N^*) = \arg \max_{(x_0, \dots, x_N) \in S^{N+1}} F_N(x_0, \dots, x_N)$$

where$F_N(x_0, \dots, x_N)$

$$\begin{aligned}F_N(x_0, \dots, x_N) &= P\left(\{X_n = x_n\}_{n=0}^{N}, \{U_n = u_n\}_{n=0}^{N}\right) \\ &= P\left(\{(X_n, U_n) = (x_n, u_n)\}_{n=0}^{N}\right). \qquad (12.1)\end{aligned}$$

With no extra structure this maximization problem is intractable even for moderately large N since it requires $|S|^{N+1}$ evaluations of F_N.[1] On the other hand, we will be able to say something about this problem if we make the following additional assumptions.

Assumption 12.1 *Keeping the notation above, let us now further assume that;*

1. *$\{X_n\}$ is a **(hidden)** Markov chain with transition probabilities $p : S \times S \to [0,1]$ and initial distribution $\nu : S \to [0,1]$ and*
2. *the "random noise" functions, $\{\mathcal{E}_n : S \to O\}_{n=0}^{\infty}$, are i.i.d. random functions which are independent of the chain $\{X_n\}_{n=0}^{\infty}$.*

[1] It is worth observing that the dimension of the space of possible functions F_N is also roughly $|S|^{N+1}$.

Notation 12.2 *Under Assumption 12.1, the **emission probabilities** refers to the function, $e : S \times O \to [0,1]$ defined by*

$$e(x, u) := P(\mathcal{E}_n(x) = u) \;\; \forall \;\; (x, u) \in S \times O.$$

Given $\mathbf{u} := \{u_n\}_{n=0}^{N} \subset O$, if we let

$$\begin{aligned}V_0(x_0) &= \nu(x_0)\, e(x_0, u_0) \;\text{ and} \\ q_n(x, y) &:= p(x, y)\, e(y, u_n), \qquad (12.2)\end{aligned}$$

then the function F_N in Eq. (12.1) becomes,

$$\begin{aligned}F_N(x_0, \dots, x_N) &= P\left(\{X_n = x_n\}_{n=0}^{N}\right) \cdot P\left(\{\mathcal{E}_n(x_n) = u_n\}_{n=0}^{N}\right) \\ &= \nu(x_0)\, e(x_0, u_0) \prod_{n=1}^{N} p(x_{n-1}, x_n)\, e(x_n, u_n) \qquad (12.3) \\ &= V_0(x_0) \prod_{n=1}^{N} q_n(x_{n-1}, x_n). \qquad (12.4)\end{aligned}$$

Example 12.3 (Occasionally Dishonest Casino). In a casino game, a die is thrown. Usually a fair die is used, but sometimes the casino swaps it out for a loaded die. To be precise: there are two dice, F (fair) and L (loaded). For the fair die F, the probability of rolling i is $\frac{1}{6}$ for $i = 1, 2, \dots, 6$. For the loaded die, however, 1 is rolled with probability 0.5 while 2 through 5 are each rolled with probability 0.1. Each roll, the casino randomly switches out F for L with probability 0.05; if L is in use, they switch back to F next roll with probability 0.9.

In this example, we may take $S = \{F, L\}$ (F =fair and L =loaded) and $(X_n)_{n \geq 0}$ to be the Markov chain keeping track of which dice is in play and has transition probabilities,

$$\mathbf{P} = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} F & L \\ \end{array}$$
$$\mathbf{P} = \begin{array}{c} \\ \end{array}\begin{bmatrix} 0.95 & 0.05 \\ 0.9 & 0.1 \end{bmatrix}\begin{array}{c} F \\ L \end{array}.$$

We further let $O = \{1, 2, 3, 4, 5, 6\}$ and we are given the emission probabilities,

$$e(x,u) = \left[\begin{array}{c|cccccc} x/u & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline F & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ L & 0.5 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{array}\right]$$

Questions: You notice that 1 is rolled 6 times in a row. How likely is it that the fair die was in use for those rolls (given the above information)? How likely is it the loaded die was used in rolls 2, 3, and 5? More generally, what is the *most likely* sequence of dice that was used to generate this sequence of rolls?

Example 12.3 is a very typical situation that occurs in many real-world problems. Another application of Hidden Markov Models (HMMs) is in *machine speech recognition*. Here we have $S = \{\text{words}\}$ and $O = \{\text{wave forms}\}$.

12.1 The most likely trajectory via dynamic programing

Let $V_0 : S \to [0,\infty)$ and $q_n : S \times S \to [0,\infty)$ for $n \in \mathbb{N}$ be given functions and then set

$$F_n(x_0, x_1, \dots, x_n) = V_0(x_0)\, q_1(x_0, x_1)\, q_2(x_1, x_2) \dots q_n(x_{n-1}, x_n). \tag{12.5}$$

The space of functions, F_n, that have the form in Eq. (12.5) is now approximately, $n|S|^2 + |S|$ which is much smaller than $|S|^{n+1}$ – the dimension of all functions, F_n on S^{n+1}. Viterbi's algorithm exploits this fact in order to find a maximizer of F_n with only order of $n|S|^2$ operations rather than $|S|^{n+1}$. The key observations is that we may describe F_n inductively as;

$$\begin{aligned} F_1(x_0, x_1) &= V_0(x_0)\, q_1(x_0, x_1) \text{ and} \\ F_n(x_0, x_1, \dots, x_n) &= F_{n-1}(x_0, x_1, \dots, x_{n-1})\, q_n(x_{n-1}, x_n). \end{aligned} \tag{12.6}$$

From Eq. (12.6) it follows that

$$\begin{aligned} &\max_{x_0,x_1,\dots,x_n \in S} F_n(x_0, x_1, \dots, x_n) \\ &\quad = \max_{x_0,x_1,\dots,x_n \in S} [F_{n-1}(x_0, x_1, \dots, x_{n-1})\, q_n(x_{n-1}, x_n)] \\ &\quad = \max_{x_n \in S} \max_{x_{n-1} \in S} \left[\max_{x_0,x_1,\dots,x_{n-2} \in S} F_{n-1}(x_0, x_1, \dots, x_{n-1})\right] q_n(x_{n-1}, x_n). \end{aligned}$$

This identity suggest that we define (for $n \in \mathbb{N}$) $V_n : S \to [0,1]$ by,

$$V_n(x_n) := \max_{x_0,x_1,\dots,x_{n-1} \in S} F_n(x_0, x_1, \dots, x_n) \text{ for } n \geq 1. \tag{12.7}$$

It then follows that

$$\max_{x_0,x_1,\dots,x_n \in S} F_n(x_0, x_1, \dots, x_n) = \max_{x_n \in S} V_n(x_n)$$

and

$$\begin{aligned} V_n(x_n) &= \max_{x_0,x_1,\dots,x_{n-1} \in S} [F_{n-1}(x_0, x_1, \dots, x_{n-1})\, q_n(x_{n-1}, x_n)] \\ &= \max_{x_{n-1} \in S} \max_{x_0,x_1,\dots,x_{n-2} \in S} [F_{n-1}(x_0, x_1, \dots, x_{n-1})\, q_n(x_{n-1}, x_n)] \\ &= \max_{x_{n-1} \in S} [V_{n-1}(x_{n-1})\, q_n(x_{n-1}, x_n)]. \end{aligned}$$

Notation 12.4 *If $G : S \to \mathbb{R}$ is a function on a finite set S, let $\arg\max_{x \in S} G(x)$ denote a point $x^* \in S$ such that $G(x^*) = \max_{x \in S} G(x)$.*

The above comments lead to the following algorithm for finding a maximizer of F_N.

Theorem 12.5 (Viterbi's most likely trajectory). *Let us continue the notation above so that for $n \geq 1$,*

$$V_n(y) = \max_{x \in S} [V_{n-1}(x)\, q_n(x, y)]$$

and further let

$$\varphi_n(y) := \arg\max_{x \in S} [V_{n-1}(x)\, q_n(x, y)].$$

Given $N \in \mathbb{N}$, if we let

$$\begin{aligned} x_N^* &= \arg\max_{x \in S} V_N(x), \\ x_{N-1}^* &= \varphi_N(x_N^*), \\ x_{N-2}^* &= \varphi_{N-1}(x_{N-1}^*), \\ &\vdots \\ x_0^* &:= \varphi_1(x_1^*), \end{aligned}$$

then

$$(x_0^*, x_1^*, \dots, x_N^*) = \arg \max_{(x_0,\dots,x_N) \in S^{N+1}} F_N(x_0, \dots, x_N) \tag{12.8}$$

and

$$M := \max_{(x_0,\dots,x_N) \in S^{N+1}} F_N(x_0, \dots, x_N) = V_N(x_N^*) = \max_x V_N(x). \tag{12.9}$$

Proof. First off Eq. (12.9) is just a rewrite of Eq. (12.7). Moreover,

$$\begin{aligned} M = V_N(x_N^*) &= \max_{x_{N-1}\in S}\left[V_{N-1}(x_{N-1})\, q_N(x_{N-1}, x_N^*)\right] \\ &= V_{N-1}\left(x_{N-1}^*\right) q_N\left(x_{N-1}^*, x_N^*\right) \end{aligned}$$

while

$$\begin{aligned} V_{N-1}\left(x_{N-1}^*\right) &= \max_{x_{N-2}\in S}\left[V_{N-2}(x_{N-2})\, q_{N-1}\left(x_{N-2}, x_{N-1}^*\right)\right] \\ &= V_{N-2}\left(x_{N-2}^*\right) q_{N-1}\left(x_{N-2}^*, x_{N-1}^*\right) \end{aligned}$$

so that

$$M = V_{N-2}\left(x_{N-2}^*\right) q_{N-1}\left(x_{N-2}^*, x_{N-1}^*\right) q_N\left(x_{N-1}^*, x_N^*\right).$$

Continuing this way inductively shows for $1 \le k \le N$

$$M = V_{N-k}\left(x_{N-k}^*\right) q_{N-k+1}\left(x_{N-k}^*, x_{N-k+1}^*\right) \dots q_{N-1}\left(x_{N-2}^*, x_{N-1}^*\right) q_N\left(x_{N-1}^*, x_N^*\right)$$

and in particular at $k = N$ we find,

$$\begin{aligned} M &= V_0(x_0^*)\, q_1(x_0^*, x_1^*) \dots q_{N-1}\left(x_{N-2}^*, x_{N-1}^*\right) q_N\left(x_{N-1}^*, x_N^*\right) \\ &= F_N(x_0^*, x_1^*, \dots, x_N^*) \end{aligned}$$

which proves Eq. (12.8).

Proof in the special case where $N = 3$. The algorithm gives;

$$\begin{aligned} V_1(y) &:= \max_x V_0(x)\, q_1(x, y) = V_0(\varphi_1(y))\, q_1(\varphi_1(y), y), \\ V_2(y) &:= \max_x V_1(x)\, q_2(x, y) = V_1(\varphi_2(y))\, q_2(\varphi_2(y), y), \\ V_3(y) &:= \max_x V_2(x)\, q_3(x, y) = V_2(\varphi_3(y))\, q_3(\varphi_3(y), y). \end{aligned}$$

Now choose x_3^* so that

$$M = \max F_3(x_0, x_1, x_2, x_3) = \max_y V_3(y) = V_3(x_3^*)$$

and then let

$$x_2^* = \varphi_3(x_3^*),\ x_1^* = \varphi_2(x_2^*),\ \text{and } x_0^* = \varphi_1(x_1^*).$$

We then find by back substitution that

$$\begin{aligned} M = \max_y V_3(y) &= V_3(x_3^*) \\ &= V_2(\varphi_3(x_3^*))\, q_3(\varphi_3(x_3^*), x_3^*) = V_2(x_2^*)\, q_3(x_2^*, x_3^*) \\ &= V_1(\varphi_2(x_2^*))\, q_2(\varphi_2(x_2^*), x_2^*)\, q_3(x_2^*, x_3^*) \\ &= V_1(x_1^*)\, q_2(x_1^*, x_2^*)\, q_3(x_2^*, x_3^*) \\ &= V_0(\varphi_1(x_1^*))\, q_1(\varphi_1(x_1^*), x_1^*)\, q_2(x_1^*, x_2^*)\, q_3(x_2^*, x_3^*) \\ &= V_0(x_0^*)\, q_1(x_0^*, x_1^*)\, q_2(x_1^*, x_2^*)\, q_3(x_2^*, x_3^*) \\ &= F_3(x_0^*, x_1^*, x_2^*, x_3^*) \end{aligned}$$

and so $(x_0^*, x_1^*, x_2^*, x_3^*)$ is a maximizer of $F_3(x_0, x_1, x_2, x_3)$. ■

12.1.1 Viterbi's algorithm in the Hidden Markov chain context

Let us now write out the pseudo code of this algorithm in the hidden Markov chain context. We now assume the setup described in Notation 12.2 and we assume that we have observed $\mathbf{u} := \{u_k\}_{k=0}^N \subset O$.

Hidden Markov chain pseudo code:

For $y \in S$,

$\quad V_0(y) = \nu(y)\, e(y, u_0)$

End $y \in S$.

For $n = 1$ to N,

$\quad$ For $y \in S$,

$$\begin{aligned} V_n(y) &:= \max_{x\in S}\left[V_{n-1}(x)\, q_n(x, y)\right] \\ &= \max_{x\in S}\left[V_{n-1}(x)\, p(x, y)\, e(x, u_n)\right] \text{ and} \\ \varphi_n(y) &= \arg\max_{x\in S}\left[V_{n-1}(x)\, q_n(x, y)\right] \\ &= \arg\max_{x\in S}\left[V_{n-1}(x)\, p(x, y)\, e(x, u_n)\right] \end{aligned}$$

$\quad$ End $y \in S$.

End $n = 1$ to N

$x_N^* := \arg\max_{y\in S} V_n(y)$

For $n = 1$ to N,

$x_{N-n}^* = \varphi_{N-n+1}\left(x_{N-n+1}^*\right)$

End $n = 1$ to N.

Output: $(x_0^*, x_1^*, \dots, x_N^*)$.

The output $\{x_k^*\}_{k=0}^N$ is **a** most likely trajectory given $\{u_k\}_{k=0}^N$, i.e.

$$(x_0^*, x_1^*, \dots, x_N^*) = \arg\max_{(x_0,\dots,x_N)\in S^{N+1}} F_N(x_0, \dots, x_N).$$

Complexity analysis: Notice that the computational complexity of this algorithm is approximately, $C \cdot N\,|S|^2$ rather than $|S|^{N+1}$. Indeed, the N comes from the for loops involving $n = 1$ to N. In this loop there is a loop on $y \in S$ and in this loop there are roughly $|S|$ evaluation needed to compute $V_n(y)$ and $\varphi_n(y)$. Thus at each index n, there are about $|S|^2$ operations needed.

12.2 The computation of the conditional probabilities

If $\mathbf{u} := (u_0, u_1, u_2, \dots, u_N) \in O^{N+1}$ is given observation sequence and $x = (x_0, x_1, x_2, \dots x_N) \in S^{N+1}$ is a state sequence we may ask how likely is it that

$\mathbf{X}^{(N)} := (X_0, X_1, X_2, \dots X_N) = \mathbf{x}$ given $\mathbf{U}^{(N)} := (U_0, U_1, U_2, \dots, U_N) = \mathbf{u}$,[2] i.e. we want to compute,

$$\rho(\mathbf{x}|\mathbf{u}) := P\left(\mathbf{X}^{(N)} = \mathbf{x}|\mathbf{U}^{(N)} = \mathbf{u}\right) = \frac{P\left(\mathbf{X}^{(N)} = \mathbf{x}, \mathbf{U}^{(N)} = \mathbf{u}\right)}{P\left(\mathbf{U}^{(N)} = \mathbf{u}\right)}$$
$$= \frac{\nu(x_0)\, e(x_0, u_0) \prod_{k=1}^{N} p(x_{k-1}, x_k)\, e(x_k, u_k)}{\sum_{\hat{\mathbf{y}} \in S^{N+1}} \nu(\hat{y}_0)\, e(\hat{y}_0, u_0) \prod_{k=1}^{N} p(\hat{y}_{k-1}, \hat{y}_k)\, e(\hat{y}_k, u_k)}.$$

The denominator, $P\left(\mathbf{U}^{(N)} = \mathbf{u}\right)$, as written is computationally to costly to compute even for moderately large N. Theorem 12.7 gives a computationally tractable algorithm for computing $P\left(\mathbf{U}^{(N)} = \mathbf{u}\right)$ hence for computing $\rho(\mathbf{x}|\mathbf{u})$. We first introduce the following notation.

Notation 12.6 *For $0 \le n \le N$, let $\mathbf{u}^{(n)} = (u_0, u_1, u_2, \dots, u_n) \in O^{n+1}$ and similarly $\mathbf{U}^{(n)} = (U_0, U_1, U_2, \dots, U_n)$.*

Theorem 12.7. *Let $N \in \mathbb{N}$, $\mathbf{u} = (u_0, u_1, \dots, u_N) \in O^{N+1}$ be an observation sequence and define*

$$\alpha_n(x) = \mathbb{P}\left(\mathbf{U}^{(n)} = \mathbf{u}^{(n)}, X_n = x\right) \text{ for } 0 \le n \le N. \tag{12.10}$$

Then

$$\alpha_0(x) = \nu(x)\, e(x, u) = \mathbb{P}(X_0 = x)\, e(x, u) \tag{12.11}$$

and $\{\alpha_n : S \to [0,1]\}_{n=1}^{N}$ may be computed inductively using;

$$\alpha_{n+1}(y) = [\alpha_n \mathbf{P}]_y\, e(y, u_{n+1}) = e(y, u_{n+1}) \sum_{x \in S} \alpha_n(x)\, p(x, y). \tag{12.12}$$

We then have

$$P\left(\mathbf{U}^{(N)} = \mathbf{u}\right) = \sum_{x \in S} \mathbb{P}(\mathbf{U}^{(N)} = \mathbf{u}, X_N = x) = \sum_{x \in S} \alpha_N(x). \tag{12.13}$$

Proof. Equation (12.13) trivially follows from Eq. (12.10) with $n = N$ while for Eq. (12.11) we simply have,

$$\alpha_0(x) = \mathbb{P}(U_0 = u_0, X_0 = x) = \mathbb{P}(\mathcal{E}_0(x) = u_0, X_0 = x)$$
$$= \mathbb{P}(\mathcal{E}_0(x) = u_0)\, \mathbb{P}(X_0 = x) = e(x, u)\, \nu(x).$$

In order to prove Eq. (12.12), let us realize the Markov chain, $\{X_n\}_{n=0}^{\infty}$, as it is described in Eq. (5.1). That is choose $X_0 \in S$ and i.i.d. random functions $\{f_n : S \to S\}_{n=0}^{\infty}$ such that $\mathbb{P}(X_0 = x) = \nu(x)$, $p(x, y) = \mathbb{P}(f_n(x) = y)$ for all $x, y \in S$, and then take$X_{n+1} = f_n(X_n)$ where and $n \in \mathbb{N}_0$. We further assume that $\{X_0\} \cup \{f_n\}_{n=0}^{\infty} \cup \{\mathcal{E}_n\}_{n=0}^{\infty}$ are all independent in order to satisfy Assumption 12.1. With this notation and making use of all the independence we find,

$$\alpha_{n+1}(y) := \mathbb{P}\left(\mathbf{U}^{(n+1)} = \mathbf{u}^{(n+1)}, X_{n+1} = y\right)$$
$$= \sum_{x \in S} \mathbb{P}\left(\mathbf{U}^{(n+1)} = \mathbf{u}^{(n+1)}, X_{n+1} = y, X_n = x\right)$$
$$= \sum_{x \in S} \mathbb{P}\left(\mathbf{U}^{(n)} = \mathbf{u}^{(n)}, \mathcal{E}_{n+1}(y) = u_{n+1}, f_n(x) = y, X_n = x\right)$$
$$= \sum_{x \in S} \mathbb{P}\left(\mathbf{U}^{(n)} = \mathbf{u}^{(n)}, X_n = x\right) \mathbb{P}(\mathcal{E}_{n+1}(y) = u_{n+1})\, \mathbb{P}(f_n(x) = y)$$
$$= \sum_{x \in S} \alpha_n(x)\, e(y, u_{n+1})\, p(x, y) = [\alpha_n \mathbf{P}]_y\, e(y, u_{n+1}).$$

■

Complexity of the algorithm. The complexity of this algorithm is again roughly $CN|S|^2$. Indeed, there are N inductive steps in Eq. (12.12) and there are $C|S|$ operations needed to compute the sum in Eq. (12.12) and this has to be done $|S|$ times, once for each $y \in S$.

12.2.1 Exercises

Exercise 12.1. Here is a summary of the Viterbi algorithm:

1. (Initialize): $\delta_0(j) = p(j) h_j(b_0)$ for $j \in S$.
2. (Recursion): $\delta_{k+1}(j) = [\max_{i \in S} \delta_k(i) P(i, j)]\, h_j(b_{k+1})$, for $j \in S$ and
$$\psi_{k+1}(j) = \arg\max_{i \in S} \delta_k(i) P(i, j) \text{ for } k = 0, 1, 2, \dots n-1.$$
3. (Termination): $P^*(b)[= P^*(b_0, \dots b_n)] = \max_{j \in S} \delta_n(j)$ and
$$j_n^* = \arg\max_{j \in S} \delta_n(j).$$
4. (Back substitution): $j_k^* = \psi_{k+1}(j_{k+1}^*)$ for $k = n-1, n-2, \dots, 2, 1, 0$.

Show, by backward induction on k, that $p(j_k^*, j_{k+1}^*) > 0$ for $k = 0, 1, 2, \dots, n-1$.

Exercise 12.2. This problems implements the Viterbi algorithm for the "crooked casino" HMM. The underlying Markov chain has state space $S = \{F, L\}$ and transition matrix

[2] Recall that $\{U_n = \mathcal{E}_n(X_n)\}_{n=0}^{\infty}$ so that U_n depends on $\mathbf{X}$ and the $\{f_n\}$.

$$\mathbf{P} = \begin{bmatrix} .95 & .05 \\ .1 & .9 \end{bmatrix}.$$

When the Markov chain is in state F, a fair die is thrown—all six faces are equally likely; when the Markov chain is in state L a loaded die is thrown—face 6 has probability .5 and all other faces have probability 0.1. In short, the "emission probabilities" are given by the matrix

$$\mathbf{H} = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/2 \end{bmatrix}.$$

We assume that at the start the fair die is being used; that is, $p(F) = 1$ and $p(L) = 0$.

Here are 100 random rolls in the crooked casino. Use the Viterbi algorithm to predict the state sequence; that is, which of the two dice was used on each roll.

3151162464 4664424531 1321631164 1521336251 4454363165
6626566666 6511664531 3265124563 6664631636 6631623264

On the second page are the dice that are tossed on each roll. Don't look until you have done the exercise!

Here are the dice that are rolled, each listed just below the roll itself.

```
3151162464 4664424531 1321631164 1521336251 4454363165
FFFFFFFFFF FFFFFFFFFF FFFFFFFFFF FFFFFFFFFF FFFFFLLLLL
6626566666 6511664531 3265124563 6664631636 6631623264
LLLLLLLLLL LLLLLLFFFF FFFFFFFFLL LLLLLLLLLL LLLLFFFLLL
```

Part III

Martingales

13

(Sub and Super) martingales

The typical setup in this chapter is that $\{X_k, Y_k, Z_k, \dots\}_{k=0}^{\infty}$ is some collection of stochastic processes on a probability space $(\Omega, \mathbb{P})$ and we let $\mathcal{F}_n = \sigma(\{X_n, Y_n, Z_n, \dots\}_{k=0}^{n})$ for $n \in \mathbb{N}_0$ and $\mathcal{F} = \sigma(\{X_n, Y_n, Z_n, \dots\}_{k=0}^{\infty})$. Notice that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n = 0, 1, 2 \dots$. We say a stochastic process, $\{U_n\}_{n=0}^{\infty}$ is **adapted** to $\{\mathcal{F}_n\}$ provided U_n is $\mathcal{F}_n$ – measurable for all $n \in \mathbb{N}_0$, i.e. $U_n = F_n(\{X_n, Y_n, Z_n, \dots\}_{k=0}^{n})$.

Definition 13.1. *Let $X := \{X_n\}_{n=0}^{\infty}$ be an $\{\mathcal{F}_n\}$-adapted sequence of integrable random variables. Then;*

1. *X is a $\{\mathcal{F}_n\}_{n=0}^{\infty}$ – **martingale** if $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ a.s. for all $n \in \mathbb{N}_0$.*
2. *X is a $\{\mathcal{F}_n\}_{n=0}^{\infty}$ – **submartingale** if $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$ a.s. for all $n \in \mathbb{N}_0$.*
3. *X is a $\{\mathcal{F}_n\}_{n=0}^{\infty}$ – **supermartingale** if $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$ a.s. for all $n \in \mathbb{N}_0$.*

It is often fruitful to view X_n as your earnings at time n while playing some game of chance. In this interpretation, your expected earnings at time $n+1$ given the history of the game up to time n is the same, greater than, less than your earnings at time n if $X = \{X_n\}_{n=0}^{\infty}$ is a martingale, submartingale or supermartingale respectively. In this interpretation, martingales are fair games, submartingales are games which are favorable to the gambler (unfavorable to the casino), and supermartingales are games which are unfavorable to the gambler (favorable to the casino), see Example 13.39.

By induction one shows that X is a supermartingale, martingale, or submartingale iff

$$\mathbb{E}[X_n|\mathcal{F}_m] \overset{\leq}{\underset{\geq}{=}} X_m \text{ a.s for all } n \geq m, \tag{13.1}$$

to be read from top to bottom respectively. This last equation may also be expressed as

$$\mathbb{E}[X_n|\mathcal{F}_m] \overset{\leq}{\underset{\geq}{=}} X_{n\wedge m} \text{ a.s for all } m, n \in \mathbb{N}_0. \tag{13.2}$$

The reader should also note that $\mathbb{E}[X_n]$ is decreasing, constant, or increasing respectively.

13.1 (Sub and Super) martingale Examples

Example 13.2. Suppose that $\{Z_n\}_{n=0}^{\infty}$ are **independent**[1] 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, $\mathcal{F}_n^Z := \sigma(Z_0, \dots, Z_n)$. Indeed,

$$\mathbb{E}[S_{n+1} - S_n|\mathcal{F}_n] = \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \mathbb{E}Z_{n+1} = 0.$$

This same computation also shows that $\{S_n\}_{n\geq 0}$ is a submartingale if $\mathbb{E}Z_n \geq 0$ and supermartingale if $\mathbb{E}Z_n \leq 0$ for all n.

Example 13.3 (Regular martingales). If X is a random variable with $\mathbb{E}|X| < \infty$, then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ is a martingale. Indeed, by the tower property of conditional expectations,

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] = X_n \text{ a.s.}$$

A martingale of this form is called a **regular martingale.** From Proposition 3.12 we know that

$$\mathbb{E}|X_n| = \mathbb{E}|\mathbb{E}[X|\mathcal{F}_n]| \leq \mathbb{E}|X|$$

which shows that $\sup_n \mathbb{E}|X_n| \leq \mathbb{E}|X| < \infty$ for regular martingales. In the next exercise you will show that not all martingales are regular.

Exercise 13.1. Construct an example of a martingale, $\{M_n\}_{n=0}^{\infty}$ such that $\mathbb{E}|M_n| \to \infty$ as $n \to \infty$. [In particular, $\{M_n\}_{n=1}^{\infty}$ will be a martingale which is not of the form $M_n = \mathbb{E}_{\mathcal{F}_n} X$ for some $X \in L^1(P)$.] **Hint:** try taking $M_n = \sum_{k=0}^{n} Z_k$ for a judicious choice of $\{Z_k\}_{k=0}^{\infty}$ which you should take to be independent, mean zero, and having $\mathbb{E}|Z_n|$ growing rather rapidly.

Example 13.4 ("Conditioning a δ – fucntion"). Suppose that $\Omega = (0, 1]$, $\mathcal{F} = \mathcal{F}_{(0,1]}$, and $P = m$ – Lebesgue measure. We then define random variables, $\{Y_n\}_{n=1}^{\infty}$, with values in $\{0, 1\}$ so that

$$\omega = \sum_{n=1}^{\infty} \frac{Y_n(\omega)}{2^n}, \text{ i.e. } \omega = .Y_1(\omega) Y_2(\omega) Y_3(\omega) \dots$$

[1] We do not need to assume that the $\{Z_n\}_{n=0}^{\infty}$ are identically distributed here!!

is the binary expansion of ω. We then let $\mathcal{F}_n := \sigma(Y_1, \ldots, Y_n)$. A little thought shows that X is $\mathcal{F}_n$ measurable iff X is constant on each of the sub-intervals, $\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$, for $1 \le k \le 2^n$. As $\left\{\sqrt{2^n} 1_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]}\right\}_{k=1}^{2^n}$ is an orthonormal basis for this collection of functions we may conclude that

$$\begin{aligned}\mathbb{E}[X|\mathcal{F}_n](\omega) &= \sum_{k=1}^{2^n}\left[\int_0^1 X(\sigma)\sqrt{2^n}1_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]}(\sigma)\,d\sigma\right]\cdot\sqrt{2^n}1_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]}(\omega) \\ &= \sum_{k=1}^{2^n}\left[2^n\int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} X(\sigma)\,d\sigma\right]\cdot 1_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]}(\omega).\end{aligned}$$

If we formally take $X = \delta_0$ to be the Dirac delta function of 0 (with all its weight to the right of zero), then

$$M_n := \text{“}\mathbb{E}[\delta_0|\mathcal{F}_n]\text{”} = 2^n 1_{(0,2^{-n}]} \text{ for } n \in \mathbb{N}.$$

In Exercise 13.2), you are asked to verify directly that $\left\{M_n := 2^n 1_{(0,2^{-n}]}\right\}_{n=0}^{\infty}$ is a martingale.

Notice that $\mathbb{E}|M_n| = 1$ for all n. However, there is no $X \in L^1(\Omega, \mathcal{F}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{F}_n]$. To verify this last assertion, suppose such an X existed. We would then have for $2^n > k > 0$ and any $m > n$, that

$$\begin{aligned}\mathbb{E}\left[X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{F}_m} X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] \\ &= \mathbb{E}\left[M_m : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = 0.\end{aligned}$$

Using $\mathbb{E}[X : A] = 0$ for all A in the π – system, $\mathcal{Q} := \cup_{n=1}^{\infty}\left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] : 0 \le k < 2^n\right\}$, an application of the “π – λ theorem” shows $\mathbb{E}[X : A] = 0$ for all $A \in \sigma(\mathcal{Q}) = \mathcal{F}$. By more general measure theoretic arguments, this suffices to show $X = 0$ a.s. which is impossible since $1 = \mathbb{E}M_n = \mathbb{E}X$.

Moral: not all L^1 – bounded martingales are regular. It is also worth noting that pointwise $M_\infty = \lim_{n\to\infty} M_n = 0$ yet

$$\lim_{n\to\infty}\mathbb{E}M_n = \lim_{n\to\infty} 1 \ne 0 = \mathbb{E}M_\infty = \mathbb{E}\lim_{n\to\infty} M_n.$$

Exercise 13.2. Show that $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}_0$ as defined in Example 13.4 is a martingale.

Notation 13.5 *Given a sequence $\{Z_k\}_{k=0}^{\infty}$, let $\Delta_k Z := Z_k - Z_{k-1}$ for $k = 1, 2, \ldots$. and by convention let $\Delta_0 Z = Z_0$.*

The next lemma is the “differential” characterization of (sub or super) martingales.

Lemma 13.6 (Submartingale Increments). *Let $X := \{X_n\}_{n=0}^{\infty}$ be an adapted process of integrable random variables on a filtered probability space, $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ and let $d_n := \Delta_n X := X_n - X_{n-1}$. Then X is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{F}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{F}_n] \ge 0$ or $\mathbb{E}[d_{n+1}|\mathcal{F}_n] \le 0$ respectively) for all $n \in \mathbb{N}_0$.*

Conversely if $\{d_n\}_{n=1}^{\infty}$ is an adapted sequence of integrable random variables and X_0 is a $\mathcal{F}_0$ -measurable integrable random variable. Then $X_n = X_0 + \sum_{j=1}^{n} d_j$ is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{F}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{F}_n] \ge 0$ or $\mathbb{E}[d_{n+1}|\mathcal{F}_n] \le 0$ respectively) for all $n \in \mathbb{N}$.

Proof. We prove the assertions for martingales only, the other all being similar. Clearly X is a martingale iff

$$0 = \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n = \mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = \mathbb{E}[d_{n+1}|\mathcal{F}_n].$$

The second assertion is an easy consequence of the first assertion. ■

Example 13.7. Suppose that $\{Y_n\}_{n=1}^{\infty}$ is any adapted sequence of integrable random variables, X_0 is any integrable $\mathcal{F}_0$ – random variable, and

$$X_n := X_0 + \sum_{k=1}^{n}\left(Y_k - \mathbb{E}[Y_k|\mathcal{F}_{k-1}]\right) \text{ for } n \ge 1.$$

Then $\{X_n\}_{n=0}^{\infty}$ is a martingale. Indeed, $\Delta_n X = Y_n - \mathbb{E}[Y_n|\mathcal{F}_{n-1}]$ and hence

$$\begin{aligned}\mathbb{E}[\Delta_n X|\mathcal{F}_{n-1}] &= \mathbb{E}[Y_n - \mathbb{E}[Y_n|\mathcal{F}_{n-1}]|\mathcal{F}_{n-1}] \\ &= \mathbb{E}[Y_n|\mathcal{F}_{n-1}] - \mathbb{E}[Y_n|\mathcal{F}_{n-1}] = 0 \text{ for } n \ge 1.\end{aligned}$$

Example 13.8. Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a sequence of independent integrable random variables, $X_n = Z_0 \ldots Z_n$, and $\mathcal{F}_n := \sigma(Z_0, \cdots, Z_n)$. (Observe that $\mathbb{E}|X_n| = \prod_{k=0}^{n}\mathbb{E}|Z_k| < \infty$.) Since

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n Z_{n+1}|\mathcal{F}_n] = X_n\mathbb{E}[Z_{n+1}|\mathcal{F}_n] = X_n \cdot \mathbb{E}[Z_{n+1}] \text{ a.s.},$$

it follows that $\{X_n\}_{n=0}^{\infty}$ is a martingale if $\mathbb{E}Z_n = 1$. If we further assume, for all n, that $Z_n \ge 0$ so that $X_n \ge 0$, then $\{X_n\}_{n=0}^{\infty}$ is a supermartingale (submartingale) provided $\mathbb{E}Z_n \le 1$ ($\mathbb{E}Z_n \ge 1$) for all n.

Let us specialize the above example even more by taking $Z_n \overset{d}{=} p + U$ where $p \ge 0$ and U is the uniform distribution on $[0, 1]$. In this case we have by the strong law of large numbers that

$$\frac{1}{n}\ln X_n = \frac{1}{n}\sum_{k=0}^{n}\ln Z_k \to \mathbb{E}\left[\ln\left(p+U\right)\right] \text{ a.s.} \tag{13.3}$$

An elementary computation shows

$$\begin{aligned}\mathbb{E}\left[\ln\left(p+U\right)\right] &= \int_0^1 \ln\left(p+x\right)dx = \int_p^{p+1}\ln\left(p+x\right)dx \\ &= \left(x\ln x - x\right)_{x=p}^{x=p+1} = \left(p+1\right)\ln\left(p+1\right) - p\ln p - 1\end{aligned}$$

The function $f\left(p\right) := \mathbb{E}\left[\ln\left(p+U\right)\right]$ has a zero at $p = p_c \cong 0.542\,21$ and

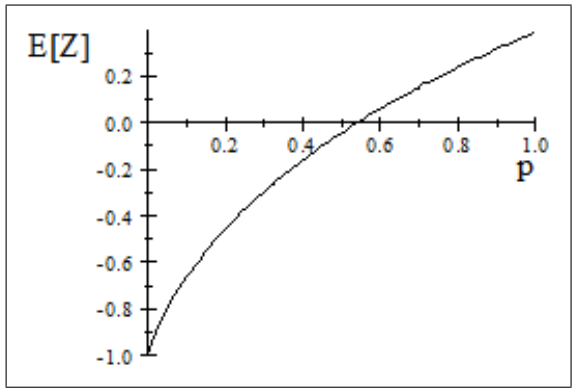


Fig. 13.1. The graph of $\mathbb{E}\left[\ln\left(p+U\right)\right]$ as a function of p. This function has a zero at $p = p_c \cong 0.542\,21$.

$f\left(p\right) < 0$ for $p < p_c$ while $f\left(p\right) > 0$ for $p > p_c$, see Figure 13.1. Combining these observations with Eq. (13.3) implies,

$$X_n \to \lim_{n\to\infty}\exp\left(n\mathbb{E}\left[\ln\left(p+U\right)\right]\right) = \begin{cases} 0 & \text{if } p < p_c \\ ? & \text{if } p = p_c \\ \infty & \text{if } p > p_c \end{cases} \text{ a.s.}$$

Notice that $\mathbb{E}Z_n = p + 1/2$ and therefore X_n is a martingale precisely when $p = 1/2$ and is a sub-martingale for $p > 1/2$. So for $1/2 < p < p_c$, $\{X_n\}_{n=1}^{\infty}$ is a positive sub-martingale, $\mathbb{E}X_n = \left(p+1/2\right)^{n+1} \to \infty$ yet $\lim_{n\to\infty} X_n = 0$ a.s.

Definition 13.9. *We say $\{C_n : \Omega \to S\}_{n=1}^{\infty}$ is* ***predictable or previsible*** *if each C_n is $\mathcal{F}_{n-1}$– measurable for all $n \in \mathbb{N}$, i.e. $C_n = F_n\left(\{X_n, Y_n, Z_n, \dots\}_{k=0}^{n-1}\right)$.*

Lemma 13.10 (Doob Decomposition). *Each adapted sequence, $\{Z_n\}_{n=0}^{\infty}$, of integrable random variables has a unique decomposition,*

$$Z_n = M_n + A_n \tag{13.4}$$

where $\{M_n\}_{n=0}^{\infty}$ is a martingale and A_n is a predictable process such that $A_0 = 0$. Moreover this decomposition is given by $A_0 = 0$,

$$A_n := \sum_{k=1}^{n}\mathbb{E}_{\mathcal{F}_{k-1}}\left[\Delta_k Z\right] \text{ for } n \geq 1 \tag{13.5}$$

and

$$M_n = Z_n - A_n = Z_n - \sum_{k=1}^{n}\mathbb{E}_{\mathcal{F}_{k-1}}\left[\Delta_k Z\right] \tag{13.6}$$

$$= Z_0 + \sum_{k=1}^{n}\left(Z_k - \mathbb{E}_{\mathcal{F}_{k-1}}Z_k\right). \tag{13.7}$$

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. (13.4), then

$$\mathbb{E}_{\mathcal{F}_n}\left[\Delta_{n+1}Z\right] = \mathbb{E}_{\mathcal{F}_n}\left[\Delta_{n+1}M + \Delta_{n+1}A\right] = \Delta_{n+1}A \tag{13.8}$$

wherein we have used M is a martingale and A is predictable so that $\mathbb{E}_{\mathcal{F}_n}\left[\Delta_{n+1}M\right] = 0$ and $\mathbb{E}_{\mathcal{F}_n}\left[\Delta_{n+1}A\right] = \Delta_{n+1}A$. Hence we must define, for $m \geq 1$,

$$A_n := \sum_{k=1}^{n}\Delta_k A = \sum_{k=1}^{n}\mathbb{E}_{\mathcal{F}_{k-1}}\left[\Delta_k Z\right]$$

which is a predictable process. This proves the uniqueness of the decomposition and the validity of Eq. (13.5). Moreover if Z is a sub-martingale then $\mathbb{E}_{\mathcal{F}_{k-1}}\left[\Delta_k Z\right] \geq 0$ from which it follows that $\{A_n\}_{n=0}^{\infty}$ is increasing.

For existence, we define A_n by Eq. (13.5) and $M_n := Z_n - A_n$ as in Eq. (13.6). From Eq. (13.5) we have $\Delta_n A = \mathbb{E}\left[\Delta_n Z|\mathcal{F}_{n-1}\right]$ and therefore,

$$\mathbb{E}\left[\Delta_n M|\mathcal{F}_{n-1}\right] = \mathbb{E}\left[\Delta_n Z - \Delta_n A|\mathcal{F}_{n-1}\right] = \mathbb{E}\left[\Delta_n Z|\mathcal{F}_{n-1}\right] - \mathbb{E}\left[\Delta_n Z|\mathcal{F}_{n-1}\right] = 0$$

which shows that M is indeed a martingale.

■

Exercise 13.3. Suppose that $\{Z_n\}_{n=0}^{\infty}$ are independent random variables such that $\sigma^2 := \mathbb{E}Z_n^2 < \infty$ and $\mathbb{E}Z_n = 0$ for all $n \geq 1$. As in Example 13.2, let $S_n :=$

$\sum_{k=0}^{n} Z_k$ be the martingale relative to the filtration, $\mathcal{F}_n^Z := \sigma(Z_0, \ldots, Z_n)$. Show

$$M_n := S_n^2 - \sigma^2 n \text{ for } n \in \mathbb{N}_0$$

is a martingale. [Hint, make use of the basic properties of conditional expectations.]

Exercise 13.4 (Quadratic Variation). Suppose $\{M_n\}_{n=0}^{\infty}$ is a square integrable martingale. Show;

1. $\mathbb{E}\left[M_{n+1}^2 - M_n^2 | \mathcal{F}_n\right] = \mathbb{E}\left[(M_{n+1} - M_n)^2 | \mathcal{F}_n\right]$. Conclude from this that the Doob decomposition of M_n^2 is of the form,
$$M_n^2 = N_n + A_n$$
where
$$A_n := \sum_{1 \le k \le n} \mathbb{E}\left[(\Delta_k M)^2 | \mathcal{F}_{k-1}\right] = \sum_{1 \le k \le n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}\right].$$
In particular we see from this that M_n^2 is a sub-martingale.
2. Show that N_n above may be expressed as, $N_0 = M_0^2$ and for $n \ge 1$,
$$N_n = M_0^2 + 2\sum_{k=1}^{n} M_{k-1}\Delta_k M + \sum_{k=1}^{n}\left((\Delta_k M)^2 - \mathbb{E}\left[(\Delta_k M)^2 | \mathcal{F}_{k-1}\right]\right).$$
[From Example 13.7 and Proposition 13.29 below one may see directly that the expression above is a martingale.]
3. If we further assume that $M_k - M_{k-1}$ is independent of $\mathcal{F}_{k-1}$ for all $k = 1, 2, \ldots$, explain why,
$$A_n = \sum_{1 \le k \le n} \mathbb{E}(M_k - M_{k-1})^2.$$

Exercise 13.5. Suppose that $\{X_n\}_{n=0}^{\infty}$ is a Markov chain on S with one step transition probabilities, $\mathbf{P}$, and $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$. Let $f : S \to \mathbb{R}$ be a bounded (for simplicity) function and set $g(x) = (\mathbf{P}f)(x) - f(x)$. Show

$$M_n := f(X_n) - \sum_{0 \le j < n} g(X_j) \text{ for } n \in \mathbb{N}_0$$

is a martingale.

Proposition 13.11 (Markov Chains and Martingales). *Suppose that S is a finite or countable set, $\mathbf{P}$ is a Markov matrix on S, and $\{X_n\}_{n=0}^{\infty}$ is the associated Markov Chain. If $f : S \to \mathbb{R}$ is a function such that either $f \ge 0$ or $\mathbb{E}|f(X_n)| < \infty$ for all $n \in \mathbb{N}_0$ and ($\mathbf{P}f \le f$) $\mathbf{P}f = f$ then $Z_n = f(X_n)$ is a (sub-martingale) martingale. More generally, if $f : \mathbb{N}_0 \times S \to \mathbb{R}$ is a such that either $f \ge 0$ or $\mathbb{E}|f(n, X_n)| < \infty$ for all $n \in \mathbb{N}_0$ and ($\mathbf{P}f(n+1, \cdot) \le f(n, \cdot)$) $\mathbf{P}f(n+1, \cdot) = f(n, \cdot)$, then $\{Z_n := f(n, X_n)\}_{n=0}^{\infty}$ is a (sub-martingale) martingale.*

Proof. Using the Markov property and the definition of $\mathbf{P}$ we find,[2]

$$\begin{aligned}\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= \mathbb{E}[f(n+1, X_{n+1})|\mathcal{F}_n] \\ &= \mathbb{E}[f(n+1, X_{n+1})|X_n] = [\mathbf{P}f(n+1, \cdot)](X_n).\end{aligned}$$

The latter expression is (less than or equal) equal to Z_n if ($\mathbf{P}f(n+1, \cdot) \le f(n \cdot)$) $\mathbf{P}f(n+1, \cdot) = f(n, \cdot)$ for all $n \ge 0$. ■

Remark 13.12. One way to find solutions to the equation $\mathbf{P}f(n+1, \cdot) = f(n, \cdot)$, at least for $0 \le n \le T$ for some $T < \infty$, is to let $g : S \to \mathbb{R}$ be an arbitrary function and $T \in \mathbb{N}$ be given and then define

$$f(n, y) := \left(\mathbf{P}^{T-n} g\right)(y) \text{ for } 0 \le n \le T.$$

Then $\mathbf{P}f(n+1, \cdot) = \mathbf{P}\left(\mathbf{P}^{T-n-1} g\right) = \mathbf{P}^{T-n} g = f(n, \cdot)$ and we will have that

$$Z_n = f(n, X_n) = \left(\mathbf{P}^{T-n} g\right)(X_n)$$

is a martingale for $0 \le n \le T$. If $f(n, \cdot)$ satisfies $\mathbf{P}f(n+1, \cdot) = f(n, \cdot)$ for all n then we must have, with $f_0 := f(0, \cdot)$,

$$f(n, \cdot) = \mathbf{P}^{-n} f_0$$

where $\mathbf{P}^{-1} g$ denotes a function h solving $\mathbf{P}h = g$. In general $\mathbf{P}$ is not invertible and hence there may be no solution to $\mathbf{P}h = g$ or there might be many solutions.

Example 13.13. Let $S = \mathbb{Z}$, $S_n = X_0 + X_1 + \cdots + X_n$, where $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with $\mathbb{P}(X_i = 1) = p \in (0, 1)$ and $\mathbb{P}(X_i = -1) = q := 1 - p$, and X_0 is S –valued random variable independent of $\{X_i\}_{i=1}^{\infty}$. Recall that $\{S_n\}_{n=0}^{\infty}$ is a time homogeneous Markov chain with transition kernel determined by $\mathbf{P}f(x) = pf(x+1) + qf(x-1)$. As we have seen if $f(x) = a + b(q/p)^x$, then $\mathbf{P}f = f$ and therefore, by Proposition 13.11,

[2] Recall from Proposition 3.10 that $\mathbb{E}[f(n+1, X_{n+1})|X_n] = h(X_n)$ where

$$\begin{aligned}h(x) &= \mathbb{E}[f(n+1, X_{n+1})|X_n = x] \\ &= \sum_{y \in S} p(x, y) f(n+1, y) = [\mathbf{P}f(n+1, \cdot)](x).\end{aligned}$$

$$M_n = a + b\,(q/p)^{S_n}$$

is a martingale for all $a, b \in \mathbb{R}$. This is easily verified directly as well;

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_n}\left(\frac{q}{p}\right)^{S_{n+1}} &= \mathbb{E}_{\mathcal{F}_n}\left(\frac{q}{p}\right)^{S_n+X_{n+1}} = \left(\frac{q}{p}\right)^{S_n}\mathbb{E}_{\mathcal{F}_n}\left(\frac{q}{p}\right)^{X_{n+1}} \\
&= \left(\frac{q}{p}\right)^{S_n}\mathbb{E}\left(\frac{q}{p}\right)^{X_{n+1}} = \left(\frac{q}{p}\right)^{S_n}\cdot\left[\left(\frac{q}{p}\right)^{1}p + \left(\frac{q}{p}\right)^{-1}q\right] \\
&= \left(\frac{q}{p}\right)^{S_n}\cdot[q+p] = \left(\frac{q}{p}\right)^{S_n}.
\end{aligned}$$

Example 13.14. We can generalize the above example by observing for $\lambda \neq 0$ that $\mathbf{P}[x \to \lambda^x] = \left(p\lambda + q\lambda^{-1}\right)\lambda^x$. Thus it follows that we may set $\mathbf{P}^{-1}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-1}\lambda^x$ and therefore conclude that

$$f(n,x) := \mathbf{P}^{-n}\lambda^x = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^x$$

satisfies $\mathbf{P}f(n+1,\cdot) = f(n,\cdot)$. So if we suppose that X_0 is a bounded so that S_n is bounded for all n, then using Proposition 13.11, it follows that $\left\{M_n = f(n,S_n) = \left(p\lambda + q\lambda^{-1}\right)^{-n}\lambda^{S_n}\right\}_{n\geq 0}$ is a martingale for all $\lambda \neq 0$. To recover the special case in Example 13.13 we choose λ so that $p\lambda + q\lambda^{-1} = 1$. This quadratic equation has 1 and $q/p = (1-p)/p$ as solutions as one easily verifies.

Exercise 13.6. For $\theta \in \mathbb{R}$ let

$$f_\theta(n,x) := \mathbf{P}^{-n}e^{\theta x} = \left(pe^{\theta} + qe^{-\theta}\right)^{-n}e^{\theta x}$$

so that $\mathbf{P}f_\theta(n+1,\cdot) = f_\theta(n,\cdot)$ for all $\theta \in \mathbb{R}$. Compute;

1. $f_\theta^{(k)}(n,x) := \left(\frac{d}{d\theta}\right)^k f_\theta(n,x)$ for $k = 1, 2$.
2. Use your results to show,

$$M_n^{(1)} := S_n - n(p-q)$$
and
$$M_n^{(2)} := (S_n - n(p-q))^2 - 4npq$$

are martingales.

(If you are ambitious you might also find $M_n^{(3)}$.)

Example 13.15 (Polya urns). Let us consider an urn that contains red and green balls. Let (R_n, G_n) denote the number of red and green balls (respectively) in the urn at time n. If at a given time 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. So (R_n, G_n) is a Markov process with transition probabilities determined by,

$$\begin{aligned}
\mathbb{P}\left((R_{n+1},G_{n+1}) = (r+c,g)\,|\,(R_n,G_n) = (r,g)\right) &= \frac{r}{r+g} \text{ and} \\
\mathbb{P}\left((R_{n+1},G_{n+1}) = (r,g+c)\,|\,(R_n,G_n) = (r,g)\right) &= \frac{g}{r+g}.
\end{aligned} \tag{13.9}$$

We now let

$$\mathcal{F}_n := \sigma\left((R_k,G_k) : k \leq n\right) = \sigma\left(X_k : k \leq n\right)$$

and recall that by the Markov property along with Proposition 3.10 (the basic formula for computing conditional expectations) we have,

$$\mathbb{E}\left[f(R_{n+1},G_{n+1})\,|\mathcal{F}_n\right] = \mathbb{E}\left[f(R_{n+1},G_{n+1})\,|\,(R_n,G_n)\right] = h(R_n,G_n)$$

where

$$\begin{aligned}
h(r,g) &:= \mathbb{E}\left[f(R_{n+1},G_{n+1})\,|\,(R_n,G_n) = (r,g)\right] \\
&= \frac{r}{r+g}f(r+c,g) + \frac{g}{r+g}f(r,g+c).
\end{aligned}$$

Thus we have shown,

$$\mathbb{E}\left[f(R_{n+1},G_{n+1})\,|\mathcal{F}_n\right] = \frac{R_n}{R_n+G_n}f(R_n+c,G_n) + \frac{G_n}{R_n+G_n}f(R_n,G_n+c). \tag{13.10}$$

Next 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 an $\mathcal{F}_n$ – martingale. Indeed, using Eq. (13.10) with $f(r,g) = \frac{g}{r+g}$ we learn

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[X_{n+1}|\,(R_n,G_n)\right] \\
&= \frac{R_n}{R_n+G_n}\cdot\frac{G_n}{R_n+c+G_n} + \frac{G_n}{R_n+G_n}\cdot\frac{G_n+c}{R_n+G_n+c} \\
&= \frac{G_n}{R_n+G_n}\cdot\frac{R_n+G_n+c}{R_n+G_n+c} = X_n.
\end{aligned}$$

Remark 13.16. The previous example is in fact a special case of Proposition 13.11. From Eq. (13.9) (or see Eq. (13.10)) the Markov matrix $\mathbf{P}$ associated to the chain in Example 13.15 is determined by,

$$(\mathbf{P}f)(r,g) = \frac{r}{r+g} f(r+c,g) + \frac{g}{r+g} f(r,g+c).$$

From this expression with $f(r,g) = \frac{g}{r+g}$ we find,

$$\begin{aligned}(\mathbf{P}f)(r,g) &= \frac{r}{r+g}\frac{g}{r+c+g} + \frac{g}{r+g}\frac{g+c}{r+c+g} \\ &= \frac{1}{(r+g)(r+c+g)} \cdot [rg + g(g+c)] = \frac{g}{r+g} = f(r,g).\end{aligned}$$

An application of Proposition 13.11 now implies $X_n = f(R_n, G_n) = \frac{G_n}{R_n+G_n}$ is necessarily a martingale.

13.2 Jensen's and Hölder's Inequalities

Theorem 13.17 (Jensen's Inequality). *Suppose that $(\Omega, \mathbb{P})$ is a probability space, $-\infty \le a < b \le \infty$, and $\varphi : (a,b) \to \mathbb{R}$ is a convex function, (i.e. $\varphi''(x) \ge 0$ on (a,b)). If $f : \Omega \to (a,b)$ is a random variable with $\mathbb{E}|f| < \infty$, then*

$$\varphi(\mathbb{E}f) \le \mathbb{E}[\varphi(f)].$$

[Here it may happen that $\varphi \circ f \notin L^1(\mathbb{P})$, in this case it is claimed that $\varphi \circ f$ is integrable in the extended sense and $\mathbb{E}[\varphi(f)] = \infty$.]

Proof. Let $t = \mathbb{E}f \in (a,b)$ and let $\beta \in \mathbb{R}$ ($\beta = \dot\varphi(t)$ when $\dot\varphi(t)$ exists), be such that

$$\varphi(s) - \varphi(t) \ge \beta(s-t) \text{ for all } s \in (a,b), \tag{13.11}$$

see Figure 13.2. Then integrating the inequality, $\varphi(f) - \varphi(t) \ge \beta(f-t)$, implies that

$$0 \le \mathbb{E}\varphi(f) - \varphi(t) = \mathbb{E}\varphi(f) - \varphi(\mathbb{E}f).$$

Moreover, if $\varphi(f)$ is not integrable, then $\varphi(f) \ge \varphi(t) + \beta(f-t)$ which shows that negative part of $\varphi(f)$ is integrable. Therefore, $\mathbb{E}\varphi(f) = \infty$ in this case. ■

Example 13.18. Since e^x for $x \in \mathbb{R}$, $-\ln x$ for $x > 0$, and x^p for $x \ge 0$ and $p \ge 1$ are all convex functions, we have the following inequalities

$$\begin{aligned}\exp(\mathbb{E}f) &\le \mathbb{E}e^f, \\ \mathbb{E}\log(|f|) &\le \log(\mathbb{E}|f|)\end{aligned} \tag{13.12}$$

and for $p \ge 1$,

$$|\mathbb{E}f|^p \le (\mathbb{E}|f|)^p \le \mathbb{E}|f|^p.$$

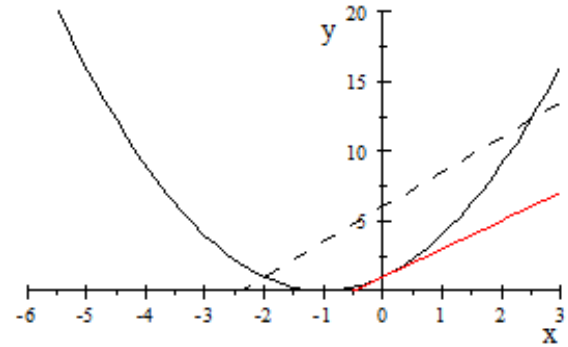


Fig. 13.2. A convex function, φ, along with a cord and a tangent line. Notice that the tangent line is always below φ and the cord lies above φ between the points of intersection of the cord with the graph of φ.

Example 13.19. As a special case of Eq. (13.12), if $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \le \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}. \tag{13.13}$$

Indeed, we have applied Eq. (13.12) with $\Omega = \{1, 2, \ldots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i}\delta_i$ and $f(i) := \ln s_i^{p_i}$. As a special case of Eq. (13.13), 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

$$st \le \frac{1}{p}s^p + \frac{1}{q}t^q. \tag{13.14}$$

(When $p = q = 1/2$, the inequality in Eq. (13.14) follows from the inequality, $0 \le (s-t)^2$.)

As another special case of Eq. (13.13), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \ldots a_n} \le \frac{1}{n}\sum_{i=1}^n a_i. \tag{13.15}$$

Example 13.20. Let $(\Omega, \mathbb{P})$ be a probability space, $0 < p < q < \infty$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then by Jensen's inequality,

$$(\mathbb{E}|f|^p)^{q/p} \le \mathbb{E}(|f|^p)^{q/p} = \mathbb{E}|f|^q$$

from which it follows that $\|f\|_p \le \|f\|_q$. In particular, $L^p(\mathbb{P}) \subset L^q(\mathbb{P})$ for all $0 < p < q < \infty$.

Theorem 13.21 (Hölder's inequality). *Suppose that $1 \le p \le \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \le \|f\|_p \cdot \|g\|_q. \tag{13.16}$$

Assuming $p \in (1,\infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (13.16) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p \quad a.s. \tag{13.17}$$

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 ∞ or $\|g\|_p = 0$ or ∞, Eq. (13.16) 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. (13.14) gives,

$$\frac{|fg|}{\|f\|_p \|g\|_q} \le \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{13.18}$$

with equality iff $|g/\|g\|_q| = |f|^{p-1} / \|f\|_p^{(p-1)} = |f|^{p/q} / \|f\|_p^{p/q}$, i.e. $|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p$. Integrating Eq. (13.18) implies

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \le \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (13.17) holds. The proof is finished since it is easily checked that equality holds in Eq. (13.16) when $|f|^p = c\,|g|^q$ of $|g|^q = c\,|f|^p$ for some constant c. ■

Theorem 13.22 (Conditional Jensen's inequality). *Let $(\Omega, \mathbb{P})$ be a probability space, $-\infty \le a < b \le \infty$, and $\varphi : (a,b) \to \mathbb{R}$ be a convex function. Assume $f \in L^1(\Omega, \mathbb{P})$ is a random variable satisfying, $f \in (a,b)$ a.s. and $\varphi(f) \in L^1(\Omega, \mathbb{P})$. Then for any sigma-algebra (of information) on Ω,*

$$\varphi(\mathbb{E}_{\mathcal{G}} f) \le \mathbb{E}_{\mathcal{G}} [\varphi(f)] \quad a.s. \tag{13.19}$$

and

$$\mathbb{E}[\varphi(\mathbb{E}_{\mathcal{G}} f)] \le \mathbb{E}[\varphi(f)]. \tag{13.20}$$

where $\mathbb{E}_{\mathcal{G}} f := \mathbb{E}[f|\mathcal{G}]$.

Proof. Not completely rigorous proof. Assume that φ is C^1 in which case Eq. (13.11) becomes,

$$\varphi(s) - \varphi(t) \ge \dot{\varphi}(t)(s-t) \text{ for all } s, t \in (a,b). \tag{13.21}$$

We now take $s = f$ and $t = \mathbb{E}_{\mathcal{G}} f$ in this inequality to find,

$$\varphi(f) - \varphi(\mathbb{E}_{\mathcal{G}} f) \ge \dot{\varphi}(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f).$$

The result now follows by taking $\mathbb{E}_{\mathcal{G}}$ of both sides of this inequality to learn,

$$\begin{aligned} &\mathbb{E}_{\mathcal{G}} \varphi(f) - \varphi(\mathbb{E}_{\mathcal{G}} f) \\ &\quad = \mathbb{E}_{\mathcal{G}} [\varphi(f) - \varphi(\mathbb{E}_{\mathcal{G}} f)] \\ &\quad \ge \mathbb{E}_{\mathcal{G}} [\dot{\varphi}(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f)] = \dot{\varphi}(\mathbb{E}_{\mathcal{G}} f) \cdot \mathbb{E}_{\mathcal{G}} [(f - \mathbb{E}_{\mathcal{G}} f)] = 0. \end{aligned}$$

The technical problem with this argument is the justification that $\mathbb{E}_{\mathcal{G}} [\varphi'(\mathbb{E}_{\mathcal{G}} f)(f - \mathbb{E}_{\mathcal{G}} f)] = \varphi'(\mathbb{E}_{\mathcal{G}} f) \mathbb{E}_{\mathcal{G}} [(f - \mathbb{E}_{\mathcal{G}} f)]$ since there is no reason for φ' to be a bounded function. The honest proof below circumvents this technical detail.

***Honest proof.** Let $\Lambda := \mathbb{Q} \cap (a,b)$ – a countable dense subset of (a,b). By the basic properties of convex functions,

$$\varphi(s) \ge \varphi(t) + \varphi'_-(t)(s-t) \text{ for all for all } t, s \in (a,b), \tag{13.22}$$

where $\varphi'_-(t)$ is the left hand derivative of φ at t. Taking $s = f$ and then taking conditional expectations imply,

$$\mathbb{E}_{\mathcal{G}} [\varphi(f)] \ge \mathbb{E}_{\mathcal{G}} \left[\varphi(t) + \varphi'_-(t)(f-t)\right] = \varphi(t) + \varphi'_-(t)(\mathbb{E}_{\mathcal{G}} f - t) \text{ a.s.} \tag{13.23}$$

Since this is true for all $t \in (a,b)$ (and hence all t in the countable set, Λ) we may conclude that

$$\mathbb{E}_{\mathcal{G}} [\varphi(f)] \ge \sup_{t \in \Lambda} \left[\varphi(t) + \varphi'_-(t)(\mathbb{E}_{\mathcal{G}} f - t)\right] \text{ a.s.}$$

Since $\mathbb{E}_{\mathcal{G}} f \in (a,b)$ we may conclude that

$$\sup_{t \in \Lambda} \left[\varphi(t) + \varphi'_-(t)(\mathbb{E}_{\mathcal{G}} f - t)\right] = \varphi(\mathbb{E}_{\mathcal{G}} f) \text{ a.s.}$$

Combining the last two estimates proves Eq. (13.19).

From Eq. (13.19) and Eq. (13.22) with $s = \mathbb{E}_{\mathcal{G}} f$ and $t \in (a,b)$ fixed we find,

$$\varphi(t) + \varphi'_-(t)(\mathbb{E}_{\mathcal{G}} f - t) \le \varphi(\mathbb{E}_{\mathcal{G}} f) \le \mathbb{E}_{\mathcal{G}} [\varphi(f)]. \tag{13.24}$$

Therefore

$$|\varphi(\mathbb{E}_{\mathcal{G}} f)| \le |\mathbb{E}_{\mathcal{G}} [\varphi(f)]| \vee \left|\varphi(t) + \varphi'_-(t)(\mathbb{E}_{\mathcal{G}} f - t)\right| \in L^1(\Omega, \mathcal{G}, \mathbb{P}) \tag{13.25}$$

which implies that $\varphi(\mathbb{E}_{\mathcal{G}} f) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. Taking expectations of Eq. (13.19) is now allowed and immediately gives Eq. (13.20). ■

Corollary 13.23. *If $\{X_n\}_{n=0}^\infty$ is a martingale and φ is a convex function such that $\mathbb{E}[\varphi(X_n)] < \infty$ for all n, then$\{\varphi(X_n)\}_{n=0}^\infty$ is a sub-martingale. [Typical examples for φ are $\varphi(x) = |x|^p$ for $p \geq 1$.]*

Proof. As $\{X_n\}_{n=0}^\infty$ is a martingale, $X_n = \mathbb{E}_{\mathcal{F}_n} X_{n+1}$. Applying φ to both sides of this equation and then using the conditional form of Jensen's inequality then shows,

$$\varphi(X_n) = \varphi(\mathbb{E}_{\mathcal{F}_n} X_{n+1}) \leq \mathbb{E}_{\mathcal{F}_n}[\varphi(X_{n+1})].$$

■

13.3 Stochastic Integrals and Optional Stopping

Notation 13.24 *Suppose that $\{c_n\}_{n=1}^\infty$ and $\{x_n\}_{n=0}^\infty$ are two sequences of numbers, let $c \cdot \Delta x = \{(c \cdot \Delta x)_n\}_{n \in \mathbb{N}_0}$ denote the sequence of numbers defined by $(c \cdot \Delta x)_0 = 0$ and*

$$(c \cdot \Delta x)_n = \sum_{j=1}^n c_j (x_j - x_{j-1}) = \sum_{j=1}^n c_j \Delta_j x \text{ for } n \geq 1.$$

(For convenience of notation later we will interpret $\sum_{j=1}^0 c_j \Delta_j x = 0$.)

Interpretation. Suppose that x_j represents the value of a stock on the time interval, $[j, j+1)$.[3] For $j \in \mathbb{N}$, let c_j denote the amount of stock you hold in the interval $(j-1, j]$ so that during this period you make a profit (or loss) of $c_j \Delta_j x = c_j x_j - c_j x_{j-1}$, and therefore $(c \cdot \Delta x)_n = \sum_{j=1}^n c_j \Delta_j x$ represents your profit (or loss) from time 0 to time n. For example if you want to buy 5 shares of the stock just after time, $n = 3$, and then sell them all at time just after $n = 9$, you would take $c_k = 5 \cdot 1_{3<k\leq 9}$ so that

$$(c \cdot \Delta x)_9 = 5 \cdot \sum_{3<k\leq 9} \Delta_k x = 5 \cdot (x_9 - x_3)$$

would represent your profit (loss) for this transaction. The next example formalizes this observation.

Example 13.25 (Coin flip game). Suppose that $\{Z_n\}_{n=1}^\infty$ are i.i.d. Bernoulli random variables with $\mathbb{P}(Z_n = \pm 1) = \frac{1}{2}$. Think of $Z_n = 1$ or -1 depending on whether the nth - flip of a coin is heads or tails respectively. Suppose that C_n is your bet on the nth coin flip and you loose your bet if $Z_n = -1$ and you double your bet if $Z_n = +1$. If $\{W_n^C\}_{n=0}^\infty$ represents your wealth just after the nth flip of the coin then $\Delta_n W^C = W_n^C - W_{n-1}^C = C_n Z_n$. [If $Z_n = 1$ your wealth goes up by C_n which represents your net winning of $2C_n$ (double your bet) - C_n (your bet) before the nth flip.] Thus

$$W_n^C = W_0 + C_1 Z_1 + \cdots + C_n Z_n.$$

If we let $X_n := W_n^1$ (i.e. take $C_n = 1$ so that we bet \$1) on every game, we will have

$$X_n = X_0 + Z_1 + \cdots + Z_n.$$

With this notation we see that

$$W_n^C = W_0 + [C \cdot \Delta X]_n.$$

[This example is generalized in Example 13.30 below.]

Example 13.26. If $\sigma, \tau \in \bar{\mathbb{N}}_0$ with $0 \leq \sigma \leq \tau$ and $c_n := 1_{\sigma<n\leq\tau}$, then

$$(c \cdot \Delta x)_n = x_{\tau\wedge n} - x_{\sigma\wedge n} \text{ for all } n \in \mathbb{N}_0.$$

[This is easily proved by considering three case, $n \leq \sigma$, $\sigma \leq n \leq \tau$, and $n \geq \tau$.] More generally if $\sigma, \tau \in \bar{\mathbb{N}}_0$ are arbitrary and $c_n := 1_{\sigma<n\leq\tau}$, then $c_n := 1_{\sigma\wedge\tau<n\leq\tau}$ and therefore

$$(c \cdot \Delta x)_n = x_{\tau\wedge n} - x_{\sigma\wedge\tau\wedge n}.$$

Notation 13.27 (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, $1_{\sigma<n\leq\tau}$.

Lemma 13.28. *If τ is a stopping time, then the processes, $f_n := 1_{\tau\leq n}$, and $f_n := 1_{\tau=n}$ are adapted and $f_n := 1_{\tau<n}$ and $f_n = 1_{n\leq\tau}$ are predictable. Moreover, if σ and τ 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 $\mathcal{F}_{n-1}$ measurable since,

$$\{\sigma < n \leq \tau\} = \{\sigma < n\} \cap \{n \leq \tau\} = \{\sigma < n\} \cap \{\tau < n\}^c \in \mathcal{F}_{n-1}.$$

■

Proposition 13.29 (The Discrete Stochastic Integral). *Let $X = \{X_n\}_{n=0}^\infty$ be an adapted integrable process, i.e. $\mathbb{E}|X_n| < \infty$ for all n. If X is a martingale and $\{C_n\}_{n=1}^\infty$ is a predictable (as in Definition 13.9) 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). (In other words, X is a sub-martingale if no matter what your (non-negative) betting strategy is you will make money on average.)*

[3] Stocks only change price at integer times in this model.

Proof. For any adapted process X, we have

$$\begin{aligned}\mathbb{E}\left[(C\cdot\Delta X)_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[(C\cdot\Delta X)_n + C_{n+1}(X_{n+1}-X_n)|\mathcal{F}_n\right] \\ &= (C\cdot\Delta X)_n + C_{n+1}\mathbb{E}\left[(X_{n+1}-X_n)|\mathcal{F}_n\right] \qquad (13.26)\end{aligned}$$

from which the result easily follows.

Remark: For $n=1$ we have,

$$\mathbb{E}\left[(C\cdot\Delta X)_1|\mathcal{F}_0\right] = C_1\cdot\mathbb{E}\left[\Delta_1 X|\mathcal{F}_0\right] = \begin{cases} 0 & \text{if } X \text{ is a mtg.} \\ \geq 0 & \text{if } X \text{ is a submtg}\end{cases}$$

and this explains why taking $(C\cdot\Delta X)_0 = 0$ is the correct choice here given that

$$(C\cdot\Delta X)_n = \sum_{k=1}^n C_k\Delta_k X \text{ when } n\geq 1.$$

■

Example 13.30. Suppose that $\{X_n\}_{n=0}^\infty$ are mean zero independent integrable random variables and $f_k:\mathbb{R}^k\to\mathbb{R}$ are bounded measurable functions for $k\in\mathbb{N}$. Then $\{Y_n\}_{n=0}^\infty$, defined by $Y_0=0$ and

$$Y_n := \sum_{k=1}^n f_k(X_0,\dots,X_{k-1})(X_k-X_{k-1}) \text{ for } n\in\mathbb{N}, \qquad (13.27)$$

is a martingale sequence relative to $\left\{\mathcal{F}_n^X\right\}_{n\geq 0}$.

Notation 13.31 *Given an $\{\mathcal{F}_n\}-$ adapted process, X, and an $\{\mathcal{F}_n\}-$ stopping time τ, let $X_n^\tau := X_{\tau\wedge n}$. We call $X^\tau := \{X_n^\tau\}_{n=0}^\infty$ the **process** X **stopped by** τ.*

Theorem 13.32 (Optional stopping theorem). *Suppose $X=\{X_n\}_{n=0}^\infty$ is a supermartingale, martingale, or submartingale with $\mathbb{E}|X_n|<\infty$ for all n. Then for every stopping time, τ, X^τ is a $\{\mathcal{F}_n\}_{n=0}^\infty$ – supermartingale, martingale, or submartingale respectively.*

Proof. If we define $C_k := 1_{k\leq\tau}$, then

$$[C\cdot\Delta X]_n = \sum_{k=1}^n 1_{k\leq\tau}\Delta_k X = \sum_{k=1}^{n\wedge\tau}\Delta_k X = X_{n\wedge\tau}-X_0.$$

This shows that

$$X_n^\tau = X_0 + [C\cdot\Delta X]_n$$

and the result now follows Proposition 13.29. ■

Corollary 13.33. *Suppose $X=\{X_n\}_{n=0}^\infty$ is a martingale and τ is a stopping time such that $\mathbb{P}(\tau=\infty)=0$. If there exists $C<\infty$ such that either $\mathbb{E}[\sup_n|X_n|]<\infty$ for all n or $\tau\leq C$, then*

$$\mathbb{E}X_\tau = \mathbb{E}X_0. \qquad (13.28)$$

Proof. From Theorem 13.32 we know that

$$\mathbb{E}X_0 = \mathbb{E}\left[X_n^\tau\right] = \mathbb{E}\left[X_{\tau\wedge n}\right] \text{ for all } n\in\mathbb{N}.$$

- If we assume that $\tau\leq C$ then by taking $n>C$ we learn that Eq. (13.28) holds.
- If we assume $\mathbb{E}[\sup_n|X_n|]<\infty$, then by the dominated convergence theorem (see Section 1.1),

$$\mathbb{E}X_0 = \lim_{n\to\infty}\mathbb{E}\left[X_{\tau\wedge n}\right] = \mathbb{E}\left[\lim_{n\to\infty}X_{\tau\wedge n}\right] = \mathbb{E}\left[X_\tau\right].$$

■

Corollary 13.34 (Martingale Expectations). *Suppose that $\{X_n\}_{n=0}^\infty$ is a martingale and τ is a stopping time so that $\mathbb{P}(\tau=\infty)=0$, $\mathbb{E}|X_\tau|<\infty$, and*

$$\lim_{n\to\infty}\mathbb{E}\left[|X_n|:\tau>n\right]=0, \qquad (13.29)$$

then $\mathbb{E}X_\tau=\mathbb{E}X_0$.

Proof. We have

$$X_\tau = X_{\tau\wedge n} + (X_\tau - X_n)1_{\tau>n}$$

and hence

$$\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[X_{\tau\wedge n}\right] + \varepsilon_n = \mathbb{E}X_0 + \varepsilon_n$$

where

$$|\varepsilon_n| = |\mathbb{E}\left[(X_\tau - X_n)1_{\tau>n}\right]| \leq \mathbb{E}\left[|X_\tau|\cdot 1_{\tau>n}\right] + \mathbb{E}\left[|X_n|\cdot 1_{\tau>n}\right].$$

The first term of the right side above goes to 0 as $n\to\infty$ by DCT while the second terms goes to zero by the assumption in Eq. (13.29) and so $\varepsilon_n\to 0$ as $n\to\infty$ and hence $\mathbb{E}X_\tau=\mathbb{E}X_0$. ■

Example 13.35. One way to "beat" the coin flip game in Example 13.25 is to use the double or nothing betting strategy. Our strategy for making $M>0$ dollars is to bet $C_k = 2^{k-1}M$ on the kth - flip, i.e we start by betting $\$M$ and then continually double our bet at each successive flip. We will quit the first time that a heads is flipped. To analyze this strategy, let

$$T = \inf\{k \geq 1 : Z_k = 1\} = \inf\{k \geq 1 : X_k - X_{k-1} = 1\}.$$

Notice that T is an $\mathcal{F}^Z$ – stopping time since

$$\{T = n\} = \{Z_1 = -1, \ldots, Z_{n-1} = -1, Z_n = 1\} \in \mathcal{F}_n^Z \text{ for all } n$$

and

$$\mathbb{P}(T > n) = \mathbb{P}(Z_1 = -1, \ldots, Z_{n-1} = -1, Z_n = -1) = \left(\frac{1}{2}\right)^n$$

so that $\mathbb{P}(T = \infty) = \lim_{n\to\infty} \mathbb{P}(T > n) = 0$. Our wealth, W_n, at time n will be

$$W_n = W_0 + M \sum_{k=1}^{n} 2^{k-1} Z_k$$

where W_0 is our initial wealth. [For the moment we suppose we have good credit with the casino and so our wealth is allowed to go negative.] The process $\{W_n\}_{n=0}^{\infty}$ is a martingale and on the event $\{T > n\}$,

$$W_n - W_0 = -M \sum_{k=1}^{n} 2^{k-1} = -M(2^n - 1)$$

while on the event $\{T = n\}$,

$$W_n - W_0 = W_n - W_{n-1} + W_{n-1} - W_0 = M\left[2^{n-1} - \left(2^{n-1} - 1\right)\right] = M.$$

Thus we have shown that $W_T = W_0 + M$ and we have indeed one \$M. It clearly now follows that

$$\mathbb{E}W_T = \mathbb{E}W_0 + M \neq \mathbb{E}W_0$$

even though $\mathbb{E}W_n = \mathbb{E}W_0$ for all n? This example shows that the assumption in Eq. (13.29) of Corollary 13.34 can not be eliminated in general. Here we have

$$\mathbb{E}\left[|W_n - W_0| : T > n\right] = M(2^n - 1)\cdot\left(\frac{1}{2}\right)^n = M\left[1 - \left(\frac{1}{2}\right)^n\right] \to M \text{ as } n \to \infty$$

showing explicitly that Eq. (13.29) does not hold.

Example 13.36. Let us now suppose that, in Example 13.35 above that are initial wealth is $W_0 = M(2^m - 1)$ and that we have **no** credit with the casino so that we must stop betting at any time n where $W_n = 0$. Keeping the same betting strategy as in Example 13.35, we now let

$$\tau_m = \inf\{n \geq 1 : W_n = 0\}.$$

Then $\tau_m \wedge T$ is a finite stopping time (a.s.) and $0 \leq W_n^{T\wedge\tau_m} \leq W_0 + M$ for all $n \in \mathbb{N}_0$. Therefor Corollary 13.33 is applicable and shows

$$M(2^m - 1) = \mathbb{E}W_0 = \mathbb{E}W_{\tau_m\wedge T} = 0\cdot\mathbb{P}(\tau_m < T) + [M(2^m - 1) + M]\cdot\mathbb{P}(T < \tau_m)$$

so that

$$\mathbb{P}(\text{you win}) = \mathbb{P}(T < \tau_m) = \frac{2^m - 1}{2^m} = 1 - \frac{1}{2^m}.$$

In hind site, this result is obvious since you loose precisely when $Z_k = -1$ for $k = 1, 2, \ldots, m$, i.e.

$$\mathbb{P}(\text{you loose}) = \mathbb{P}(\tau_m < T) = \left(\frac{1}{2}\right)^m = \frac{1}{2^m}.$$

Moral: the more wealth you start out with the more likely you will win $\$M$ before going bust. Nevertheless, if you repeatedly use this strategy over and over again with a fixed m, then you expected winnings are going to be zero!! [Over many trials you will win most of the time and gain $\$M$ but you will also loose on rare occasions at a heavy cost of loosing \$ $M(2^m - 1)$.]

Example 13.37. 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 $C_n X_n - C_n X_{n-1} = C_n \Delta_n X$ and your wealth at time n will be

$$W_n = W_0 + \sum_{k=1}^{n} C_k \Delta_k X.$$

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.

The next lemma gives the optimal strategy (relative to expected return) for buying stock which is known to be a submartingale.

Lemma 13.38 (*"All In"). *If $\{X_n\}_{n=0}^{\infty}$ is a sub-martingale and $\{C_k\}_{k=1}^{\infty}$ is a previsible process with values in $[0, 1]$, then*

$$\mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right) \leq \mathbb{E}\left[X_n - X_0\right]$$

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 13.29,

$$\mathbb{E}\left(\sum_{k=1}^{n} (1 - C_k)\Delta_k X\right) \geq 0.$$

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

$$\mathbb{E}[X_n - X_0] = \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right) + \mathbb{E}\left(\sum_{k=1}^{n} (1 - C_k) \Delta_k X\right) \geq \mathbb{E}\left(\sum_{k=1}^{n} C_k \Delta_k X\right).$$

■

*Example 13.39 (*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[4]. Let $\{Z_n\}_{n=1}^{\infty}$ be random functions with values in S such that $p(s) := \mathbb{P}(Z_n = s)$ for all $s \in S$. (Z_n represents the outcome of the nth – game.) Also let $\alpha : S \to [0, \infty)$ be the house's payoff function, i.e. for each dollar you (the gambler) bets on $s \in S$, the house will pay $\alpha(s)$ dollars back if s is rolled. Further let $W : \Omega \to \mathbb{W}$ be measurable function into some other measure space, $(\mathbb{W}, \mathcal{F})$ which is to represent your random (or not so random) "whims.". We now assume that Z_n is independent of $(W, Z_1, \ldots, Z_{n-1})$ for each n, i.e. the dice are not influenced by the previous plays or your whims. If we let $\mathcal{F}_n := \sigma(W, Z_1, \ldots, Z_n)$ with $\mathcal{F}_0 = \sigma(W)$, then we are assuming the Z_n is independent of $\mathcal{F}_{n-1}$ for each $n \in \mathbb{N}$.

As a gambler, you are allowed to choose **before** the nth – game is played, the amounts $\left(\{C_n(s)\}_{s \in S}\right)$ that you want to bet on each of the possible outcomes of the nth – game. Assuming that you are not clairvoyant (i.e. can not see the future), these amounts may be random but must be $\mathcal{F}_{n-1}$ – 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 13.9 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 nth – 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 nth – game and $C_n(Z_n)\, \alpha(Z_n)$ represents the house's payoff to you for the nth – game. Therefore it follows that

$$X_n = X_0 + \sum_{k=1}^{n} \left[-\sum_{s \in S} C_k(s) + C_n(Z_k)\, \alpha(Z_k) \right],$$

[4] To be concrete, take $S = \{2, \ldots, 12\}$ representing the possible values for the sums of the upward pointing faces of two dice. Assuming the dice are independent and fair then determines $p : S \to (0,1)$. For example $p(2) = p(12) = 1/36$, $p(3) = p(11) = 1/18$, $p(7) = 1/6$, etc.

X_n is $\mathcal{F}_n$ – measurable for each n, and

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_{n-1}}[X_n - X_{n-1}] &= -\sum_{s \in S} C_n(s) + \mathbb{E}_{\mathcal{F}_{n-1}}[C_n(Z_n)\, \alpha(Z_n)] \\
&= -\sum_{s \in S} C_n(s) + \sum_{s \in S} C_n(s)\, \alpha(s)\, p(s) \\
&= \sum_{s \in S} C_n(s) \left(\alpha(s)\, p(s) - 1\right).
\end{aligned}$$

Thus it follows, that no matter the choice of the betting "strategy," $\{C_n(s) : s \in S\}_{n=1}^{\infty}$, we will have

$$\mathbb{E}_{\mathcal{F}_{n-1}}[X_n - X_{n-1}] = \begin{cases} \geq 0 \text{ if } \alpha(\cdot)\, p(\cdot) \geq 1 \\ = 0 \text{ if } \alpha(\cdot)\, p(\cdot) = 1 \\ \leq 0 \text{ if } \alpha(\cdot)\, p(\cdot) \leq 1 \end{cases},$$

that is $\{X_n\}_{n \geq 0}$ is a sub-martingale, martingale, or supermartingale depending on whether $\alpha \cdot p \geq 1$, $\alpha \cdot p = 1$, or $\alpha \cdot p \leq 1$.

Moral: If the Casino wants to be guaranteed to make money on average, it had better choose $\alpha : S \to [0, \infty)$ such that $\alpha(s) < 1/p(s)$ for all $s \in S$. In this case the expected earnings of the gambler will be decreasing which means the expected earnings of the Casino will be increasing.

13.4 Martingale Convergence Theorems

Definition 13.40. *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a sequence of extended real numbers and $-\infty < a < b < \infty$. To each $N \in \mathbb{N}$, let $U_N(a,b)$ denote the number of* ***up-crossings of*** *$[a,b]$, i.e. the number times that $\{X_n\}_{n=0}^{N}$ goes from below a to above b.*

Notice that if $\{C_n\}_{n=1}^{\infty}$ is the betting strategy of buy at or below a and sell at or above b then $U_N(a,b) = [C \cdot \Delta X]_N$. In more detail, $C_n = 1$ iff $X_j \leq a$ for some $0 \leq j < n$ and following $\{X_k\}_{k=0}^{n-1}$ backwards we have $X_{n-k} < b$ for $k \geq 1$ until the first time that $X_k \leq a$. In any other scenario, we have $C_n = 0$.

Lemma 13.41. *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a sequence of extended real numbers such that $U_{\infty}^{X}(a,b) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$. Then $X_{\infty} := \lim_{n \to \infty} X_n$ exists in $\bar{\mathbb{R}}$.*

Proof. If $\lim_{n \to \infty} X_n$ does not exists in $\bar{\mathbb{R}}$, then there would exists $a, b \in \mathbb{Q}$ such that

$$\liminf_{n \to \infty} X_n < a < b < \limsup_{n \to \infty} X_n$$

and for this choice of a and b, we must have $X_n < a$ and $X_n > b$ infinitely often. Therefore, $U_{\infty}^{X}(a,b) = \infty$. ■

Theorem 13.42 (Doob's Upcrossing Inequality). *If $\{X_n\}_{n=0}^\infty$ is a martingale and $-\infty < a < b < \infty$, then for all $N \in \mathbb{N}$,*

$$\mathbb{E}\left[U_N^X(a,b)\right] \le \frac{1}{b-a}\left[\mathbb{E}\left(X_N - a\right)_+\right] \le \frac{1}{b-a}\left[\mathbb{E}\left|X_N - a\right|\right]. \tag{13.30}$$

Proof. Let $\{C_k\}_{k=1}^\infty$ be the buy at or below a and sell at or above b as explained after Definition 13.40. Our net winnings of this strategy up to time N is $[C \cdot \Delta X]_N$ which satisfies,

$$[C \cdot \Delta X]_N \ge (b-a)\, U_N^X(a,b) - [a - X_N]_+ . \tag{13.31}$$

In words, 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 of times we buy low and sell high plus a possible penalty for holding the stock below a at the end of the day. [The worst case scenario for this penalty is that we bought the stock for \$ a and at time N, $X_N < a$, so that we are down \$ $(a - X_N)$.] From Proposition 13.29 we know that $\mathbb{E}[C \cdot \Delta X]_N = 0$ for all N and therefore taking expectation of Eq. (13.31) gives Eq. (13.30). ■

Theorem 13.43 (Martingale Convergence Theorem). *If $\{X_n\}_{n=0}^\infty$ is a martingale such that $C := \sup_n \mathbb{E}|X_n| < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $\mathbb{E}|X_\infty| \le C$.*

Proof. Sketch. If $-\infty < a < b < \infty$, it follows from Eq. (13.30) and Fatou's Lemma or Monotone convergence theorem (see Section 1.1) that

$$\mathbb{E}\left[U_\infty^X(a,b)\right] = \mathbb{E}\left[\lim_{N\to\infty} U_N^X(a,b)\right] \le \liminf_{N\to\infty} \mathbb{E}\left[U_N^X(a,b)\right] \le \frac{C+a}{b-a} < \infty.$$

From this we learn that $\mathbb{P}\left(U_\infty^X(a,b) < \infty\right) = 0$ for all $a, b \in \mathbb{Q}$ with $-\infty < a < b < \infty$ and hence

$$\mathbb{P}\left(U_\infty^X(a,b) < \infty \ \forall -\infty < a < b < \infty\right) = 0.$$

From this inequality and Lemma 13.41 it follows that $\lim_{n\to\infty} X_n = X_\infty$ exists a.s. as a random variable with values in $\mathbb{R} \cup \{\pm\infty\}$. Another application of Fatou's lemma then shows

$$\mathbb{E}|X_\infty| = \mathbb{E}\lim_{n\to\infty}|X_n| \le \liminf_{N\to\infty}\mathbb{E}|X_n| \le C < \infty.$$

■

Corollary 13.44 (Submartingale convergence theorem). *If $\{X_n\}_{n=0}^\infty$ is a submartingale such that $C := \sup_n \mathbb{E}|X_n| < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $\mathbb{E}|X_\infty| \le C$.*

Proof. By Doob's decomposition (Lemma 13.10), we have $X_n = M_n + A_n$ where $\{M_n\}_{n=0}^\infty$ is a martingale and

$$A_n = \sum_{k=1}^n \mathbb{E}\left[\Delta_k X | \mathcal{F}_{k-1}\right].$$

As $\{X_n\}_{n=0}^\infty$ is a submartingale we know $\mathbb{E}\left[\Delta_k X|\mathcal{F}_{k-1}\right]$ and hence $A_n \ge 0$ for all n and A_n is increasing (i.e. non-decreasing) as n increases. Furthermore by the monotone convergence theorem,

$$\begin{aligned}\mathbb{E}A_\infty &= \lim_{n\to\infty}\mathbb{E}A_n = \lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\mathbb{E}\left[\Delta_k X|\mathcal{F}_{k-1}\right] \\ &= \lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\left[\Delta_k X\right] = \lim_{n\to\infty}\mathbb{E}\left[X_n - X_0\right] \le 2C.\end{aligned}$$

Thus we may conclude for all $n \in \mathbb{N}_0$ that

$$\mathbb{E}|M_n| = \mathbb{E}|X_n - A_n| \le \mathbb{E}|X_n| + \mathbb{E}A_n \le \mathbb{E}|X_n| + \mathbb{E}A_\infty \le 3C < \infty.$$

Therefore $\lim_{n\to\infty} M_n = M_\infty$ exists by Theorem 13.43 and hence

$$\lim_{n\to\infty} X_n = \lim_{n\to\infty}\left[M_n + A_n\right] = M_\infty + A_\infty$$

exists as well. ■

13.5 Uniform Integrability and Optional stopping

Definition 13.45. *A sequence of random functions, $\{X_n\}_{n=0}^\infty$, is uniformly integrable (U.I. for short) if the tails are uniformly small in the following sense,*

$$\varepsilon_K := \sup_n \mathbb{E}\left[|X_n| : |X_n| \ge K\right] \to 0 \text{ as } K \to \infty.$$

Lemma 13.46. *If $\{X_n\}$ is U.I. then $\sup_n \mathbb{E}\left[|X_n|\right] < \infty$.*

Proof. Fix a $K < \infty$ such that

$$\varepsilon_K := \sup_n \mathbb{E}\left[|X_n| : |X_n| \ge K\right] \le 1.$$

For this K it follows that

$$\mathbb{E}\left[|X_n|\right] = \mathbb{E}\left[|X_n| : |X_n| \ge K\right] + \mathbb{E}\left[|X_n| : |X_n| < K\right] \le 1 + K$$

and so

$$M := \sup_n \mathbb{E}\left[|X_n|\right] \le 1 + K < \infty.$$

■

Example 13.47. Let $X_n = M_n = 2^n 1_{[0,2^{-n}]}$ for $\omega \in \Omega := [0,1]$ as in Exercise 13.2 where $\mathbb{P}$ is the length measure on Ω. Then $\mathbb{E}|X_n| = 1$ for all n but $\{X_n\}$ is **not** uniformly integrable. Indeed for any $K > 0$ we will have $\mathbb{E}[|X_n| : |X_n| \geq K] = 1$ provided $2^n > K$, i.e. $n > \ln K / \ln 2$.

On the other hand if $1 < \alpha < 2$ and $Y_n := \alpha^n 1_{[0,2^{-n}]}$, then

$$\sup_n \mathbb{E}[|Y_n| : |Y_n| \geq K] = \sup_n \left(\frac{\alpha}{2}\right)^n 1_{n > \frac{\ln K}{\ln \alpha}} \leq \left(\frac{\alpha}{2}\right)^{\frac{\ln K}{\ln \alpha}} \to 0 \text{ as } K \to \infty$$

and so $\{Y_n\}_{n=1}^\infty$ is uniformly integrable. Intuitively, an $L^1(\mathbb{P})$ – bounded sequence, $\{X_n\}_{n=0}^\infty$, is uniformly integrable provided the X_n 's are not allowed to form "to sharp" of peaks.

Proposition 13.48. *If $\{X_n\}$ is U.I., then for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_n \mathbb{E}[|X_n| : A] < \varepsilon$ whenever $A \subset \Omega$ with $\mathbb{P}(A) < \delta$.*

Proof. For any $A \subset \Omega$,

$$\begin{aligned}\mathbb{E}[|X_n| : A] &= \mathbb{E}[|X_n| : A, |X_n| \geq K] + \mathbb{E}[|X_n| : A, |X_n| < K] \\ &\leq \mathbb{E}[|X_n| : |X_n| \geq K] + K\mathbb{P}(A)\end{aligned}$$

and so

$$\sup_n \mathbb{E}[|X_n| : A] \leq \varepsilon_K + K\mathbb{P}(A). \tag{13.32}$$

Thus given $\varepsilon > 0$ choose K so large that $\varepsilon_K < \varepsilon/2$ and choose $\delta := \frac{\varepsilon}{2K}$ so that when $\mathbb{P}(A) < \delta$ we will have $K\mathbb{P}(A) < \varepsilon/2$ which then combined with Eq. (13.32) implies, $\sup_n \mathbb{E}[|X_n| : A] < \varepsilon$. ■

Corollary 13.49. *Suppose that $\{X_n\}_{n=0}^\infty$ is a U.I. martingale and τ is a stopping time so that $\mathbb{P}(\tau = \infty) = 0$ and $\mathbb{E}|X_\tau| < \infty$, then $\mathbb{E}X_\tau = \mathbb{E}X_0$.*

Proof. According to Corollary 13.34 we need to use the U.I. assumption to show

$$\lim_{n\to\infty} \mathbb{E}[|X_n| : \tau > n] = 0.$$

But using $\mathbb{P}(\tau > n) \to 0$ as $n \to \infty$ we may use Proposition 13.48 to conclude,

$$\mathbb{E}[|X_n| : \tau > n] \leq \sup_k \mathbb{E}[|X_k| : \tau > n] \to 0 \text{ as } n \to \infty.$$

■

Let us end this section with a couple of criteria for checking when $\{X_n\}$ is U.I.

Proposition 13.50 (DCT and U.I.). *If $F = \sup_n |X_n| \in L^1(\mathbb{P})$ then $\{X_n\}$ is U.I.*

Proof. We have

$$\sup_n \mathbb{E}[|X_n| : |X_n| \geq K] \leq \mathbb{E}[F : F \geq K] \to 0 \text{ as } K \to \infty$$

by DCT. ■

Proposition 13.51 (p – integrability and U.I.). *If there exists $p > 1$ such that $M := \sup_n \mathbb{E}|X_n|^p < \infty$ then $\{X_n\}$ is U.I.*

Proof. If we let $\varepsilon := p - 1 > 0$, then

$$\begin{aligned}\mathbb{E}[|X_n| : |X_n| \geq K] &= \mathbb{E}\left[|X_n|^{1+\varepsilon} \frac{1}{|X_n|^\varepsilon} : |X_n| \geq K\right] \\ &\leq \frac{1}{K^\varepsilon}\mathbb{E}\left[|X_n|^{1+\varepsilon} : |X_n| \geq K\right] \leq \frac{1}{K^\varepsilon}\mathbb{E}[|X_n|^p]\end{aligned}$$

and so

$$\sup_n \mathbb{E}[|X_n| : |X_n| \geq K] \leq \frac{M}{K^\varepsilon} \to 0 \text{ as } K \uparrow \infty.$$

■

Proposition 13.52. *If $\{X_n\}_{n=0}^\infty$, is uniformly integrable and $X = \lim_{n\to\infty} X_n$ exists a.s., then $\lim_{n\to\infty} \mathbb{E}|X - X_n| = 0$.*

Proof. By Fatou's lemma, $\mathbb{E}|X| \leq M := \sup_n \mathbb{E}[|X_n|] < \infty$ so that X is integrable. Let $Y_n := |X - X_n|$ so that $Y_n \to 0$ a.s. and $\mathbb{E}Y_n \leq 2M$ for all n. Using DCT twice we find, for any $K > 0$ that

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}Y_n &\leq \limsup_{n\to\infty} \mathbb{E}[Y_n : Y_n < K] + \limsup_{n\to\infty} \mathbb{E}[Y_n : Y_n \geq K] \\ &= \limsup_{n\to\infty} \mathbb{E}[Y_n : Y_n \geq K]\end{aligned}$$

where

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}[Y_n : Y_n \geq K] &\leq \limsup_{n\to\infty} \mathbb{E}[|X| : Y_n \geq K] + \limsup_{n\to\infty} \mathbb{E}[|X_n| : Y_n \geq K] \\ &= \limsup_{n\to\infty} \mathbb{E}[|X_n| : Y_n \geq K] \leq \sup_n \mathbb{E}[|X_n| : Y_n \geq K].\end{aligned}$$

Since $\mathbb{P}(Y_n \geq K) \leq 2M/K$, it follows by Proposition 13.51 that

$$\limsup_{n\to\infty} \mathbb{E}Y_n \leq \sup_n \mathbb{E}[|X_n| : Y_n \geq K] \to 0 \text{ as } K \to \infty.$$

■

Corollary 13.53. *If $X = \lim_{n\to\infty} X_n$ exists a.s. and there exists $p > 1$ such that $M := \sup_n \mathbb{E}|X_n|^p < \infty$, then $\lim_{n\to\infty} \mathbb{E}|X - X_n| = 0$.*

Proof. First proof. This is a direct consequence of Proposition 13.51 and 13.52.

Second direct proof. By Fatou's lemma,

$$\mathbb{E}\,|X|^p = \mathbb{E}\liminf_{n\to\infty}|X_n|^p \le \liminf_{n\to\infty}\mathbb{E}\,|X_n|^p \le M < \infty.$$

Let $Y_n := |X - X_n|$ so that $Y_n \to 0$ a.s. and

$$\mathbb{E}Y_n^p \le C_p\left[\mathbb{E}\,|X|^p + \mathbb{E}\,|X_n|^p\right] \le M' = 2C_pM.$$

Now for any $K > 0$ we have,

$$\begin{aligned}\mathbb{E}Y_n &= \mathbb{E}\left[Y_n : Y_n \le K\right] + \mathbb{E}\left[Y_n : Y_n > K\right]\\ &\le \mathbb{E}\left[Y_n 1_{Y_n\le K}\right] + \mathbb{E}\left[Y_n\left(\frac{Y_n}{K}\right)^{p-1} : Y_n > K\right]\\ &\le \mathbb{E}\left[Y_n 1_{Y_n\le K}\right] + \frac{1}{K^{p-1}}\mathbb{E}Y_n^p \le \mathbb{E}\left[Y_n 1_{Y_n\le K}\right] + \frac{M'}{K^{p-1}}.\end{aligned}$$

Thus by an application of DCT,

$$\limsup_{n\to\infty}\mathbb{E}Y_n \le \limsup_{n\to\infty}\mathbb{E}\left[Y_n 1_{Y_n\le K}\right] + \frac{M'}{K^{p-1}} = \frac{M'}{K^{p-1}} \to 0 \text{ as } K\to\infty.$$

■

Here is the theorem summarizing the key facts about uniformly integrable martingales.

Theorem 13.54 (Uniformly integrable martingales). *If $\{X_n\}_{n=0}^\infty$ is a U.I. martingale then*

1. *$X_\infty = \lim_{n\to\infty} X_n$ exists ($\mathbb{P}$ – a.s.) and $\lim_{n\to\infty}\mathbb{E}\,|X_n - X_\infty| = 0$, i.e. $X_n \to X_\infty$ in $L^1(\mathbb{P})$. [This in particular implies that $\lim_{n\to\infty}\mathbb{E}X_n = \mathbb{E}X_\infty$.]*
2. *$\{X_n\}_{n=0}^\infty$ is a regular martingale and in fact $\{X_n = \mathbb{E}\left[X_\infty|\mathcal{F}_n\right]\}_{n=0}^\infty$.*
3. *For any stopping time τ (no need to assume $\mathbb{P}(\tau=\infty) = 0$) we have $\mathbb{E}\,|X_\tau| < \infty$ and $\mathbb{E}X_\tau = \mathbb{E}X_0$. [This even holds if $\tau = \infty$ a.s.]. [This result extends Corollary 13.49.]*

Proof. We take each item in turn.

1. The first item is a direct consequence of Lemma 13.46, Theorem 13.43, and Proposition 13.52.
2. Passing to the limit, $N \to \infty$, in the equality, $X_n = \mathbb{E}\left[X_N|\mathcal{F}_n\right]$, using the fact that conditional expectation is a $L^1(\mathbb{P})$ – contraction proves item 2.
3. From Theorem 4.17,

$$\mathbb{E}\left[X_\infty|\mathcal{F}_\tau\right] = \sum_{n\in\bar{\mathbb{N}}_0}\mathbb{E}\left[X_\infty|\mathcal{F}_n\right]1_{\tau=n} = \sum_{n\in\bar{\mathbb{N}}_0}X_n 1_{\tau=n} = X_\tau.$$

Combining this equality with the contractive and tower properties of conditional expectations shows $\mathbb{E}\,|X_\tau| \le \mathbb{E}\,|X_\infty| < \infty$ and

$$\mathbb{E}X_\tau = \mathbb{E}\left(\mathbb{E}\left[X_\infty|\mathcal{F}_\tau\right]\right) = \mathbb{E}X_\infty = \mathbb{E}\mathbb{E}\left[X_\infty|\mathcal{F}_n\right] = \mathbb{E}X_n \;\forall\; n\in\mathbb{N}_0.$$

■

Example 13.55 (Generalized Random Sums). Suppose that $\{X_k\}_{k=1}^\infty$ are independent random variables such that $\mathbb{E}X_k = 0$. If we further know that $K := \sum_{k=1}^\infty \mathbb{E}X_k^2 < \infty$, then $\sum_{k=1}^\infty X_k$ exists a.s., i.e.

$$\mathbb{P}\left(\sum_{k=1}^\infty X_k \text{ converges in } \mathbb{R}\right) = 1.$$

To prove this, note that $S_0 = 0$ and $S_n = \sum_{k=1}^n X_k$ is a martingale. Moreover, using the fact that variances of sums of independent random variables add we find,

$$\mathbb{E}S_n^2 = \mathrm{Var}\left(S_n\right) = \sum_{k=1}^n \mathrm{Var}\left(X_k\right) = \sum_{k=1}^n \mathbb{E}X_k^2 \le K < \infty.$$

From Proposition 13.51 we conclude that $\{S_n\}_{n=0}^\infty$ is a uniformly integrable martingale and in particular $\{S_n\}_{n=0}^\infty$ is L^1 - bounded.[5] Therefore the result now follows from the martingale convergence theorem, Theorem 13.43. Moreover, making use of Theorem 13.54 we may conclude that

$$\mathbb{E}\left[S_\infty\right] = \mathbb{E}\left[\sum_{k=1}^\infty X_k\right] = \mathbb{E}S_0 = 0$$

and more generally if τ is any stopping time we have $\mathbb{E}\left[S_\tau\right] = \mathbb{E}\left[S_0\right] = 0$.

For a more explicit example, if $X_k = \frac{1}{k}Z_k$ where $\{Z_k\}_{k=1}^\infty$ are i.i.d. such that $\mathbb{P}\left(Z_k = \pm 1\right) = \frac{1}{2}$, then $\sum_{k=1}^\infty \mathbb{E}X_k^2 = \sum_{k=1}^\infty \frac{1}{k^2} < \infty$ and therefore the random harmonic series, $\sum_{k=1}^\infty \frac{1}{k}Z_k$ is almost surely convergent. However, notice that

$$\sum_{k=1}^\infty\left|\frac{1}{k}Z_k\right| = \sum_{k=1}^\infty\frac{1}{k} = \infty,$$

i.e. the series $\sum_{k=1}^\infty \frac{1}{k}Z_k$ is never absolutely convergent. For the full generalization of this result look up Kolmogorov's three series theorem.

[5] Alternatively, by Jensen's inequality, $\left[\mathbb{E}\,|S_n|\right]^2 \le \mathbb{E}S_n^2 \le K$ which again shows $\{S_n\}$ is L^1 – bounded.

13.6 Submartingale Maximal Inequalities

Notation 13.56 (Running Maximum) *If $X = \{X_n\}_{n=0}^\infty$ is a sequence of (extended) real numbers, we let*

$$X_N^* := \max\{X_0, \dots, X_N\}. \tag{13.33}$$

Proposition 13.57 (Submartingale Maximal Inequalities). *Let $\{X_n\}$ be a submartingale on a filtered probability space, $(\Omega, \{\mathcal{F}_n\}_{n=0}^\infty, \mathbb{P})$. Then for any $a \ge 0$ and $N \in \mathbb{N}_0$,*

$$a\mathbb{P}(X_N^* \ge a) \le \mathbb{E}[X_N : X_N^* \ge a] \le \mathbb{E}\left[X_N^+\right], \tag{13.34}$$

where $X_N^+ := X_N \vee 0 = \max(X_N, 0)$.

Proof. Let $\tau := \inf\{n : X_n \ge a\}$ and observe $X_N^* \ge a$ iff $\tau \le N$ and $a \le X_k$ on $\{\tau = k\}$. Therefore,

$$\begin{aligned} a\mathbb{P}(X_N^* \ge a) =& a\mathbb{P}(\tau \le N) = \sum_{k=0}^N \mathbb{E}[a1_{\tau=k}] \\ \le& \sum_{k=0}^N \mathbb{E}[X_k 1_{\tau=k}] \le \sum_{k=0}^N \mathbb{E}[X_N 1_{\tau=k}] \\ &= \mathbb{E}[X_N : \tau \le N] = \mathbb{E}[X_N : X_N^* \ge a] \\ &\le \mathbb{E}\left[X_N^+ : X_N^* \ge a\right] \le \mathbb{E}X_N^+, \end{aligned}$$

wherein we have used $\{\tau = k\} \in \mathcal{F}_k$ and $\{X_n\}_{n=0}^\infty$ is a submartingale in the second inequality.

■

Corollary 13.58. *If $\{M_n\}_{n=0}^\infty$ is a martingale and $M_N^* := \max\{M_0, \dots, M_N\}$, then for $a > 0$,*

$$\mathbb{P}\left(|M|_N^* \ge a\right) \le \frac{1}{a^p}\mathbb{E}|M_N|^p \text{ for all } p \ge 1 \tag{13.35}$$

and

$$\mathbb{P}(M_N^* \ge a) \le \frac{1}{e^{\lambda a}}\mathbb{E}\left[e^{\lambda M_N}\right] \text{ for all } \lambda \ge 0. \tag{13.36}$$

[This corollary has fairly obvious generalizations to other convex functions, φ, other than $\varphi(x) = |x|^p$ and $\varphi(x) = e^{\lambda x}$.]

Proof. By the conditional Jensen's inequality, it follows that $X_n := |M_n|^p$ is a submartingale and so Eq. (13.35) follows from Eq. (13.34) with a replaced by a^p. Again by the conditional Jensen's inequality, $\left\{X_n := e^{\lambda M_n}\right\}_{n=0}^\infty$ is a submartingale. Since $\lambda > 0$, we know that $x \to e^{\lambda x}$ is an increasing function and hence,

$$\{M_N^* \ge a\} = \left\{X_N^* \ge e^{\lambda a}\right\}$$

and so Eq. (13.36) follows from Eq. (13.34) with a replaced by $e^{\lambda a}$. ■

The following result is a substantial improvement of Proposition 2.18.

Corollary 13.59 (L^2 – SSLN). *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, and $\sigma^2 = \mathbb{E}X_m^2 < \infty$. Letting $S_m = \sum_{k=1}^m X_k$ and $p > 1/2$, we have*[6]

$$\mathbb{P}(S_m = O(m^p)) = 1 \tag{13.37}$$

As this is true for all $p > 1/2$ we may restate Eq. (13.37) as; if $p > 1/2$, then $\lim_{m\to\infty} \frac{S_m}{m^p} = 0$ a.s..

Proof. From Eq. (13.35) of Corollary 13.58 with $M_n = S_n$; if $N \in \mathbb{N}$ then,

$$\mathbb{P}\left(|S|_N^* \ge N^p\right) \le \frac{1}{N^{2p}}\mathbb{E}\left[S_N^2\right] = \frac{1}{N^{2p}}\sigma^2 N = \frac{\sigma^2}{N^{(2p-1)}}. \tag{13.38}$$

Choose $\alpha \in \mathbb{N}$ large enough so that $\alpha(2p-1) > 1$ and take $N = n^\alpha$ in Eq. (13.38) and then sum on n to find,

$$\mathbb{E}\left[\sum_{n=1}^\infty 1_{|S|_{n^\alpha}^* \ge n^{\alpha p}}\right] = \sum_{n=1}^\infty \mathbb{P}\left(|S|_{n^\alpha}^* \ge n^{\alpha p}\right) \le \sum_{n=1}^\infty \frac{\sigma^2}{n^{\alpha(2p-1)}} < \infty.$$

Therefore (as in the proof of the first Borel - Cantelli lemma),

$$\sum_{n=1}^\infty 1_{|S|_{n^\alpha}^* \ge n^{\alpha p}} < \infty \text{ a.s.} \implies \mathbb{P}\left(|S|_{n^\alpha}^* < n^{\alpha p} \text{ for a.a. } n\right) = 1$$

and hence, almost surely, we have

$$\frac{|S|_{n^\alpha}^*}{n^{\alpha p}} < 1 \text{ for almost all } n. \tag{13.39}$$

As $[1, \infty)$ is the disjoint union of the intervals, $\{[n^\alpha, (n+1)^\alpha)\}_{n=1}^\infty$, to each $m \in \mathbb{N}$ there exists a unique $n = n(m)$ such that $n^\alpha \le m < n^\alpha$. This inequality along with the (obvious) inequality, $|S|_m^* \le |S|_{(n+1)^\alpha}^*$,[7] then implies for sufficiently large m (with $n = n(m)$),

$$\frac{|S|_m^*}{m^p} \le \frac{|S|_{(n+1)^\alpha}^*}{n^{\alpha p}} = \frac{|S|_{(n+1)^\alpha}^*}{(n+1)^\alpha}\left(\frac{n+1}{n}\right)^{\alpha p} < 1 \cdot 2^{\alpha p}$$

which suffices to prove Eq. (13.37).

[6] Here we say $S_m = O(m^p)$ iff there exists $C < \infty$ so that $|S_m| \le Cm^p$ for all m, where C is allowed to be random here.

[7] This is where we need the maximal inequality since it would not necessarily be true that $|S|_m \le |S|_{(n+1)^\alpha}$!

■

Corollary 13.60. *If $\{Y_n\}$ is a sequence of independent random variables $\mathbb{E}Y_n = \mu$ and $\sigma^2 = \mathrm{Var}(X_n) < \infty$, then for any $\beta \in (0, 1/2)$,*

$$\frac{1}{m}\sum_{n=1}^{m} Y_n - \mu = O\left(\frac{1}{m^\beta}\right) \quad a.s.$$

Proof. Let $X_n = Y_n - \mu$. Then for $p > 1/2$ we have

$$\sum_{n=1}^{m} Y_n - m\mu = \sum_{n=1}^{m}(Y_n - \mu) = S_m = O(m^p) \text{ a.s.}$$

and therefore,

$$\frac{1}{m}\sum_{n=1}^{m} Y_n - \mu = O\left(\frac{1}{m^\alpha}\right)$$

where $\beta = 1 - p$ which can be any number in $(0, 1/2)$. ■

Remark 13.61 (Law of the iterated logarithm). Under the assumptions of Corollary 13.59, $\mathbb{E}S_m^2 = \sigma^2 m$ and so in the root mean square sense, $|S_m| \lesssim \sigma\sqrt{m}$. The result in Corollary 13.59 makes this statement honest at the expense of replacing $\sqrt{m}$ by m^p for any $p > \frac{1}{2}$. The more precise statement is the so called **law of the iterated logarithm** which states;

$$\limsup_{m\to\infty} \frac{S_m}{\sqrt{m \cdot \ln(\ln m)}} = 2\sigma \text{ a.s.}$$

Here is another (not so great) example of using Corollary 13.58.

Corollary 13.62. *Let $\{X_n\}$ be a sequence of Bernoulli random variables with $\mathbb{P}(X_n = \pm 1) = \frac{1}{2}$ and let $S_0 = 0$, $S_n := X_1 + \cdots + X_n$ for $n \in \mathbb{N}$. Then*

$$\lim_{N\to\infty} \frac{|S|_N^*}{\sqrt{N}\ln N} := \lim_{N\to\infty} \frac{\max\{|S_0|, \ldots, |S_N|\}}{\sqrt{N}\ln N} = 0 \ a.s. \tag{13.40}$$

[This result actually holds for an sequence $\{X_n\}$ of i.i.d. random variable such that $\mathbb{E}X_n = 0$ and $\mathrm{Var}(X_n) = \mathbb{E}X_n^2 < \infty$.]

Proof. By Eq. (13.36), if $a, \lambda > 0$ then

$$\begin{aligned}\mathbb{P}(S_N^* \geq a) &\leq \frac{1}{e^{\lambda a}}\mathbb{E}\left[e^{\lambda S_N}\right] = \frac{1}{e^{\lambda a}}\mathbb{E}\left[\prod_{j=1}^{N} e^{\lambda X_j}\right] \\ &= e^{-\lambda a}\prod_{j=1}^{N}\mathbb{E}\left[e^{\lambda X_j}\right] = e^{-\lambda a}\left[\cos(\lambda)\right]^N \end{aligned} \tag{13.41}$$

wherein we have used,

$$\mathbb{E}\left[e^{\lambda X_j}\right] = \frac{1}{2}\left[e^{\lambda} + e^{-\lambda}\right] = \cosh(\lambda).$$

Making the substitutions, $a \to a\sqrt{N}\cdot\ln N$ and $\lambda \to \lambda/\sqrt{N}$ in Eq. (13.41) shows

$$\begin{aligned}&\mathbb{P}\left(S_N^* \geq a\sqrt{N}\cdot\ln N\right) \\ &\quad \leq e^{-\lambda a \ln N}\left[\cos\left(\frac{\lambda}{\sqrt{N}}\right)\right]^N = \frac{1}{N^{\lambda a}}\left[\cos\left(\frac{\lambda}{\sqrt{N}}\right)\right]^N.\end{aligned}$$

Given $a > 0$ we now choose λ so that $\lambda a > 1$ in which case it follows that

$$\sum_{N=1}^{\infty}\mathbb{P}\left(S_N^* \geq a\sqrt{N}\cdot\ln N\right) \leq \sum_{N=1}^{\infty}\frac{1}{N^{\lambda a}}\left[\cos\left(\frac{\lambda}{\sqrt{N}}\right)\right]^N$$

wherein we have used

$$\lim_{N\to\infty}\left[\cos\left(\frac{\lambda}{\sqrt{N}}\right)\right]^N = \lim_{N\to\infty}\left(1 + \frac{\lambda}{2N} + O\left(\frac{1}{N^2}\right)\right)^N = e^{\lambda/2}.$$

From this inequality we learn

$$\mathbb{P}\left(S_N^* \geq a\sqrt{N}\ln N \text{ i.o. } N\right) = 0$$

and so with probability one we conclude, $\frac{S_N^*}{\sqrt{N}\cdot\ln N} < a$ for almost all N. Since a was arbitrary it follows that

$$\limsup_{N\to\infty}\frac{S_N^*}{\sqrt{N}\cdot\ln N} = 0 \text{ a.s.}$$

By replacing X_n by $-X_n$ for all n we may further conclude that

$$\limsup_{N\to\infty}\frac{(-S)_N^*}{\sqrt{N}\cdot\ln N} = 0$$

and together these two statements proves Eq. (13.40) since $|S|_N^* = \max\left\{S_N^*, (-S)_N^*\right\}$.

■

Remark 13.63. The above corollary is not a very good example as the results follow just as easily without the use of Doob's maximal inequality. Indeed, the proof above goes through the same with S_N^* replaced by S_N everywhere in which case we would conclude that $\lim_{N\to\infty}\frac{S_N}{\sqrt{N}\ln N} = 0$ or equivalently that

for all $\varepsilon > 0$, $|S_N| \le \varepsilon\sqrt{N}\ln N$ for all $N \ge N_\varepsilon$ where N_ε is a finite random variable. For $N \ge N_\varepsilon$, we have

$$|S|_N^* \le \left(\varepsilon\sqrt{N}\ln N\right) \vee |S|_{N_\varepsilon}^* \implies |S|_N^* \le \left(\varepsilon\sqrt{N}\ln N\right) \text{ for almost all } N.$$

This then shows that

$$\mathbb{P}\left(\limsup_{N\to\infty} \frac{|S|_N^*}{\sqrt{N}\ln N} \le \varepsilon\right) = 1$$

and as $\varepsilon > 0$ was arbitrary we may conclude that

$$\mathbb{P}\left(\limsup_{N\to\infty} \frac{|S|_N^*}{\sqrt{N}\ln N} = 0\right) = 1.$$

13.7 * L^p – inequalities

Lemma 13.64. *Suppose that X and Y are two non-negative random variables such that $\mathbb{P}(Y \ge y) \le \frac{1}{y}\mathbb{E}[X : Y \ge y]$ for all $y > 0$. Then for all $p \in (1,\infty)$,*

$$\mathbb{E}Y^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p. \tag{13.42}$$

Proof. We will begin by proving Eq. (13.42) under the additional assumption that $Y \in L^p(\Omega, \mathcal{B}, \mathbb{P})$. Since

$$\begin{aligned}
\mathbb{E}Y^p &= p\mathbb{E}\int_0^\infty 1_{y\le Y}\cdot y^{p-1}dy = p\int_0^\infty \mathbb{E}[1_{y\le Y}]\cdot y^{p-1}dy \\
&= p\int_0^\infty \mathbb{P}(Y \ge y)\cdot y^{p-1}dy \le p\int_0^\infty \frac{1}{y}\mathbb{E}[X : Y \ge y]\cdot y^{p-1}dy \\
&= p\mathbb{E}\int_0^\infty X1_{y\le Y}\cdot y^{p-2}dy = \frac{p}{p-1}\mathbb{E}\left[XY^{p-1}\right].
\end{aligned}$$

Now apply Hölder's inequality, with $q = p(p-1)^{-1}$, to find

$$\mathbb{E}\left[XY^{p-1}\right] \le \|X\|_p \cdot \left\|Y^{p-1}\right\|_q = \|X\|_p \cdot \left[\mathbb{E}|Y|^p\right]^{1/q}.$$

Combining thew two inequalities shows and solving for $\|Y\|_p$ shows $\|Y\|_p \le \frac{p}{p-1}\|X\|_p$ which proves Eq. (13.42) under the additional restriction of Y being in $L^p(\Omega, \mathcal{B}, \mathbb{P})$.

To remove the integrability restriction on Y, for $M > 0$ let $Z := Y \wedge M$ and observe that

$$\mathbb{P}(Z \ge y) = \mathbb{P}(Y \ge y) \le \frac{1}{y}\mathbb{E}[X : Y \ge y] = \frac{1}{y}\mathbb{E}[X : Z \ge y] \text{ if } y \le M$$

while

$$\mathbb{P}(Z \ge y) = 0 = \frac{1}{y}\mathbb{E}[X : Z \ge y] \text{ if } y > M.$$

Since Z is bounded, the special case just proved shows

$$\mathbb{E}[(Y \wedge M)^p] = \mathbb{E}Z^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p.$$

We may now use the MCT to pass to the limit, $M \uparrow \infty$, and hence conclude that Eq. (13.42) holds in general. ■

Corollary 13.65 (Doob's Inequality). *If $X = \{X_n\}_{n=0}^\infty$ be a non-negative submartingale and $1 < p < \infty$, then*

$$\mathbb{E}X_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_N^p. \tag{13.43}$$

Proof. Equation 13.43 follows by applying Lemma 13.64 with the aid of Proposition 13.57. ■

Corollary 13.66 (Doob's Inequality). *If $\{M_n\}_{n=0}^\infty$ is a martingale and $1 < p < \infty$, then for all $a > 0$ we have*

$$\mathbb{P}\left(|M|_N^* \ge a\right) \le \frac{1}{a}\mathbb{E}[|M|_N : M_N^* \ge a] \le \frac{1}{a}\mathbb{E}[|M_N|] \tag{13.44}$$

and

$$\mathbb{E}|M|_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_N|^p. \tag{13.45}$$

Proof. By the conditional Jensen's inequality (Theorem 13.22), it follows that $X_n := |M_n|$ is a submartingale. Hence Eq. (13.44) follows from Eq. (13.34) and Eq. (13.45) follows from Eq. (13.43). ■

Example 13.67. 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 $|S|_n^* = \max_{j\le n}|S_j|$. Since $\{S_n\}_{n=0}^\infty$ is a martingale, by cJensen's inequality (Theorem 13.22), $\{|S_n|^p\}_{n=1}^\infty$ is a (possibly extended) submartingale for any $p \in [1,\infty)$. Therefore an application of Eq. (13.34) of Proposition 13.57 show

$$\mathbb{P}\left(|S|_N^* \ge \alpha\right) = \mathbb{P}\left(|S|_N^{*p} \ge \alpha^p\right) \le \frac{1}{\alpha^p}\mathbb{E}[|S_N|^p : S_N^* \ge \alpha].$$

(When $p = 2$, this is Kolmogorov's inequality.) From Corollary 13.66 we also know that

$$\mathbb{E}\left|S\right|_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left|S_N\right|^p .$$

In particular when $p = 2$, this inequality becomes,

$$\mathbb{E}\left|S\right|_N^{*2} \le 4 \cdot \mathbb{E}\left|S_N\right|^2 = 4 \cdot \sum_{n=1}^N \mathbb{E}\left|X_n\right|^2 .$$

13.8 Martingale Exercises

(The next four problems were taken directly from
http://math.nyu.edu/~sheff/martingalenote.pdf.)

Exercise 13.7. Suppose Harriet has 7 dollars. Her plan is to make one dollar bets on fair coin tosses until her wealth reaches either 0 or 50, and then to go home. What is the expected amount of money that Harriet will have when she goes home? What is the probability that she will have 50 when she goes home?

Exercise 13.8. Consider a contract that at time N will be worth either 100 or 0. Let S_n be its price of the contract at time $0 \le n \le N$. If S_n is a martingale, and $S_0 = 47$, then what is the probability that the contract will be worth 100 at time N?

Exercise 13.9. Pedro plans to buy the contract in the previous problem at time 0 and sell it the first time T at which the price goes above 55 or below 15. What is the expected value of S_T? You may assume that the value, S_n, of the contract is bounded – there is only a finite amount of money in the world up to time N. Also note, by assumption, $T \le N$.

Exercise 13.10. 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.)

13.8.1 More Random Walk Exercises

For the next four exercises, let $\{Z_n\}_{n=1}^\infty$ be a sequence of Bernoulli random variables with $\mathbb{P}(Z_n = \pm 1) = \frac{1}{2}$ and let $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$. Then S becomes a martingale relative to the filtration, $\mathcal{F}_n := \sigma(Z_1, \ldots, Z_n)$ with $\mathcal{F}_0 := \{\emptyset, \Omega\}$ – of course S_n is the (fair) simple random walk on $\mathbb{Z}$. For any $a \in \mathbb{Z}$, let

$$\sigma_a := \inf\{n : S_n = a\} .$$

Exercise 13.11. For $a < 0 < b$ with $a, b \in \mathbb{Z}$, let $\tau = \sigma_a \wedge \sigma_b$. Explain why $\{S_n^\tau\}_{n=0}^\infty$ is a bounded martingale use this to show $\mathbb{P}(\tau = \infty) = 0$. **Hint:** make use of the fact that $|S_n - S_{n-1}| = |Z_n| = 1$ for all n and hence the only way $\lim_{n\to\infty} S_n^\tau$ can exist is if it stops moving!

Exercise 13.12. In this exercise, you are asked to use the central limit Theorem to prove again that $\mathbb{P}(\tau = \infty) = 0$, Exercise 13.11. **Hints:** Use the central limit theorem to show

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-x^2/2} dx \ge f(0) \mathbb{P}(\tau = \infty) \tag{13.46}$$

for all $f \in C^3(\mathbb{R} \to [0, \infty))$ with $M := \sup_{x\in\mathbb{R}} \left|f^{(3)}(x)\right| < \infty$. Use this inequality to conclude that $\mathbb{P}(\tau = \infty) = 0$.

Exercise 13.13. Show

$$\mathbb{P}(\sigma_b < \sigma_a) = \frac{|a|}{b + |a|} \tag{13.47}$$

and use this to conclude $\mathbb{P}(\sigma_b < \infty) = 1$, i.e. every $b \in \mathbb{N}$ is almost surely visited by S_n.

Hint: As in Exercise 13.11 notice that $\{S_n^\tau\}_{n=0}^\infty$ is a bounded martingale where $\tau := \sigma_a \wedge \sigma_b$. Now compute $\mathbb{E}[S_\tau] = \mathbb{E}[S_\tau^\tau]$ in two different ways.

Exercise 13.14. 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 13.4 above to conclude $N_n := S_n^2 - n$ is a martingale.
2. Now show
$$0 = \mathbb{E}N_0 = \mathbb{E}N_{\tau\wedge n} = \mathbb{E}S_{\tau\wedge n}^2 - \mathbb{E}[\tau \wedge n] . \tag{13.48}$$
3. Now use DCT and MCT along with Exercise 13.13 to compute the limit as $n \to \infty$ in Eq. (13.48) to find
$$\mathbb{E}[\sigma_a \wedge \sigma_b] = \mathbb{E}[\tau] = b|a| . \tag{13.49}$$
4. By considering the limit, $a \to -\infty$ in Eq. (13.49), show $\mathbb{E}[\sigma_b] = \infty$.

For the next group of exercise we are now going to suppose that $\mathbb{P}(Z_n = 1) = p > \frac{1}{2}$ and $\mathbb{P}(Z_n = -1) = q = 1 - p < \frac{1}{2}$. As before let $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$, $S_0 = 0$ 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 6.4 above.

In order to follow the procedures above, we start by looking for a function, φ, such that $\varphi(S_n)$ is a martingale. Such a function must satisfy,

$$\varphi(S_n) = \mathbb{E}_{\mathcal{F}_n}\varphi(S_{n+1}) = \varphi(S_n + 1)p + \varphi(S_n - 1)q,$$

and this then leads us to try to solve the following difference equation for φ;

$$\varphi(x) = p\varphi(x+1) + q\varphi(x-1) \text{ for all } x \in \mathbb{Z}. \tag{13.50}$$

Similar to the theory of second order ODE's this equation has two linearly independent solutions which could be found by solving Eq. (13.50) 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 λ to be determined. Doing so leads to the equation, $\lambda^x = p\lambda^{x+1} + q\lambda^{x-1}$, or equivalently to the **characteristic equation,**

$$p\lambda^2 - \lambda + q = 0.$$

The solutions to this equation are

$$\begin{aligned}\lambda &= \frac{1 \pm \sqrt{1-4pq}}{2p} = \frac{1 \pm \sqrt{1-4p(1-p)}}{2p} \\ &= \frac{1 \pm \sqrt{4p^2-4p+1}}{2p} = \frac{1 \pm \sqrt{(2p-1)^2}}{2p} = \{1, (1-p)/p\} = \{1, q/p\}.\end{aligned}$$

The most general solution to Eq. (13.50) is then given by

$$\varphi(x) = A + B(q/p)^x.$$

Below we will take $A = 0$ and $B = 1$. As before let $\sigma_a = \inf\{n \geq 0 : S_n = a\}$.

Exercise 13.15. Let $a < 0 < b$ and $\tau := \sigma_a \wedge \sigma_b$ and recall that $p > 1/2$ and $q = 1 - p$.

1. Apply the method in Exercise 13.11 with S_n replaced by $M_n := (q/p)^{S_n}$ to show $\mathbb{P}(\tau = \infty) = 0$. [Recall that $\{M_n\}_{n=1}^\infty$ is a martingale as explained in Example 13.13.]
2. Now use the method in Exercise 13.13 to show
$$\mathbb{P}(\sigma_a < \sigma_b) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{13.51}$$
3. By letting $a \to -\infty$ in Eq. (13.51), conclude $\mathbb{P}(\sigma_b = \infty) = 0$.
4. By letting $b \to \infty$ in Eq. (13.51), conclude $\mathbb{P}(\sigma_a < \infty) = (q/p)^{|a|}$.

Exercise 13.16. 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 13.4 or Exercise 13.6.)

Exercise 13.17. Using exercise 13.16, show

$$\mathbb{E}(\sigma_a \wedge \sigma_b) = \left(\frac{b[1-(q/p)^a] + a\left[(q/p)^b - 1\right]}{(q/p)^b - (q/p)^a}\right)(p-q)^{-1}. \tag{13.52}$$

By considering the limit of this equation as $a \to -\infty$, show

$$\mathbb{E}[\sigma_b] = \frac{b}{p-q}$$

and by considering the limit as $b \to \infty$, show $\mathbb{E}[\sigma_a] = \infty$.

13.8.2 More advanced martingale exercises

Exercise 13.18. 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 $\lambda > 0$,

$$\mathbb{P}\left(\max_{1\leq m\leq n} M_m \geq \lambda\right) \leq \frac{E[M_n^2]}{E[M_n^2] + \lambda^2}.$$

Hints: First show that for any $c > 0$ that $\left\{X_n := (M_n + c)^2\right\}_{n=0}^\infty$ is a submartingale and then observe,

$$\left\{\max_{1\leq m\leq n} M_m \geq \lambda\right\} \subset \left\{\max_{1\leq m\leq n} X_n \geq (\lambda + c)^2\right\}.$$

Now use Doob' Maximal inequality (Proposition 13.57) to estimate the probability of the last set and then choose c so as to optimize the resulting estimate you get for $\mathbb{P}(\max_{1\leq m\leq n} M_m \geq \lambda)$. (Notice that this result applies to $-M_n$ as well so it also holds that;

$$\mathbb{P}\left(\min_{1\leq m\leq n} M_m \leq -\lambda\right) \leq \frac{E[M_n^2]}{E[M_n^2] + \lambda^2} \text{ for all } \lambda > 0.$$

Exercise 13.19. Let $\{Z_n\}_{n=1}^\infty$ be independent random variables, $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$, and $f_n(\lambda) := \mathbb{E}\left[e^{i\lambda Z_n}\right]$. Suppose $\mathbb{E}e^{i\lambda S_n} = \prod_{n=1}^N f_n(\lambda)$ converges to a continuous function, $F(\lambda)$, as $N \to \infty$. Show for each $\lambda \in \mathbb{R}$ that

$$\mathbb{P}\left(\lim_{n\to\infty} e^{i\lambda S_n} \text{ exists}\right) = 1. \tag{13.53}$$

Hints:

1. Show it is enough to find an $\varepsilon > 0$ such that Eq. (13.53) holds for $|\lambda| \leq \varepsilon$.

2. Choose $\varepsilon > 0$ such that $|F(\lambda) - 1| < 1/2$ for $|\lambda| \le \varepsilon$. For $|\lambda| \le \varepsilon$, show $M_n(\lambda) := \frac{e^{i\lambda S_n}}{\mathbb{E}e^{i\lambda S_n}}$ is a bounded complex[8] martingale relative to the filtration, $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$.

Lemma 13.68 (Protter [16, See the lemma on p. 22.]). *Let $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ such that $\left\{e^{iux_n}\right\}_{n=1}^{\infty}$ is convergent for Lebesgue almost every $u \in \mathbb{R}$. Then $\lim_{n\to\infty} x_n$ exists in $\mathbb{R}$.*

Proof. Let U be a uniform random variable with values in $[0,1]$. By assumption, for any $t \in \mathbb{R}$, $\lim_{n\to\infty} e^{itUx_n}$ exists a.s. Thus if n_k and m_k are any increasing sequences we have

$$\lim_{k\to\infty} e^{itUx_{n_k}} = \lim_{n\to\infty} e^{itUx_n} = \lim_{k\to\infty} e^{itUx_{m_k}} \text{ a.s.}$$

and therefore,

$$e^{it\left(Ux_{n_k} - Ux_{m_k}\right)} = \frac{e^{itUx_{n_k}}}{e^{itUx_{m_k}}} \to 1 \text{ a.s. as } k \to \infty.$$

Hence by DCT it follows that

$$\mathbb{E}\left[e^{it\left(Ux_{n_k} - Ux_{m_k}\right)}\right] \to 1 \text{ as } k \to \infty$$

and therefore

$$(x_{n_k} - x_{m_k}) \cdot U = Ux_{n_k} - Ux_{m_k} \to 0$$

in distribution and hence in probability. But his can only happen if $(x_{n_k} - x_{m_k}) \to 0$ as $k \to \infty$. As $\{n_k\}$ and $\{m_k\}$ were arbitrary, this suffices to show $\{x_n\}$ is a Cauchy sequence. ■

Exercise 13.20 (Continuation of Exercise 13.19 – See Doob [5, Chapter VII.5]). Let $\{Z_n\}_{n=1}^{\infty}$ be independent random variables. Use Exercise 13.19 an Lemma 13.68 to prove the series, $\sum_{n=1}^{\infty} Z_n$, converges in $\mathbb{R}$ a.s. iff $\prod_{n=1}^{N} f_n(\lambda)$ converges to a continuous function, $F(\lambda)$ as $N \to \infty$. Conclude from this that $\sum_{n=1}^{\infty} Z_n$ is a.s. convergent iff $\sum_{n=1}^{\infty} Z_n$ is convergent in distribution.

[8] 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.

Some martingale Examples and Applications

Exercise 14.1. 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 σ^2, i.e. $Z_n \overset{d}{=} \mu + \sigma N$ where N is a standard normal random variable.[1] Let

$$\begin{aligned} \tau &= \inf\{n : S_n \leq 0\} \text{ and} \\ R &= \{\tau < \infty\} = \{S_n \leq 0 \text{ for some } n\} \end{aligned}$$

be the event that the company eventually becomes bankrupt, i.e. is **R**uined. Show

$$\mathbb{P}(\text{Ruin}) = \mathbb{P}(R) \leq e^{-2(c-\mu)S_0/\sigma^2}.$$

Outline:

1. Show that $\lambda = -2(c-\mu)/\sigma^2 < 0$ satisfies, $\mathbb{E}\left[e^{\lambda \xi_n}\right] = 1$.
2. With this λ show (using $S_n = S_0 + \xi_1 + \cdots + \xi_n$)

$$Y_n := \exp(\lambda S_n) = e^{\lambda S_0} \prod_{j=1}^{n} e^{\lambda \xi_j} \tag{14.1}$$

   is a non-negative $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$ – 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 $\mathbb{E}Y_\tau \leq e^{\lambda S_0}$.
4. Finally conclude that

$$\mathbb{P}(R) = \mathbb{P}(\tau < \infty) \leq \mathbb{E}[Y_\tau : \tau < \infty] \leq \mathbb{E}Y_\tau \leq e^{\lambda S_0} = e^{-2(c-\mu)S_0/\sigma^2}.$$

Observe that by the strong law of large numbers that $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}\xi_1 = c - \mu > 0$ a.s. Thus for large n we have $S_n \sim n(c-\mu) \to \infty$ as $n \to \infty$. The question we have addressed is what happens to the S_n for intermediate values – in particular what is the likelyhood that S_n makes a sufficiently "large deviation" from the "typical" value of $n(c-\mu)$ in order for the company to go bankrupt. The next section is a detour which gives a taste of how such probabilities may be estimated.

[1] The number c is to be interpreted as the yearly premium, μ represents the mean payout in claims per year, and σN represents the random fluctuations which can occur from year to year.

14.1 A Large Deviations Primer

Definition 14.1. *A real valued random variable, Z, is said to be* ***exponentially integrable*** *if $\mathbb{E}\left[e^{\theta Z}\right] < \infty$ for all $\theta \in \mathbb{R}$. Under this condition we let $M(\theta) = \mathbb{E}\left[e^{\theta Z}\right]$ be the* ***moment generating function*** *and*

$$\psi(\theta) = \ln M(\theta) = \ln \mathbb{E}\left[e^{\theta Z}\right]$$

be the ***log-moment generating function.***

Summary. The upshot of Theorems 14.2 and 14.3 and of this section is; if $S_n := Z_1 + \cdots + Z_n$ where $\{Z_n\}_{n=1}^{\infty}$ be i.i.d. exponentially integrable mean zero random variables, then

$$\mathbb{P}(S_n \geq n\ell) \sim e^{-n \cdot \max_\theta [\theta\ell - \psi(\theta)]}$$

Theorem 14.2 (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 $\mathbb{E}Z = 0$ where $Z \overset{d}{=} Z_n$. Then for all $\ell > 0$,*

$$\mathbb{P}(S_n \geq n\ell) \leq e^{-nI(\ell)} \tag{14.2}$$

where

$$I(\ell) = \sup_{\theta \geq 0} (\theta\ell - \psi(\theta)) = \sup_{\theta \in \mathbb{R}} (\theta\ell - \psi(\theta)) \geq 0. \tag{14.3}$$

In particular,

$$\limsup_{n\to\infty} \ln \mathbb{P}(S_n \geq n\ell) \leq -I(\ell) \text{ for all } \ell > 0. \tag{14.4}$$

Proof. Let $\{Z_n\}_{n=1}^{\infty}$ be i.i.d. exponentially integrable random variables such that $\mathbb{E}Z = 0$ where $Z \overset{d}{=} Z_n$. Then for $\ell > 0$ we have for any $\theta \geq 0$ that

$$\mathbb{P}(S_n \geq n\ell) = \mathbb{P}\left(e^{\theta S_n} \geq e^{\theta n \ell}\right) \leq e^{-\theta n \ell} \mathbb{E}\left[e^{\theta S_n}\right] = \left(e^{-\theta\ell}\mathbb{E}\left[e^{\theta Z}\right]\right)^n.$$

Let $M(\theta) := \mathbb{E}\left[e^{\theta Z}\right]$ be the moment generating function for Z and $\psi(\theta) = \ln M(\theta)$ be the log – moment generating function. Then we have just shown,

$$\mathbb{P}(S_n \geq n\ell) \leq \exp(-n(\theta\ell - \psi(\theta))) \text{ for all } \theta \geq 0.$$

Minimizing the right side of this inequality over $\theta \geq 0$ gives the upper bound in Eq. (14.2) where $I(\ell)$ is given as in the first equality in Eq. (14.3).

To prove the second equality in Eq. (14.3), we use the fact that $e^{\theta x}$ is a convex function in x for all $\theta \in \mathbb{R}$ and therefore by Jensen's inequality,

$$M(\theta) = \mathbb{E}\left[e^{\theta Z}\right] \geq e^{\theta \mathbb{E} X} = e^{\theta 0} = 1 \text{ for all } \theta \in \mathbb{R}.$$

This then implies that $\psi(\theta) = \ln M(\theta) \geq 0$ for all $\theta \in \mathbb{R}$. In particular, $\theta\ell - \psi(\theta) < 0$ for $\theta < 0$ while $[\theta\ell - \psi(\theta)]|_{\theta=0} = 0$ and therefore

$$\sup_{\theta\in\mathbb{R}}(\theta\ell - \psi(\theta)) = \sup_{\theta\geq 0}(\theta\ell - \psi(\theta)) \geq 0.$$

This completes the proof as Eq. (14.4) easily follows from Eq. (14.2). ■

Theorem 14.3 (Large Deviation Lower Bound). *If there is a maximizer, θ_0, for the the function $\theta \to \theta\ell - \psi(\theta)$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}(S_n \geq n\ell) \geq -I(\ell) = \theta_0\ell - \psi(\theta_0). \tag{14.5}$$

Proof. If there is a maximizer, θ_0, for the the function $\theta \to \theta\ell - \psi(\theta)$, then

$$0 = \ell - \psi'(\theta_0) = \ell - \frac{M'(\theta_0)}{M(\theta_0)} = \ell - \frac{\mathbb{E}\left[Ze^{\theta_0 Z}\right]}{M(\theta_0)}.$$

Thus if W is a random variable with Law determined by

$$\mathbb{E}[f(W)] = M(\theta_0)^{-1}\mathbb{E}\left[f(Z)e^{\theta_0 Z}\right]$$

for all non-negative functions $f : \mathbb{R} \to [0,\infty]$ then $\mathbb{E}[W] = \ell$.

Suppose that $\{W_n\}_{n=1}^\infty$ has been chosen to be a sequence of i.i.d. random variables such that $W_n \overset{d}{=} W$ for all n. Then, for all non-negative functions $f : \mathbb{R}^n \to [0,\infty]$ we have

$$\begin{aligned}\mathbb{E}[f(W_1,\ldots,W_n)] &= M(\theta_0)^{-n}\mathbb{E}\left[f(Z_1,\ldots,Z_n)\prod_{i=1}^n e^{\theta_0 Z_i}\right] \\ &= M(\theta_0)^{-n}\mathbb{E}\left[f(Z_1,\ldots,Z_n)e^{\theta_0 S_n}\right].\end{aligned}$$

This is easily verified by showing the right side of this equation gives the correct expectations when f is a product function. Replacing $f(z_1,\ldots,z_n)$ by $M(\theta_0)^n e^{-\theta_0(z_1+\ldots z_n)} f(z_1,\ldots,z_n)$ in the previous equation then shows

$$\mathbb{E}[f(Z_1,\ldots,Z_n)] = M(\theta_0)^n \mathbb{E}\left[f(W_1,\ldots,W_n)e^{-\theta_0 T_n}\right] \tag{14.6}$$

where $T_n := W_1 + \cdots + W_n$.

Taking $\delta > 0$ and $f(z_1,\ldots,z_n) = 1_{z_1+\ldots z_n \geq n\ell}$ in Eq. (14.6) shows

$$\begin{aligned}\mathbb{P}(S_n \geq n\ell) &= M(\theta_0)^n \mathbb{E}\left[e^{-\theta_0 T_n} : n\ell \leq T_n\right] \\ &\geq M(\theta_0)^n \mathbb{E}\left[e^{-\theta_0 T_n} : n\ell \leq T_n \leq n(\ell+\delta)\right] \\ &\geq M(\theta_0)^n e^{-n\theta_0(\ell+\delta)}\mathbb{P}\left[n\ell \leq T_n \leq n(\ell+\delta)\right] \\ &= e^{-nI(\ell)}e^{-n\theta_0\delta}\mathbb{P}\left[n\ell \leq T_n \leq n(\ell+\delta)\right].\end{aligned}$$

Taking logarithms of this equation, then dividing by n, then letting $n \to \infty$ we learn

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}(S_n \geq n\ell) &\geq -I(\ell) - \theta_0\ell\delta + \lim_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left[n\ell \leq T_n \leq n(\ell+\delta)\right] \\ &= -I(\ell) - \theta_0\ell\delta + 0\end{aligned} \tag{14.7}$$

wherein have used the central limit theorem to argue that

$$\begin{aligned}\mathbb{P}\left[n\ell \leq T_n \leq n(\ell+\delta)\right] &= \mathbb{P}\left[0 \leq T_n - n\ell \leq n\delta\right] \\ &= \mathbb{P}\left[0 \leq \frac{T_n - n\ell}{\sqrt{n}} \leq \sqrt{n}\delta\right] \to \frac{1}{2} \text{ as } n \to \infty.\end{aligned}$$

Equation (14.5) now follows from Eq. (14.7) as $\delta > 0$ is arbitrary. ■

Example 14.4. Suppose that $Z \overset{d}{=} N(0,\sigma^2) \overset{d}{=} \sigma N$ where $N \overset{d}{=} N(0,1)$, then

$$M(\theta) = \mathbb{E}\left[e^{\theta Z}\right] = \mathbb{E}\left[e^{\theta\sigma N}\right] = \exp\left(\frac{1}{2}(\sigma\theta)^2\right)$$

and therefore $\psi(\theta) = \ln M(\theta) = \frac{1}{2}\sigma^2\theta^2$. Moreover for $\ell > 0$,

$$\ell = \psi'(\theta) \implies \ell = \sigma^2\theta \implies \theta_0 = \frac{\ell}{\sigma^2}.$$

Thus it follows that

$$I(\ell) = \theta_0\ell - \psi(\theta_0) = \frac{\ell^2}{\sigma^2} - \frac{1}{2}\sigma^2\left(\frac{\ell}{\sigma^2}\right)^2 = \frac{1}{2}\frac{\ell^2}{\sigma^2}.$$

In this Gaussian case we actually know that $S_n \overset{d}{=} N(0,n\sigma^2)$ and therefore by Mill's ratio,

$$\begin{aligned}\mathbb{P}(S_n \geq n\ell) &= \mathbb{P}\left(\sqrt{n}\sigma N \geq n\ell\right) = \mathbb{P}\left(N \geq \sqrt{n}\frac{\ell}{\sigma}\right) \\ &\sim \frac{1}{\sqrt{2\pi n}\frac{\ell}{\sigma}}e^{-n\frac{1}{2}\frac{\ell^2}{\sigma^2}} = \frac{\sigma}{\sqrt{2\pi n}\ell}e^{-nI(\ell)} \text{ as } n \to \infty.\end{aligned}$$

Remark 14.5. The technique used in the proof of Theorem 14.3 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 variable of the form

$$\mathbb{E}\left[f\left(W\right)\right]=\frac{\mathbb{E}\left[f\left(Z\right)\rho\left(Z\right)\right]}{\mathbb{E}\left[\rho\left(Z\right)\right]}$$

where ρ is some positive function. Under this change of measure the analogue of Eq. (14.6) is

$$\mathbb{E}\left[f\left(Z_1,\dots,Z_n\right)\right]=\left(\mathbb{E}\left[\rho\left(Z\right)\right]\right)^n\cdot\mathbb{E}\left[f\left(W_1,\dots,W_n\right)\prod_{j=1}^n\frac{1}{\rho\left(W_i\right)}\right].$$

However to make this change of variable easy to deal with in the setting at hand we would like to further have

$$\prod_{j=1}^n\frac{1}{\rho\left(W_j\right)}=f_n\left(T_n\right)=f_n\left(W_1+\cdots+W_n\right)$$

for some function f_n. Equivalently we would like, for some function g_n, that

$$\prod_{j=1}^n\rho\left(w_j\right)=g_n\left(w_1+\cdots+w_n\right)$$

for all w_i. Taking logarithms of this equation and differentiating in the w_j and w_k variables shows,

$$\frac{\rho'\left(w_j\right)}{\rho\left(w_j\right)}=\left(\ln g_n\right)'\left(w_1+\cdots+w_n\right)=\frac{\rho'\left(w_k\right)}{\rho\left(w_k\right)}.$$

From the extremes of this last equation we conclude that $\rho'\left(w_j\right)/\rho\left(w_j\right)=c$ (for some constant c) and therefore $\rho\left(w\right)=Ke^{cw}$ for some constant K. This helps to explain why the exponential function is used in the above proof.

14.2 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 Example 13.15. Recall that if the urn contains r red balls and g green balls at a given time we draw one of these balls at random and replace it and add c more balls of the same color drawn. Let (r_n,g_n) be the number of red and green balls in the earn at time n. Then we have

$$\mathbb{P}\left(\left(r_{n+1},g_n\right)=\left(r+c,g\right)\mid\left(r_n,g_n\right)=\left(r,g\right)\right)=\frac{r}{r+g}\text{ and}$$
$$\mathbb{P}\left(\left(r_{n+1},g_n\right)=\left(r,g+c\right)\mid\left(r_n,g_n\right)=\left(r,g\right)\right)=\frac{g}{r+g}.$$

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

$$X_n:=\frac{g_n}{r_n+g_n}=\frac{g_n}{r_0+g_0+nc}$$

and recall from Example 13.15 or Remark 13.16 that $\{X_n\}_{n=0}^\infty$ is a martingale relative to

$$\mathcal{F}_n:=\sigma\left(\left(r_k,g_k\right):k\le n\right)=\sigma\left(X_k:k\le n\right).$$

Since $X_n\ge 0$ and $\mathbb{E}X_n=\mathbb{E}X_0<\infty$ for all n it follows by Theorem 13.43 that $X_\infty:=\lim_{n\to\infty}X_n$ exists a.s. The distribution of X_∞ is described in the next theorem.

Theorem 14.6. *Let $\gamma:=g/c$ and $\rho:=r/c$ and $\mu:=\mathrm{Law}_{\mathbb{P}}\left(X_\infty\right)$. Then μ is the* ***beta distribution on*** $[0,1]$ *with parameters, γ,ρ, i.e.*

$$d\mu\left(x\right)=\frac{\Gamma\left(\rho+\gamma\right)}{\Gamma\left(\rho\right)\Gamma\left(\gamma\right)}x^{\gamma-1}\left(1-x\right)^{\rho-1}dx\text{ for }x\in[0,1]. \tag{14.8}$$

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

$$\frac{g\cdot\left(g+c\right)\cdot\dots\cdot\left(g+\left(n-1\right)c\right)\cdot r\cdot\left(r+c\right)\cdot\dots\cdot\left(r+\left(n-m-1\right)c\right)}{\left(r+g\right)\cdot\left(r+g+c\right)\cdot\dots\cdot\left(r+g+\left(n-1\right)c\right)}.$$

Since this is the same probability for any of the $\binom{n}{m}$ – ways of drawing m greens and $n-m$ reds in n draws we have

$$\begin{aligned}&\mathbb{P}\left(\text{Draw }m-\text{greens}\right)\\&=\binom{n}{m}\frac{g\cdot\left(g+c\right)\cdot\dots\cdot\left(g+\left(m-1\right)c\right)\cdot r\cdot\left(r+c\right)\cdot\dots\cdot\left(r+\left(n-m-1\right)c\right)}{\left(r+g\right)\cdot\left(r+g+c\right)\cdot\dots\cdot\left(r+g+\left(n-1\right)c\right)}\\&=\binom{n}{m}\frac{\gamma\cdot\left(\gamma+1\right)\cdot\dots\cdot\left(\gamma+\left(m-1\right)\right)\cdot\rho\cdot\left(\rho+1\right)\cdot\dots\cdot\left(\rho+\left(n-m-1\right)\right)}{\left(\rho+\gamma\right)\cdot\left(\rho+\gamma+1\right)\cdot\dots\cdot\left(\rho+\gamma+\left(n-1\right)\right)}.\end{aligned} \tag{14.9}$$

Before going to the general case let us warm up with the special case, $g = r = c = 1$. In this case Eq. (14.9) becomes,

$$\mathbb{P}(\text{Draw } m - \text{greens}) = \binom{n}{m} \frac{1 \cdot 2 \cdot \cdots \cdot m \cdot 1 \cdot 2 \cdot \cdots \cdot (n-m)}{2 \cdot 3 \cdot \cdots \cdot (n+1)} = \frac{1}{n+1}.$$

On the set, $\{\text{Draw } m - \text{greens}\}$, we have $X_n = \frac{1+m}{2+n}$ and hence it follows that for any $f \in C([0,1])$ that

$$\begin{aligned}\mathbb{E}[f(X_n)] &= \sum_{m=0}^{n} f\left(\frac{m+1}{n+2}\right) \cdot \mathbb{P}(\text{Draw } m - \text{greens}) \\ &= \sum_{m=0}^{n} f\left(\frac{m+1}{n+2}\right) \frac{1}{n+1}.\end{aligned}$$

Therefore

$$\mathbb{E}[f(X)] = \lim_{n\to\infty} \mathbb{E}[f(X_n)] = \int_0^1 f(x)\,dx \tag{14.10}$$

and hence we may conclude that X_∞ has the uniform distribution on $[0,1]$.

For the general case, recall that $n! = \Gamma(n+1)$, $\Gamma(t+1) = t\Gamma(t)$, and therefore for $m \in \mathbb{N}$,

$$\Gamma(x+m) = (x+m-1)(x+m-2)\ldots(x+1)\,x\Gamma(x). \tag{14.11}$$

Also recall Stirling's formula,

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x}[1 + r(x)] \tag{14.12}$$

where $|r(x)| \to 0$ as $x \to \infty$. To finish the proof we will follow the strategy of the proof of Eq. (14.10) using Stirling's formula to estimate the expression for $\mathbb{P}(\text{Draw } m - \text{greens})$ in Eq. (14.9).

On the set, $\{\text{Draw } m - \text{greens}\}$, we have

$$X_n = \frac{g + mc}{r + g + nc} = \frac{\gamma + m}{\rho + \gamma + n} =: x_m,$$

where $\rho := r/c$ and $\gamma := g/c$. For later notice that $\Delta_m x = \frac{\gamma}{\rho+\gamma+n}$.

Using this notation we may rewrite Eq. (14.9) as

$$\begin{aligned}&\mathbb{P}(\text{Draw } m - \text{greens}) \\ &= \binom{n}{m} \frac{\frac{\Gamma(\gamma+m)}{\Gamma(\gamma)} \cdot \frac{\Gamma(\rho+n-m)}{\Gamma(\rho)}}{\frac{\Gamma(\rho+\gamma+n)}{\Gamma(\rho+\gamma)}} \\ &= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)} \frac{\Gamma(\gamma+m)\Gamma(\rho+n-m)}{\Gamma(\rho+\gamma+n)}.\end{aligned} \tag{14.13}$$

Now by Stirling's formula,

$$\begin{aligned}\frac{\Gamma(\gamma+m)}{\Gamma(m+1)} &= \frac{(\gamma+m)^{\gamma+m-1/2} e^{-(\gamma+m)}[1 + r(\gamma+m)]}{(1+m)^{m+1-1/2} e^{-(m+1)}[1 + r(1+m)]} \\ &= (\gamma+m)^{\gamma-1} \cdot \left(\frac{\gamma+m}{m+1}\right)^{m+1/2} e^{-(\gamma-1)} \frac{1 + r(\gamma+m)}{1 + r(m+1)}. \\ &= (\gamma+m)^{\gamma-1} \cdot \left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2} e^{-(\gamma-1)} \frac{1 + r(\gamma+m)}{1 + r(m+1)}\end{aligned}$$

We will keep m fairly large, so that

$$\begin{aligned}\left(\frac{1+\gamma/m}{1+1/m}\right)^{m+1/2} &= \exp\left((m+1/2)\ln\left(\frac{1+\gamma/m}{1+1/m}\right)\right) \\ &\cong \exp((m+1/2)(\gamma/m - 1/m)) \cong e^{\gamma-1}.\end{aligned}$$

Hence we have

$$\frac{\Gamma(\gamma+m)}{\Gamma(m+1)} \asymp (\gamma+m)^{\gamma-1}.$$

Similarly, keeping $n - m$ fairly large, we also have

$$\begin{aligned}\frac{\Gamma(\rho+n-m)}{\Gamma(n-m+1)} &\asymp (\rho+n-m)^{\rho-1} \text{ and} \\ \frac{\Gamma(\rho+\gamma+n)}{\Gamma(n+1)} &\asymp (\rho+\gamma+n)^{\rho+\gamma-1}.\end{aligned}$$

Combining these estimates with Eq. (14.13) gives,

$$\begin{aligned}&\mathbb{P}(\text{Draw } m - \text{greens}) \\ &\asymp \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot \frac{(\gamma+m)^{\gamma-1} \cdot (\rho+n-m)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\ &= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot \frac{\left(\frac{\gamma+m}{\rho+\gamma+n}\right)^{\gamma-1} \cdot \left(\frac{\rho+n-m}{\rho+\gamma+n}\right)^{\rho-1}}{(\rho+\gamma+n)^{\rho+\gamma-1}} \\ &= \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot (x_m)^{\gamma-1} \cdot (1-x_m)^{\rho-1} \Delta_m x.\end{aligned}$$

Therefore, for any $f \in C([0,1])$, it follows that

$$\begin{aligned}\mathbb{E}[f(X_\infty)] &= \lim_{n\to\infty} \mathbb{E}[f(X_n)] \\ &= \lim_{n\to\infty} \sum_{m=0}^{n} f(x_m) \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} \cdot (x_m)^{\gamma-1} \cdot (1-x_m)^{\rho-1} \Delta_m x \\ &= \int_0^1 f(x) \frac{\Gamma(\rho+\gamma)}{\Gamma(\rho)\Gamma(\gamma)} x^{\gamma-1}(1-x)^{\rho-1}\,dx.\end{aligned}$$

14.3 Galton Watson Branching Process

This section is taken from [6, p. 245 –249].

Notation 14.7 *Let $\{p_k\}_{k=0}^{\infty}$ be a probability on $\mathbb{N}_0$ where $p_k := \mathbb{P}(\xi = k)$ is the **off-spring** distribution.*

We assume that the mean number of offspring,

$$\mu := \mathbb{E}\xi = \sum_{k=0}^{\infty} kp_k < \infty.$$

Notation 14.8 *Let $\{\xi_i^n : i, n \geq 1\}$ be a sequence of i.i.d. non-negative integer valued random variables with $\mathbb{P}(\xi_i^n = k) = p_k$ for all $k \in \mathbb{N}_0$. We also let $Y_n := \{\xi_1^n\}_{n=1}^{\infty}$ in order to shorten notation later.*

If Z_n is the number of "people" (or organisms) alive in the nth – generation then we assume the ith organism has ξ_i^{n+1} – offspring so that

$$\begin{aligned} Z_{n+1} &= \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} \\ &= \sum_{k=1}^{\infty} \left(\xi_1^{n+1} + \cdots + \xi_k^{n+1}\right) 1_{Z_n = k}. \end{aligned} \tag{14.14}$$

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 14.1.

Standing assumption: We suppose that $p_1 < 1$ for otherwise we will have $Z_n = 1$ for all n.

To understand Z_n a bit better observe that

$$\begin{aligned} Z_0 &= 1, \\ Z_1 &= \xi_{Z_0}^1 = \xi_1^1, \\ Z_2 &= \xi_1^2 + \cdots + \xi_{\xi_1^1}^2, \\ Z_3 &= \xi_1^3 + \cdots + \xi_{Z_2}^3, \\ &\vdots \end{aligned}$$

The sample path in Figure 14.1 corresponds to

$$\begin{aligned} &\xi_1^1 = 3, \\ &\xi_1^2 = 2,\ \xi_2^2 = 0,\ \xi_3^2 = 3, \\ &\xi_1^3 = \xi_2^3 = \xi_3^3 = \xi_4^3 = 0,\ \xi_5^3 = 4, \text{ and} \\ &\xi_1^4 = \xi_2^4 = \xi_3^4 = \xi_4^4 = 0. \end{aligned}$$

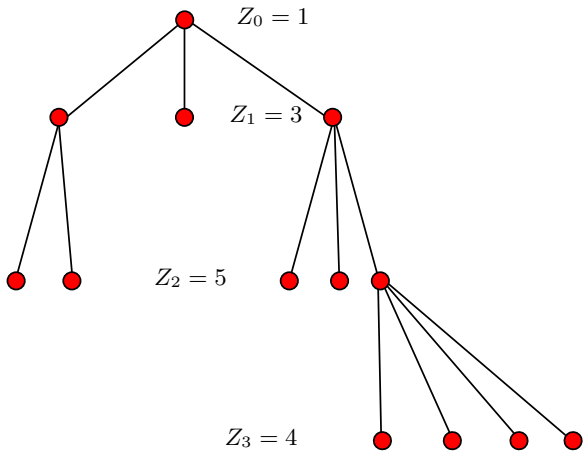


Fig. 14.1. A possible realization of a Galton Watson "tree."

In order to shorten the exposition we will make use of the following two intuitive **facts**:

1. The different branches of the Galton-Watson tree evolve independently of one another.
2. The process $\{Z_n\}_{n=0}^{\infty}$ is a Markov chain with transition probabilities,

$$\begin{aligned} p(k,l) &= \mathbb{P}(Z_{n+1} = l | Z_n = k) \\ &= \begin{cases} \delta_{0,l} & \text{if } k = 0 \\ \mathbb{P}(Y_1 + \cdots + Y_k = l) & \text{if } k \geq 0. \end{cases} \\ &= \delta_{k,0}\delta_{0,l} + 1_{k\geq 1} \cdot \mathbb{P}(Y_1 + \cdots + Y_k = l). \end{aligned} \tag{14.15}$$

Lemma 14.9. *If $f : \mathbb{N}_0 \to \mathbb{R}$ is a function then*

$$(\mathbf{P}f)(k) = \mathbb{E}[f(Y_1 + \cdots + Y_k)]$$

where $Y_1 + \cdots + Y_k := 0$ if $k = 0$.

Proof. This follows by the simple computation;

$$\begin{aligned} (\mathbf{P}f)(k) &= \sum_{l=0}^{\infty} p(k,l) f(l) = \sum_{l=0}^{\infty} [\delta_{k,0}\delta_{0,l} + 1_{k\geq 1} \cdot \mathbb{P}(Y_1 + \cdots + Y_k = l)] f(l) \\ &= \delta_{k,0} f(0) + 1_{k\geq 1} \sum_{l=0}^{\infty} [\mathbb{P}(Y_1 + \cdots + Y_k = l)] f(l) \\ &= \delta_{k,0} f(0) + 1_{k\geq 1} \mathbb{E}[f(Y_1 + \cdots + Y_k)] \\ &= \mathbb{E}[f(Y_1 + \cdots + Y_k)]. \end{aligned}$$

■

Let us evaluate $\mathbf{P}f$ for a couple of f. If $f(k) = k$, then

$$\mathbf{P}f(k) = \mathbb{E}[Y_1 + \cdots + Y_k] = k \cdot \mu \implies \mathbf{P}f = \mu f. \tag{14.16}$$

If $f(k) = \lambda^k$ for some $|\lambda| \leq 1$, then

$$\left(\mathbf{P}\lambda^{(\cdot)}\right)(k) = \mathbb{E}\left[\lambda^{Y_1 + \cdots + Y_k}\right] = \varphi(\lambda)^k. \tag{14.17}$$

Corollary 14.10. *The process $\{M_n := Z_n/\mu^n\}_{n=0}^{\infty}$ is a positive martingale and in particular*

$$\mathbb{E}Z_n = \mu^n < \infty \text{ for all } n \in \mathbb{N}_0. \tag{14.18}$$

Proof. If $f(n) = n$ for all n, then $\mathbf{P}f = \mu f$ by Eq. (14.16) and therefore

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = \mathbf{P}f(Z_n) = \mu \cdot f(Z_n) = \mu Z_n.$$

Dividing this equation by μ^{n+1} then shows $\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n$ as desired. As $M_0 = 1$ it then follows that $\mathbb{E}M_n = 1$ for all n and this gives Eq. (14.18). ■

Theorem 14.11. *If $\mu < 1$, then, almost surely, $Z_n = 0$ for a.a. n. In fact,*

$$\mathbb{E}[\textit{Total \# of organisms ever alive}] = \mathbb{E}\left[\sum_{n=0}^{\infty} Z_n\right] < \infty.$$

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 14.12. *If $\mu = 1$ and $\mathbb{P}(\xi_i^m = 1) < 1$, then again, almost surely, $Z_n = 0$ for a.a. n.*

Proof. First note the assumption $\mu = 1$ and the standing assumption that $p_1 < 1$ implies $p_0 > 0$. This then further implies that for any $k \geq 1$ we have

$$\mathbb{P}(Y_1 + \cdots + Y_k = 0) \geq \mathbb{P}(Y_1 = 0, \ldots, Y_k = 0) = p_0^k > 0$$

which then implies,

$$p(k,k) = \mathbb{P}(Y_1 + \cdots + Y_k = k) < 1. \tag{14.19}$$

Because of Corollary 14.10 and the assumption that $\mu = 1$, we know $\{Z_n\}_{n=1}^{\infty}$ is a martingale. This martingale being positive is L^1 bounded as

$$\mathbb{E}\left[|Z_n|\right] = \mathbb{E}Z_n = \mathbb{E}Z_0 = 1 \text{ for all } n.$$

Therefore the martingale convergence theorem guarantees that $Z_\infty = \lim_{n\to\infty} Z_n =: Z_\infty$ exists with $\mathbb{E}Z_\infty \le 1$. Because Z_n is integer valued, it must happen that $Z_n = Z_\infty$ a.a. If $k \in \mathbb{N}$, the event $\{Z_\infty = k\}$ can be expressed as

$$\{Z_\infty = k\} = \{Z_n = k \text{ a.a. } n\} = \cup_{M=1}^{\infty}\{Z_n = k \text{ for all } n \ge M\}.$$

As the sets in the union of this expression are increasing, it follows by the monotone (and then DCT) that

$$\begin{aligned}\mathbb{P}(Z_\infty = k) &= \lim_{M\to\infty}\mathbb{P}(Z_n = k \text{ for all } n \ge M)\\ &= \lim_{M\to\infty}\lim_{N\to\infty}\mathbb{P}(Z_n = k \text{ for } M \le n \le N)\\ &= \lim_{M\to\infty}\lim_{N\to\infty}\mathbb{P}(Z_M = k)\,p(k,k)^{N-M} = 0,\end{aligned}$$

wherein we have made use of Eq. (14.19) in order to evaluate the limit. Thus it follows that

$$\mathbb{P}(Z_\infty > 0) = \sum_{k=1}^{\infty}\mathbb{P}(Z_\infty = k) = 0.$$

The above argument does not apply to $k = 0$ since $p(0,0) = 1$ by definition. ■

Remark 14.13. By the way, the branching process, $\{Z_n\}_{n=0}^{\infty}$ with $\mu = 1$ and $\mathbb{P}(\xi = 1) < 1$ gives a nice example of a non regular martingale. Indeed, if Z were regular, we would have

$$Z_n = \mathbb{E}\left[\lim_{m\to\infty} Z_m|\mathcal{F}_n\right] = \mathbb{E}\left[0|\mathcal{F}_n\right] = 0$$

which is clearly false.

We now wish to consider the case where $\mu := \mathbb{E}Y_k = \mathbb{E}\left[\xi_i^m\right] > 1$. For $\lambda \in \mathbb{C}$ with $|\lambda| \le 1$ let

$$\varphi(\lambda) := \mathbb{E}\left[\lambda^{Y_1}\right] = \sum_{k\ge 0} p_k\lambda^k \tag{14.20}$$

be the moment generating function of $\{p_k\}_{k=0}^{\infty}$. Notice that $\varphi(1) = 1$ and for $\lambda = s \in (-1,1)$ we have

$$\varphi'(s) = \sum_{k\ge 0} kp_k s^{k-1} \text{ and } \varphi''(s) = \sum_{k\ge 0} k(k-1)\,p_k s^{k-2} \ge 0$$

with

$$\begin{aligned}\lim_{s\uparrow 1}\varphi'(s) &= \sum_{k\ge 0} kp_k = \mathbb{E}[\xi] =: \mu \text{ and}\\ \lim_{s\uparrow 1}\varphi''(s) &= \sum_{k\ge 0} k(k-1)\,p_k = \mathbb{E}[\xi(\xi-1)].\end{aligned}$$

Therefore φ is convex with $\varphi(0) = p_0$, $\varphi(1) = 1$ and $\varphi'(1) = \mu$.

Lemma 14.14. *If $\mu = \varphi'(1) > 1$, there exists a unique $\rho < 1$ so that $\varphi(\rho) = \rho$.*

Proof. See Figure 14.2 below.

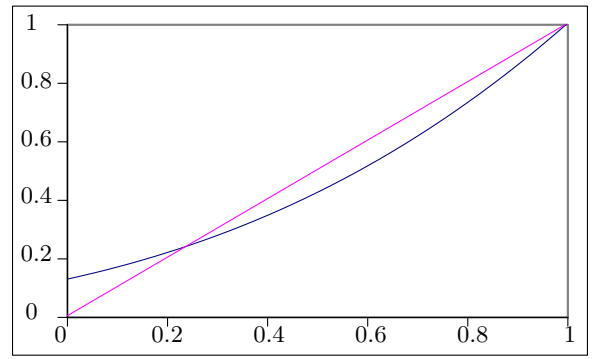


Fig. 14.2. Figure associated to $\varphi(s) = \frac{1}{8}\left(1 + 3s + 3s^2 + s^3\right)$ which is relevant for Exercise 3.13 of Durrett on p. 249. In this case $\rho \cong 0.236\,07$.

■

Theorem 14.15 (See Durrett [6], p. 247-248.). *If $\mu > 1$, then*

$$\mathbb{P}_1(\textit{Extinction}) = \mathbb{P}_1\left(\left\{\lim_{n\to\infty} Z_n = 0\right\}\right) = \mathbb{P}_1(\{Z_n = 0 \textit{ for some } n\}) = \rho.$$

Proof. Since $\{Z_m = 0\} \subset \{Z_{m+1} = 0\}$, it follows that $\{Z_m = 0\} \uparrow \{Z_n = 0 \text{ for some } n\}$ and therefore if $\theta_m := \mathbb{P}_1(Z_m = 0)$,then

$$\mathbb{P}_1(\{Z_n = 0 \text{ for some } n\}) = \lim_{m\to\infty}\theta_m.$$

Notice that $\theta_1 = \mathbb{P}_1(Z_1 = 0) = p_0 \in (0,1)$. We now show; $\theta_m = \varphi(\theta_{m-1})$. To see this, conditioned on the set $\{Z_1 = k\}$, $Z_m = 0$ iff all k – families die out in the remaining $m-1$ time units. Since each family evolves independently, the probability of this event is θ_{m-1}^k. [See Example 14.16 for a formal justification

of this fact.] Combining this with, $\mathbb{P}_1(\{Z_1 = k\}) = \mathbb{P}(\xi_1^1 = k) = p_k$, allows us to conclude by the first step analysis that

$$\begin{aligned}\theta_m = \mathbb{P}_1(Z_m = 0) &= \sum_{k=0}^{\infty} \mathbb{P}_1(Z_m = 0, Z_1 = k) \\ &= \sum_{k=0}^{\infty} \mathbb{P}_1(Z_m = 0 | Z_1 = k)\, \mathbb{P}_1(Z_1 = k) = \sum_{k=0}^{\infty} \theta_{m-1}^k p_k = \varphi(\theta_{m-1}).\end{aligned}$$

It is now easy to see that $\theta_m \uparrow \rho$ as $m \uparrow \infty$, again see Figure 14.3.

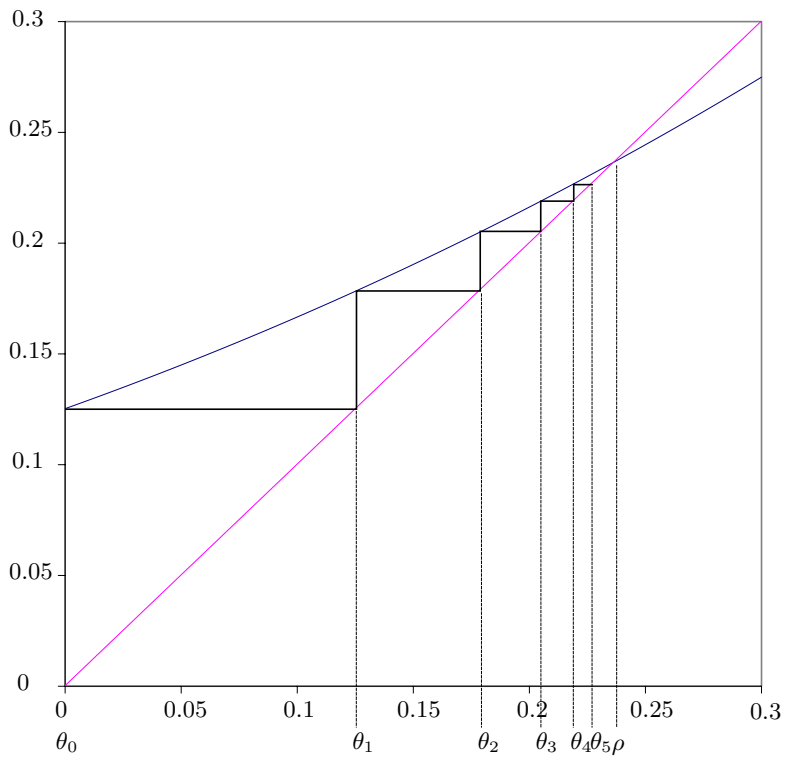


Fig. 14.3. The graphical interpretation of iterating $\theta_m = \varphi(\theta_{m-1})$ starting from $\theta_0 = \mathbb{P}_1(Z_0 = 0) = 0$ and then $\theta_1 = \varphi(0) = p_0 = \mathbb{P}_1(Z_1 = 0)$.

■

14.3.1 Appendix: justifying assumptions

Exercise 14.2 (Markov Chain Products). Suppose that $\mathbf{P}$ and $\mathbf{Q}$ are Markov matrices on state space S and T. Then $\mathbf{P} \otimes \mathbf{Q}$ defined by

$$\mathbf{P} \otimes \mathbf{Q}((s,t),(s',t')) := \mathbf{P}(s,s') \cdot \mathbf{Q}(t,t')$$

defines a Markov transition matrix on $S \times T$. Moreover, if $\{X_n\}$ and $\{Y_n\}$ are Markov chains on S and T respectively with transition matrices $\mathbf{P}$ and $\mathbf{Q}$ respectively, then $Z_n = (X_n, Y_n)$ is a Markov chain with transition matrix $\mathbf{P} \otimes \mathbf{Q}$.

Exercise 14.3 (Markov Chain Projections). Suppose that $\{X_n\}$ is the Markov chain on a state space S with transition matrix $\mathbf{P}$ and $\pi : S \to T$ is a surjective (for simplicity) map. Then there is a transition matrix $\mathbf{Q}$ on T such that $Y_n := \pi \circ X_n$ is a Markov chain on T for all starting distributions on S iff for all $t \in T$,

$$\mathbf{P}\left(x, \pi^{-1}(\{t\})\right) = \mathbf{P}\left(y, \pi^{-1}(\{t\})\right) \text{ whenever } \pi(x) = \pi(y). \tag{14.21}$$

In this case,

$$\mathbf{Q}(s,t) = \mathbf{P}\left(x, \pi^{-1}(t)\right) := \sum_{y \in S : \pi(y) = t} \mathbf{P}(x,y) \text{ where } x \in \pi^{-1}(s). \tag{14.22}$$

Example 14.16. Using the above results we may now justify the fact that "the different branches of the Galton-Watson tree evolve independently of one another." Indeed, suppose that $\{X_n\}$ and $\{Y_n\}$ are two independent copies of the Galton-Watson tree starting with k and l off-spring respectively. Let $\pi : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}_0$ be the addition map, $\pi(a,b) = a + b$ and let

$$Z_n := X_n + Y_n = \pi \circ (X_n, Y_n).$$

We claim that $\{Z_n\}$ is the Galton-Watson tree starting with $k + l$ off spring. To see this we notice that for $c \in \mathbb{N}_0$ (using the transition functions in Eq. (14.15) that

$$\begin{aligned}\sum_{(x,y) \in \pi^{-1}(c)} &\mathbf{P} \otimes \mathbf{P}((a,b),(x,y)) \\ &= \sum_{x+y=c} \mathbf{P}(a,x)\, \mathbf{P}(b,y) \\ &= \sum_{x+y=c} \left[\delta_{a,0}\delta_{0,x} + 1_{a \geq 1} \cdot \mathbb{P}(Y_1 + \cdots + Y_a = x)\right] \\ &\qquad \cdot \left[\delta_{b,0}\delta_{0,y} + 1_{b \geq 1} \cdot \mathbb{P}(Y_1 + \cdots + Y_b = y)\right]\end{aligned}$$

Hence if $\left\{\tilde{Y}_n\right\}$ is an independent copy of the $\{Y_n\}$ then assuming the $a, b \geq 1$ we find,

$$\begin{aligned}
&\sum_{x+y=c} \mathbb{P}\left(Y_1+\cdots+Y_a=x\right) \cdot \mathbb{P}\left(Y_1+\cdots+Y_b=y\right) \\
&\quad=\sum_{x+y=c} \mathbb{P}\left(Y_1+\cdots+Y_a=x, \tilde{Y}_1+\cdots+\tilde{Y}_b=y\right) \\
&\quad=\mathbb{P}\left(Y_1+\cdots+Y_a+\tilde{Y}_1+\cdots+\tilde{Y}_b=c\right) \\
&\quad=\mathbf{P}(a+b, c)=\mathbf{P}(\pi(a, b), c) .
\end{aligned}$$

When $a=0$, we get,

$$\begin{aligned}
&\sum_{(x,y)\in\pi^{-1}(c)} \mathbf{P} \otimes \mathbf{P}((a, b),(x, y)) \\
&\quad=\sum_{x+y=c} \delta_{0,x} \cdot\left[\delta_{b,0} \delta_{0,y}+1_{b \geq 1} \cdot \mathbb{P}\left(Y_1+\cdots+Y_b=y\right)\right] \\
&\quad=\left[\delta_{b,0} \delta_{0,c}+1_{b \geq 1} \cdot \mathbb{P}\left(Y_1+\cdots+Y_b=c\right)\right]=\mathbf{P}(a+b, c)
\end{aligned}$$

and similarly the statement holds if $b=0$. Thus we may now apply Exercise 14.3 to complete the proof.

Using the above results, if $\theta_m(k):=\mathbb{P}_k\left(Z_m=0\right)$, then

$$\begin{aligned}
\theta_m(k+l) &=\mathbb{P}_{k+l}\left(Z_m=0\right)=\mathbb{P}_{(k,l)}\left(X_m+Y_m=0\right) \\
&=\mathbb{P}_{(k,l)}\left(X_m=0, Y_m=0\right) \\
&=\mathbb{P}_k\left(X_m=0\right) \mathbb{P}_l\left(Y_m=0\right)=\theta_m(k) \theta_m(l)
\end{aligned}$$

and therefore, $\theta_m(k)=\theta_m(1)^k$.

Part IV

Continuous Time Theory

15

Discrete State Space/Continuous Time

In this chapter, we continue to assume that S is a finite or countable state space and $(\Omega, \mathbb{P})$ is a probability space.

15.1 Geometric, Exponential, and Poisson Distributions

Notation 15.1 *Let $N : \Omega \to \mathbb{N}_0$ be an integer valued random variable. We write:*

1. $N \overset{d}{=} Bern(p)$ *if* $\mathbb{P}(N = 1) = p$ *and* $\mathbb{P}(N = 0) = 1 - p$ *where* $0 \le p \le 1$ *and say N is a* ***Bernoulli random*** *variable.*
2. $N \overset{d}{=} Geo(p)$ *if* $\mathbb{P}(N = k) = p(1-p)^{k-1}$ *for* $k \in \mathbb{N}$ *and say N is a* ***geometric random*** *variable.*
3. $N \overset{d}{=} Pois(\lambda)$ *if* $\mathbb{P}(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ *and say N is a* ***Poisson random*** *variable.*

Remark 15.2. Here is how the geometric distribution often arises. If $\{X_k\}_{k=1}^{\infty}$ are i.i.d. Bern(p) – random variables (i.e. $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$), then

$$N := \inf\{k \ge 1 : X_k = 1\} \overset{d}{=} Geo(p).$$

Indeed,

$$\begin{aligned}\mathbb{P}(N = k) &= \mathbb{P}(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1) \\ &= (1-p)^{k-1} p.\end{aligned}$$

Here are two simple consequences of this representation.

Lemma 15.3 (Memoryless property of the geometric distribution). *If $N \overset{d}{=} Geo(p)$ is a geometric random variable then;*

1. $\mathbb{P}(N > n) = \mathbb{P}(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 0) = (1-p)^n$
2. *for all* $k, n \in \mathbb{N}$,

$$\mathbb{P}(N = n + k | N > n) = \mathbb{P}(N = k).$$

Proof. We take each item in turn.

1. Using the representation in Remark 15.2,

$$\mathbb{P}(N > n) = \mathbb{P}(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 0) = (1-p)^n.$$

Alternatively, we can sum the geometric series,

$$\mathbb{P}(N > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = (1-p)^n p \cdot \frac{1}{1-(1-p)} = (1-p)^n.$$

Conversely if $\mathbb{P}(N > n) = (1-p)^n$ for all $n \in \mathbb{N}$ then

$$\mathbb{P}(N = n) = \mathbb{P}(N > n-1) - \mathbb{P}(N > n) = (1-p)^{n-1} - (1-p)^n = p(1-p)^{n-1}.$$

2. Using the representation in Remark 15.2, we easily (and intuitively) see that

$$\begin{aligned}\mathbb{P}(N = n + k | N > n) &= \frac{\mathbb{P}(X_1 = 0, \ldots, X_{n+k-1} = 0, X_{n+k} = 1)}{\mathbb{P}(X_1 = 0, \ldots, X_n = 0)} \\ &= \mathbb{P}(X_{n+1} = 0, \ldots, X_{n+k-1} = 0, X_{n+k} = 1) \\ &= \mathbb{P}(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1) = \mathbb{P}(N = k).\end{aligned}$$

Alternatively by first principles,

$$\begin{aligned}\mathbb{P}(N = n + k | N > n) &= \frac{\mathbb{P}(N = n + k)}{\mathbb{P}(N > n)} = \frac{p(1-p)^{n+k-1}}{(1-p)^n} \\ &= p(1-p)^{k-1} = \mathbb{P}(N = k).\end{aligned}$$

■

Exercise 15.1 (Some Discrete Distributions). Let $p \in (0, 1]$ and $\lambda > 0$. In the two parts below, the distribution of N will be described. In each case find the generating function (see Proposition 1.2) and use this to verify the stated values for $\mathbb{E}N$ and $\mathrm{Var}(N)$.

1. Geometric(p) : $\mathbb{P}(N = k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$. ($\mathbb{P}(N = k)$ is the probability that the kth – trial is the first time of success out a sequence of independent trials with probability of success being p.) You should find $\mathbb{E}N = 1/p$ and $\mathrm{Var}(N) = \frac{1-p}{p^2}$.

2. Poisson(λ) : $\mathbb{P}(N=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ for all $k\in\mathbb{N}_0$. You should find $\mathbb{E}N=\lambda=\mathrm{Var}(N)$.

Definition 15.4. *A random variable $T\geq 0$ is said to be* ***exponential with parameter*** *$\lambda\in[0,\infty)$ provided, $\mathbb{P}(T>t)=e^{-\lambda t}$ for all $t\geq 0$. We will write $T\overset{d}{=}E(\lambda)$ for short.*

Alternatively we can express this distribution as

$$\begin{aligned}\mathbb{P}(T\in t,t+dt])&=\mathbb{P}(T>t)-\mathbb{P}(T>t+dt)=e^{-\lambda t}-e^{-\lambda(t+dt)}\\&=e^{-\lambda t}\left[1-e^{-\lambda dt}\right]=e^{-\lambda t}\lambda dt+o(dt)\ \text{ with } t\geq 0.\end{aligned}$$

To be more precise

$$\mathbb{E}[f(T)]=\int_0^\infty f(\tau)\lambda e^{-\lambda\tau}d\tau.$$

Notice that taking $f(\tau)=1_{\tau>t}$ in the above expression again gives,

$$\begin{aligned}\mathbb{P}(T>t)=\mathbb{E}[1_{T>t}]&=\int_0^\infty 1_{\tau>t}\lambda e^{-\lambda\tau}d\tau\\&=\int_t^\infty \lambda e^{-\lambda\tau}d\tau=e^{-\lambda t}.\end{aligned}$$

Notice that T has the following memoryless property,

$$\mathbb{P}(T>s+t|T>s)=\mathbb{P}(T>t)\ \text{ for all } s,t\geq 0.$$

Indeed,

$$\begin{aligned}\mathbb{P}(T>s+t|T>s)&=\frac{\mathbb{P}(T>s+t,T>s)}{\mathbb{P}(T>s)}\\&=\frac{\mathbb{P}(T>s+t)}{\mathbb{P}(T>s)}=\frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}\\&=e^{-\lambda t}=\mathbb{P}(T>t).\end{aligned}$$

See Theorem 15.8 below for the converse assertion.

Exercise 15.2. Let $T\overset{d}{=}E(\lambda)$ be as in Definition 15.4. Show;

1. $\mathbb{E}T=\lambda^{-1}$ and
2. $\mathrm{Var}(T)=\lambda^{-2}$.

In the next two propositions we are going $N_p\overset{d}{=}Geo(p)$ in the limit that we make p – small so there are many trials before success but at the same time we are going to suppose that time between trials is decreasing as well. Taking this limit with the appropriate scalings will limit to an exponential random variable.

Notation 15.5 *For $t\geq 0$ we let*

$$[t]:=\max\{n\in\mathbb{N}_0:n\leq t\}=\sum_{k=1}^\infty(k-1)1_{[k-1,k)}(t)$$

and more generally if $\varepsilon>0$ we let

$$[t]_\varepsilon:=\varepsilon\left[\frac{t}{\varepsilon}\right]=\sum_{k=1}^\infty 1_{[(k-1)\varepsilon,k\varepsilon)}(t).$$

Proposition 15.6 (Scaling limits of geometrics). *Let $\lambda>0$ be given and let $N_\varepsilon\overset{d}{=}Geo(\lambda\varepsilon)$ for $0<\varepsilon<\lambda^{-1}$. Then $\varepsilon N_\varepsilon\implies T\overset{d}{=}E(\lambda)$ as $\varepsilon\downarrow 0$, i.e.*

$$\lim_{\varepsilon\downarrow 0}\mathbb{P}(\varepsilon N_\varepsilon>t)=e^{-\lambda t}\ \text{ for all } t>0.$$

Proof. We have

$$\mathbb{P}(\varepsilon N_\varepsilon>t)=\mathbb{P}\left(N_\varepsilon>\frac{t}{\varepsilon}\right)=(1-\varepsilon\lambda)^{\left[\frac{t}{\varepsilon}\right]}.$$

Since $\left[\frac{t}{\varepsilon}\right]=\frac{t}{\varepsilon}+O(1)$ and $\ln(1-\varepsilon\lambda)=-\varepsilon\lambda+O(\varepsilon^2)$ we find

$$\ln\mathbb{P}(\varepsilon N_\varepsilon>t)=\left[\frac{t}{\varepsilon}\right]\ln(1-\varepsilon\lambda)=\left(\frac{t}{\varepsilon}+O(1)\right)\left(-\varepsilon\lambda+O(\varepsilon^2)\right)=-\lambda t+O(\varepsilon)$$

and the result follows. ■

We can substantially improve on the previous proposition.

Proposition 15.7. *For $p>0$ and $\lambda>0$ let $G_p:(0,1]\to\mathbb{N}$ be defined by*

$$G_p(x):=\sum_{n=1}^\infty n\cdot 1_{((1-p)^n,(1-p)^{n-1}]}(x)$$

and $T_\lambda:(0,1]\to\mathbb{R}_+$ be defined by $T_\lambda(x)=-\frac{1}{\lambda}\ln x$. If U is a $(0,1]$-valued uniformly distributed random variable, then;

1. *$G_p(U)\overset{d}{=}Geo(p)$,*
2. *$T_\lambda(U)\overset{d}{=}E(\lambda)$, and*
3. *$\lim_{\varepsilon\downarrow 0}\varepsilon G_{\varepsilon\lambda}(U)=T_\lambda(U)$.*

Proof. Since the law of U is Lebesgue measure (m) on $(0,1]$ and we find

$$\begin{aligned}\mathbb{P}(G_p(U)=n)&=m(x\in[0,1]:G_p(x)=n)\\&=(1-p)^{n-1}-(1-p)^n=p(1-p)^{n-1}\end{aligned}$$

which shows $G_p(U) \overset{d}{=} Geo(p)$. Similarly, as $T_\lambda(x) = -\frac{1}{\lambda}\ln x > t$ iff $x < e^{-\lambda t}$ it follows that

$$m(\{x \in (0,1] : T_\lambda(x) > t\}) = e^{-\lambda t},$$

i.e. $T_\lambda(U) \overset{d}{=} E(\lambda)$.

For item 3. we refer the reader to Figure 15.1. For the reader not convinced

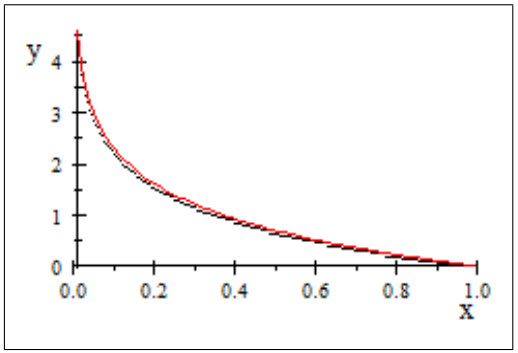


Fig. 15.1. This graph shows $\varepsilon G_\varepsilon$ for $\varepsilon = 1/20$ in black along with the graph of $-\ln x$ in red.

by the figure here are some more details. Since $G_p(x) = n$ iff

$$n\ln(1-p) < \ln x \le (n-1)\ln(1-p)$$

or equivalently (keep in mind $\ln(1-p) < 0$) iff

$$(n-1) \le \frac{\ln x}{\ln(1-p)} < n \iff \left[\frac{\ln x}{\ln(1-p)}\right] = n-1,$$

it follows that

$$G_p(x) = \left[\frac{\ln x}{\ln(1-p)}\right]_1 + 1.$$

Using $\varepsilon[t]_1 = [\varepsilon t]_\varepsilon$ we conclude

$$\varepsilon G_{\lambda\varepsilon}(x) = \varepsilon\left[\frac{1}{\ln(1-\lambda\varepsilon)}\ln x\right]_1 + \varepsilon = \left[\frac{\varepsilon}{\ln(1-\lambda\varepsilon)}\ln x\right]_\varepsilon + \varepsilon$$

from which it easily follows that

$$\lim_{\varepsilon\downarrow 0}\varepsilon G_{\lambda\varepsilon}(x) = -\frac{1}{\lambda}\ln x := T_\lambda(x)$$

This suffices to prove item 3. ■

Theorem 15.8 (Memoryless property of the exponential distribution). *A random variable, $T \in (0,\infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$\mathbb{P}(T > s+t | T > s) = \mathbb{P}(T > t) \text{ for all } s,t \ge 0,$$

where as usual, $\mathbb{P}(A|B) := \mathbb{P}(A\cap B)/\mathbb{P}(B)$ when $p(B) > 0$. (Note that $T \overset{d}{=} E(0)$ means that $\mathbb{P}(T > t) = e^{0t} = 1$ for all $t > 0$ and therefore that $T = \infty$ a.s.)

Proof. (The following proof is taken from [14].) Suppose first that $T \overset{d}{=} E(\lambda)$ for some $\lambda > 0$. Then

$$\mathbb{P}(T > s+t | T > s) = \frac{\mathbb{P}(T > s+t)}{\mathbb{P}(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(T > t).$$

For the converse, let $g(t) := \mathbb{P}(T > t)$, then by assumption,

$$\frac{g(t+s)}{g(s)} = \mathbb{P}(T > s+t | T > s) = \mathbb{P}(T > t) = g(t)$$

whenever $g(s) \ne 0$ 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

$$g(t+s) = g(t)g(s) \text{ for all } s,t \ge 0.$$

Since $T > 0$, we know that $g(1/n) = \mathbb{P}(T > 1/n) > 0$ for some n and therefore, $g(1) = g(1/n)^n > 0$ and we may write $g(1) = e^{-\lambda}$ for some $0 \le \lambda < \infty$.

Observe for $p,q \in \mathbb{N}$, $g(p/q) = g(1/q)^p$ and taking $p = q$ then shows, $e^{-\lambda} = g(1) = g(1/q)^q$. Therefore, $g(p/q) = e^{-\lambda p/q}$ so that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{Q}_+ := \mathbb{Q}\cap\mathbb{R}_+$. Given $r,s \in \mathbb{Q}_+$ and $t \in \mathbb{R}$ such that $r \le t \le s$ we have, since g is decreasing, that

$$e^{-\lambda r} = g(r) \ge g(t) \ge g(s) = e^{-\lambda s}.$$

Hence letting $s \uparrow t$ and $r \downarrow t$ in the above equations shows that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$ and therefore $T \overset{d}{=} E(\lambda)$. ■

15.2 Scaling Limits of Markov Chains

The next exercise deals with how to describe a "lazy" Markov chain. We will say a chain is lazy if $p(x,x) > 0$ for some $x \in S$. The point being that if $p(x,x) > 0$,

then the chain starting at x may be lazy and stay at x for some period of time before deciding to jump to a new site. The next exercise describes lazy chains in terms of a non-lazy chain and the random times that the lazy chain will spend lounging at each site $x \in S$. We will refer to this as the jump-hold description of the chain. We will give a similar description of chains on S in the context of continuous time in Corollary 15.42 below. Let $\{X_n\}_{n=0}^{\infty}$ be the Markov chain with transition kernel $\mathbf{P}$ and let $\mathbf{j}_1$ be the first jump time of the chain and $q(x,y)$ be defined in Exercise 15.3 below. We then have

$$\begin{aligned}\mathbb{P}_x(\mathbf{j}_1 = k, X_{\mathbf{j}_1} = y) &= \mathbb{P}_x(X_1 = x, \dots, X_{k-1} = x, X_k = y) \\ &= p(x,x)^{k-1} p(x,y) = p(x,x)^{k-1}(1 - p(x,x))\, q(x,y)\end{aligned}$$

from which we learn $\mathbf{j}_1$ and $Y_1 := X_{\mathbf{j}_1}$ are independent, $\mathbf{j}_1 \overset{d}{=} Geo((1 - p(x,x)))$, and $\mathbb{P}_x(Y_1 = y) = q(x,y)$. These computations along with induction and the strong Markov property leads to the results in the next exercise. [You are asked to give an elementary proof of these results.]

Notation 15.9 (Jump and hold times) *Let $\mathbf{j}_k$ denote the time of the kth – jump of the chain $\{X_n\}_{n=0}^{\infty}$ so that*

$$\begin{aligned}\mathbf{j}_1 &:= \inf\{n > 0 : X_n \neq X_0\} \text{ and} \\ \mathbf{j}_{k+1} &:= \inf\{n > \mathbf{j}_k : X_n \neq X_{\mathbf{j}_k}\}\end{aligned}$$

with the convention that $\mathbf{j}_0 = 0$. Further let $\sigma_k := \mathbf{j}_{k+1} - \mathbf{j}_k$ denote the time spent between the kth and $(k+1)^{th}$ jump of the chain $\{X_n\}_{n=0}^{\infty}$, see Figure 15.2.

The sample path of the chain $\{X_n\}$ associated to Figure 15.2 is

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_n	2	2	4	4	4	1	3	3	3	2	2	2	1	1	1	...

and correspondingly we have $\sigma_0 = 2$, $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 3$, $\sigma_4 = 3$, and $\sigma_5 > 2$ and similarly $j_1 = 2$, $j_2 = 5$, $j_3 = 6$, $j_4 = 9$, $j_5 = 12$, and $j_6 > 14$.

Exercise 15.3 (Discrete time M.C. jump-hold description). Let S be a countable or finite set $(\Omega, \mathbb{P}, \{X_n\}_{n=0}^{\infty})$ be a Markov chain with transition kernel, $\mathbf{P} := \{p(x,y)\}_{x,y \in S}$ and let $\nu(x) := \mathbb{P}(X_0 = x)$ for all $x \in S$. For simplicity let us assume there are no **absorbing states**,[1] (i.e. $p(x,x) < 1$ for all $x \in S$) and then define $\mathbf{Q}_{x,y} = q(x,y)$ where

$$q(x,y) := \begin{cases} \frac{p(x,y)}{1-p(x,x)} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}.$$

[1] A state x is absorbing if $p(x,x) = 1$ since in this case there is no chance for the chain to leave x once it hits x.

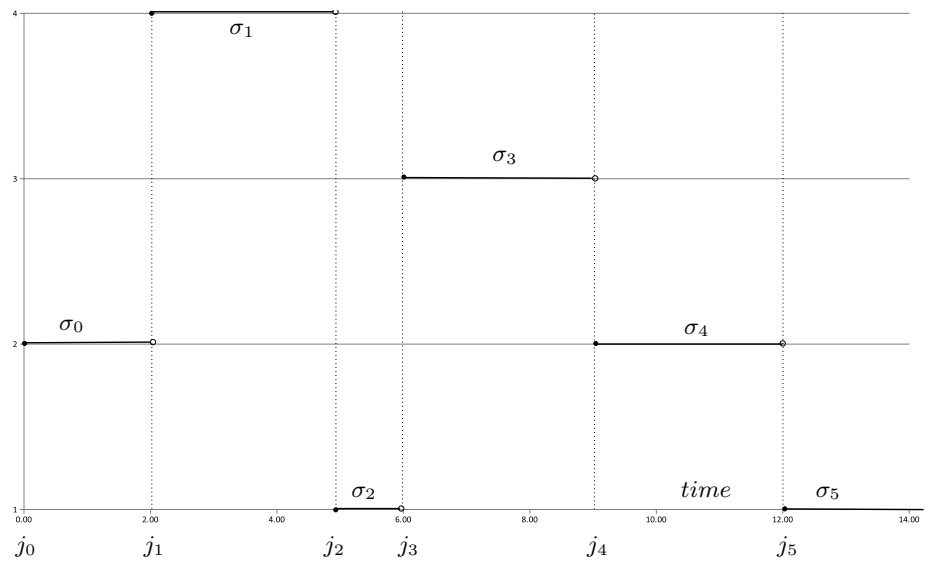


Fig. 15.2. A typical sample path for a lazy discrete time Markov chain with $S = \{1,2,3,4\}$. The picture indicates the jump and sojourn times for this random graph.

Let $\mathbf{j}_k$ denote the time of the kth – jump of the chain $\{X_n\}_{n=0}^{\infty}$ so that

$$\begin{aligned}\mathbf{j}_1 &:= \inf\{n > 0 : X_n \neq X_0\} \text{ and} \\ \mathbf{j}_{k+1} &:= \inf\{n > \mathbf{j}_k : X_n \neq X_{\mathbf{j}_k}\}\end{aligned}$$

with the convention that $\mathbf{j}_0 = 0$. Further let $\sigma_k := \mathbf{j}_{k+1} - \mathbf{j}_k$ denote the time spent between the kth and $(k+1)^{\text{th}}$ jump of the chain $\{X_n\}_{n=0}^{\infty}$, see Figure 15.2. Show;

1. For $\{x_k\}_{k=0}^{n} \subset S$ with $x_k \neq x_{k-1}$ for $k = 1, \dots, n$ and $m_0, \dots, m_{n-1} \in \mathbb{N}$, show

$$\begin{aligned}&\mathbb{P}\left(\left[\cap_{k=0}^{n}\{X_{\mathbf{j}_k} = x_k\}\right] \cap \left[\cap_{k=0}^{n-1}\{\sigma_k = m_k\}\right]\right) \\ &= \nu(x_0) \prod_{k=0}^{n-1} p(x_k, x_k)^{m_k - 1}(1 - p(x_k, x_k)) \cdot q(x_k, x_{k+1}). \end{aligned} \tag{15.1}$$

2. Summing the previous formula on $m_0, \dots, m_{n-1} \in \mathbb{N}$, conclude

$$\mathbb{P}\left(\left[\cap_{k=0}^{n}\{X_{\mathbf{j}_k} = x_k\}\right]\right) = \nu(x_0) \cdot \prod_{k=0}^{n-1} q(x_k, x_{k+1}),$$

i.e. this shows $\{Y_k := X_{\mathbf{j}_k}\}_{k=0}^{\infty}$ is a Markov chain with transition kernel, $\mathbf{Q}_{x,y} = q(x,y)$.

3. Conclude, relative to the conditional probability measure, $\mathbb{P}\left(\cdot \middle| \left[\cap_{k=0}^{n}\{X_{\mathbf{j}_k}=x_k\}\right]\right)$, that $\{\sigma_k\}_{k=0}^{n-1}$ are independent geometric $\sigma_k \overset{d}{=} Geo\left(1-p\left(x_k,x_k\right)\right)$ for $0\le k\le n-1$.

Corollary 15.10. *Let $\mathbf{P}$ and $\mathbf{Q}$ be the Markov matrices as described in Exercise 15.3, $\{Y_n\}_{n=0}^{\infty}$ be the non-lazy Markov chain associated to $\mathbf{Q}$, and let $\{U_n\}_{n=0}^{\infty}$ be i.i.d. random variables uniformly distributed on $(0,1]$. Let $\left\{\hat{X}_n\right\}_{n\ge 0}$ be the stochastic process which visits the sites (in order) determined by $\{Y_n\}_{n=0}^{\infty}$ and remains at site Y_n for $\sigma_n := G_{(1-p(Y_n,Y_n))}\left(U_n\right)$ many steps. Then $\left\{\hat{X}_n\right\}_{n\ge 0}$ and $\{X_n\}_{n\ge 0}$ have the same distribution assuming we start each chain with the same initial distributions.*

The reason that geometric distribution appears in the above description is because it has the forgetful property described in Lemma 15.3 above. This forgetfulness property of the jump time intervals is a consequence of the memoryless property of a Markov chain. For continuous time chains we will see that we essentially need only replace the geometric distribution in the above description by the exponential distribution (see Definition 15.4) which is the unique continuous distribution with the same forgetfulness property, see Theorem 15.8 above. Theorem 15.14, we are going to start with a non-lazy Markov chain $\{Y_n\}_{n=0}^{\infty}$ and make it lazier and lazier while at the same time scaling the amount of times between the steps of the chain to be smaller and smaller. By choosing the appropriate scalings will get a limiting process, $\{X_t\}$, which is going to be a typical continuous time Markov chain. Theorem 15.14 will in fact give most of the key results in this chapter. We begin with some basic facts about matrix exponentials which will be needed in the proof of Theorem 15.14 and for the rest of this chapter as well.

Definition 15.11. *If $\{\mathbf{B}_{x,y}\}_{x,y\in S}$ is a matrix on S with $K := \sup_{x\in S}\sum_{y\in S}|\mathbf{B}_{x,y}|<\infty$, then we define $e^{\mathbf{B}}$ by the convergent sum,*

$$e^{\mathbf{B}}=\sum_{n=0}^{\infty}\frac{1}{n!}\mathbf{B}^n.$$

Remark 15.12. We will take without proof that the exponentiation function satisfies the following properties;

1. $\frac{d}{dt}e^{t\mathbf{B}}=\mathbf{B}e^{t\mathbf{B}}=e^{t\mathbf{B}}\mathbf{B}$ with $e^{0\mathbf{B}}=\mathbf{I}$ and this differential equation uniquely determines $e^{t\mathbf{B}}$.
2. $e^{t\mathbf{B}}e^{s\mathbf{B}}=e^{(t+s)\mathbf{B}}$ for all $s,t\in\mathbb{R}$ and more generally $e^{\mathbf{B}}e^{\mathbf{C}}=e^{\mathbf{B}+\mathbf{C}}$ provided $\mathbf{BC}=\mathbf{CB}$.
3. If $K<1$, then

$$\ln\left(\mathbf{I}+\mathbf{B}\right)=\sum_{k=1}^{\infty}\left(-1\right)^{k+1}\frac{1}{k}\mathbf{B}^k \text{ exists}$$

and

$$e^{\ln(\mathbf{I}+\mathbf{B})}=\mathbf{I}+\mathbf{B}.$$

Lemma 15.13 (Exponential product formula). *If $\mathbf{B}$ is as in Definition 15.11 and $t\ge 0$, then*

$$\lim_{\varepsilon\downarrow 0}\left(\mathbf{I}+\varepsilon\mathbf{B}\right)^{\left[\frac{t}{\varepsilon}\right]}=e^{t\mathbf{B}}. \tag{15.2}$$

[This result is also easily proved for $t<0$ as well.]

Proof. If $t=0$ both sides of Eq. (15.2) are equal to the identity and so there is nothing to prove and we may and do assume $t>0$. From item 3. of Remark 15.12, for ε sufficiently small we have

$$\ln\left(\mathbf{I}+\varepsilon\mathbf{B}\right)=\varepsilon\mathbf{B}+O\left(\varepsilon^2\right) \text{ and}$$
$$\mathbf{I}+\varepsilon\mathbf{B}=e^{\ln(\mathbf{I}+\varepsilon\mathbf{B})}=e^{\varepsilon\mathbf{B}+O\left(\varepsilon^2\right)}.$$

Raising this equation to $\left[\frac{t}{\varepsilon}\right]^{\text{th}}$ power gives

$$\left(\mathbf{I}+\varepsilon\mathbf{B}\right)^{\left[\frac{t}{\varepsilon}\right]}=e^{\left[\frac{t}{\varepsilon}\right]\varepsilon\mathbf{B}+\left[\frac{t}{\varepsilon}\right]O\left(\varepsilon^2\right)}=e^{t\mathbf{B}+O(\varepsilon)}\to e^{t\mathbf{B}} \text{ as } \varepsilon\downarrow 0,$$

wherein we have used,

$$\left[\frac{t}{\varepsilon}\right]\varepsilon=\left(\frac{t}{\varepsilon}+O\left(1\right)\right)\varepsilon=t+O\left(\varepsilon\right)$$

for the last equality. ■

Theorem 15.14 (Continuous time MC as a scaling limit). *Let $M<\infty$ and $a:S\to(0,M)$ be a given function, $\mathbf{D}_{x,y}:=a_x\delta_{xy}$, and $\mathbf{Q}$ be a Markov matrix such that $\mathbf{Q}_{xx}=0$ for all $x\in S$ so that $\mathbf{Q}$ generates a non-lazy chain $\{Y_n\}_{n=0}^{\infty}$. For $\varepsilon>0$ let*

$$\mathbf{P}^{(\varepsilon)}:=\left(I-\varepsilon\mathbf{D}\right)+\varepsilon\mathbf{DQ} \text{ i.e. } \mathbf{P}_{xy}^{(\varepsilon)}=\left(1-\varepsilon a_x\right)\delta_{xy}+\varepsilon a_x\mathbf{Q}_{xy}.$$

Let $\left\{X_t^{(\varepsilon)}=\hat{X}_{\left[\frac{t}{\varepsilon}\right]}^{(\varepsilon)}\right\}_{t\ge 0}$ where $\left\{\hat{X}_n^{(\varepsilon)}\right\}_{n=0}^{\infty}$ is the Markov chain generated by $\mathbf{P}^{(\varepsilon)}$.[2] Then

[2] For $\varepsilon>0$ small, the lazy chain, $\left\{\hat{X}_n^{(\varepsilon)}\right\}_{n=0}^{\infty}$, is in fact very lazy and only leaves a site $x\in S$ at any given step with probability $\varepsilon\lambda_x << 1$. On the other hand,

1. *$X^{(\varepsilon)} \implies X$ where X is a "continuous time" Markov process with "Markov transition semi-group,"*

$$\mathbf{P}_t = e^{t\mathbf{A}} \text{ where } \mathbf{A} := -\mathbf{D} + \mathbf{DQ}.$$

[See the proof for the meaning of the assertion that $X^{(\varepsilon)} \implies X$.]

2. *$\mathbb{E}_x [f(X_t)|\mathcal{F}_s] = (\mathbf{P}_{t-s} f)(X_s)$ for all $t > s$ and $f : S \to \mathbb{R}$ bounded or positive, where f is viewed as a column vector, i.e.*

$$(\mathbf{P}_t f)(x) := \sum_{y \in S} \mathbf{P}_t(x,y) f(y) \ \ \forall\, x \in S.$$

3. *We can construct the process $X^{(\varepsilon)}$ is such a way that $\lim_{\varepsilon \downarrow 0} X_t^{(\varepsilon)} = X_t$, $\mathbb{P}_x$ – a.s. for all $x \in S$.*
4. *The trajectories of the X process may be described as follows;*
   a) *the jump locations are determined by the non-lazy chain $\{Y_n\}$.*
   b) *Given $\{Y_0 = x_0, \dots, Y_n = x_n\}$ with $x_j \in S$, let $\{S_j\}_{j=0}^n$ be independent exponential random variables with $S_j \overset{d}{=} E(a_{x_j})$. The trajectory of X for $0 \le t \le \sum_{j=0}^n S_j$ is determined by holding the chain at x_j for time S_j for $0 \le j \le n$, see Figure 15.5 below.*

Proof. Since $\mathbf{P}^{(\varepsilon)} \to I$ as $\varepsilon \downarrow 0$, $\left[\mathbf{P}^{(\varepsilon)}\right]^{n+1} \cong \left[\mathbf{P}^{(\varepsilon)}\right]^n$ when ε is small. Moreover by Lemma 15.13[3] we find for any $t > 0$ that

$$\lim_{\varepsilon \downarrow 0} \left[\mathbf{P}^{(\varepsilon)}\right]^{\left[\frac{t}{\varepsilon}\right]} = \lim_{\varepsilon \downarrow 0} \left[I + \varepsilon(-\mathbf{D} + \mathbf{DQ})\right]^{\left[\frac{t}{\varepsilon}\right]} = e^{t(-\mathbf{D}+\mathbf{DQ})} = \mathbf{P}_t.$$

We now go to the proof using these limiting results below without further comment.

1. Let $0 = t_0 < t_1 < t_2 < \cdots < t_m < \infty$, then

$$\begin{aligned}\mathbb{P}_x\left(X_{t_1}^{(\varepsilon)} = x_1, \dots, X_{t_n}^{(\varepsilon)} = x_n\right) &= \mathbb{P}_x\left(\hat{X}_{[t_1/\varepsilon]}^{(\varepsilon)} = x_1, \dots, \hat{X}_{[t_n/\varepsilon]}^{(\varepsilon)} = x_n\right) \\ &\cong \left[\mathbf{P}^{(\varepsilon)}\right]_{x,x_1}^{[t_1/\varepsilon]} \cdots \left[\mathbf{P}^{(\varepsilon)}\right]_{x,x_1}^{[t_n - t_{n-1}/\varepsilon]}\end{aligned}$$

and hence

$$\begin{aligned}\lim_{\varepsilon \downarrow 0} &\mathbb{P}_x\left(X_{t_1}^{(\varepsilon)} = x_1, \dots, X_{t_n}^{(\varepsilon)} = x_n\right) \\ &= \mathbf{P}_{t_1}(x, x_1)\, \mathbf{P}_{t_2 - t_1}(x_1, x_2) \dots \mathbf{P}_{t_n - t_{n-1}}(x_{n-1}, x_n).\end{aligned}$$

The latter expression is by definition, $\mathbb{P}_x(X_{t_1} = x_1, \dots, X_{t_n} = x_n)$.

2. The Markov property for X_t follows rather directly from the formula in the proof of item 1. above. Intuitively, for $\varepsilon > 0$ we have by the Markov property for $\hat{X}^{(\varepsilon)}$ that

$$\mathbb{E}_x\left[f\left(X_t^{(\varepsilon)}\right)|\mathcal{F}_s^{(\varepsilon)}\right] \cong \left(\left[\mathbf{P}^{(\varepsilon)}\right]^{\left[\frac{t-s}{\varepsilon}\right]} f\right)\left(X_s^{(\varepsilon)}\right). \tag{15.3}$$

Passing the limit in Eq. (15.3) "shows"

$$\mathbb{E}_x[f(X_t)|\mathcal{F}_s] = (\mathbf{P}_{t-s} f)(X_s).$$

3. Let us now construct $X^{(\varepsilon)}$ so that $\lim_{\varepsilon \downarrow 0} X^{(\varepsilon)}$ is a.s. convergent. We begin by letting $\{Y_n\}_{n=0}^\infty$ be the non-lazy chain associated to $\mathbf{Q}$ and then let $\{U_n\}_{n=0}^\infty$ be i.i.d. random variables uniformly distributed on $(0,1]$ which are independent of the chain $\{Y_n\}_{n=0}^\infty$. Then let $S_n^\varepsilon := \varepsilon \sigma_n^\varepsilon = \varepsilon G_{\varepsilon a_{Y_n}}(U_n)$ be the nth - holding time, $J_0^{(\varepsilon)} = 0$, and for $k \ge 1$, $J_k^{(\varepsilon)} := \sum_{0 \le n < k} S_n^\varepsilon$ be the time of the nth – jump and therefore set[4]

$$X_t^{(\varepsilon)} = Y_n \text{ when } J_n^{(\varepsilon)} \le t < J_{n+1}^{(\varepsilon)}.$$

From Exercise 15.3 we may conclude that $\left\{X_t^{(\varepsilon)}\right\}_{t \ge 0}$ has the same distribution as the chain we described in the beginning of the statement of the Theorem. Making use of Proposition 15.7 we will have $\lim_{\varepsilon \downarrow 0} X_t^{(\varepsilon)} = X_t$ a.s. where now,

$$S_n := -\frac{1}{a_{Y_n}} \ln U_n, \ J_n = \sum_{0 \le k < n} S_k, \text{ and} \tag{15.4}$$

$$X_t = Y_n \text{ when } J_n \le t < J_{n+1}. \tag{15.5}$$

4. Item 4. now follows directly from what we have done in item 3. Indeed, given $\{Y_0 = x_0, \dots, Y_n = x_n\}$ with $x_j \in S$ we have that the holding times, $\left\{S_k = -\frac{1}{a_{x_k}} \ln U_k\right\}_{k=0}^n$, are independent random times with $S_k \overset{d}{=} E(a_{x_k})$. Moreover, by Theorem 15.22 below, $\lim_{n \to \infty} J_n = \infty$ a.s. and therefore Eq. (15.5) makes sense almost surely.

■

[3] The boundeness of the rates $\{\lambda_x\}_{x \in S}$ implies $\sup_{x \in S} \sum_{y \in S} |\mathbf{A}_{xy}| < \infty$.

[4] Since $J_n^{(\varepsilon)} = \varepsilon \mathbf{j}_n^{(\varepsilon)}$ we have $J_n^{(\varepsilon)} \le t < J_{n+1}^{(\varepsilon)}$ iff $\varepsilon \mathbf{j}_n^{(\varepsilon)} \le t < \varepsilon \mathbf{j}_{n+1}^{(\varepsilon)}$ iff $\mathbf{j}_n^{(\varepsilon)} \le \frac{t}{\varepsilon} < \mathbf{j}_{n+1}^{(\varepsilon)}$ iff $\mathbf{j}_n^{(\varepsilon)} \le \left[\frac{t}{\varepsilon}\right] < \mathbf{j}_{n+1}^{(\varepsilon)}$ and therefore $X_t^{(\varepsilon)} = \hat{X}_{\left[\frac{t}{\varepsilon}\right]}^{(\varepsilon)}$.

Notation 15.15 *We refer Eqs. (15.4) and (15.5) as the* ***jump hold*** *of construction of the process* $\{X_t\}_{t\geq 0}$.

Definition 15.16 (Infinitesimal Generator). *Let* S *be a finite or countable set and* $\mathbf{A} = \{\mathbf{A}_{x,y} \in \mathbb{R}\}_{x,y\in S}$ *be a matrix indexed by* S. *We say* $\mathbf{A}$ *is an* ***infinitesimal Markov generator*** *if* $\mathbf{A}_{x,y} \geq 0$ *for all* $x \neq y$ *and* $\sum_{y\in S} \mathbf{A}_{x,y} = 0$ *for all* $x \in S$.

Example 15.17. Let $\mathbf{Q}$ and $\mathbf{D}$ be as in Theorem 15.14, then $\mathbf{A} := -\mathbf{D} + \mathbf{DQ}$, satisfies,

$$\begin{aligned} \mathbf{A}_{x,y} &= a_x \mathbf{Q}_{xy} \geq 0 \text{ if } x \neq y, \\ \mathbf{A}_{x,x} &= -a_x < 0, \text{ and} \\ \sum_{y\in S} \mathbf{A}_{xy} &= 0, \end{aligned}$$

i.e. $\mathbf{A}$ is an infinitesimal generator. Conversely if we are given a matrix $\mathbf{A}$ with the above properties then we can define $a_x := -\mathbf{A}_{x,x}$ and

$$\mathbf{Q}_{xy} := 1_{x\neq y} \cdot \frac{1}{a_x} \mathbf{A}_{x,y}.$$

We will then have $\mathbf{Q}$ is the Markov matrix of a non-lazy chain such that $\mathbf{A} := -\mathbf{D} + \mathbf{DQ}$ where $\mathbf{D} = \operatorname{diag}\left(\{a_x\}_{x\in S}\right)$.

More generally if $a_x = 0$ for some x we make the following definition.

Definition 15.18. *Given a Markov-semi-group infinitesimal generator,* $\mathbf{A}$, *on* S, *let* $a_x := -\mathbf{A}_{x,x}$ *for* $x \in S$ *and* $\mathbf{Q} = \tilde{\mathbf{A}}$ *be the Markov matrix indexed by* S *defined by*

$$\mathbf{Q}_{x,y} := 1_{a_x\neq 0} \cdot 1_{x\neq y} \frac{\mathbf{A}_{x,y}}{a_x} + 1_{a_{x=0}} \cdot 1_{x=y}.$$

Example 15.19. If

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left[\begin{array}{ccc} -3 & 1 & 2 \\ 0 & 0 & 0 \\ 2 & 7 & -9 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

then

$$\mathbf{D} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left[\begin{array}{ccc} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 9 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \text{ and } \mathbf{Q} = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left[\begin{array}{ccc} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 1 & 0 \\ \frac{2}{9} & \frac{7}{9} & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}.$$

Let us summarize the key points of the above results in the following theorem.

Theorem 15.20. *Let* S *be a finite or countable set,* $\mathbf{A}$ *be an infinitesimal Markov generator on* S, $\{a_x\}_{x\in S}$, *and* $\mathbf{Q}$ *be defined as in Definition 15.18 and assume* $\sup_{x\in S} a_x < \infty$ *so that*

$$\mathbf{P}_t := e^{t\mathbf{A}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{A}^n \text{ exists.}$$

Further let ;

1. $\{Y_n\}_{n=0}^{\infty}$ *be the non-lazy Markov chain associated to* $\mathbf{Q}$.
2. $\{U_n\}_{n=0}^{\infty}$ *be i.i.d. random variables uniformly distributed on* $(0,1]$ *which are independent of the chain* $\{Y_n\}$.
3. *Set* $S_n := -\frac{1}{a_{Y_n}} \ln U_n \overset{d}{=} E\left(a_{Y_n}\right)$ *where* $S_n \equiv \infty$ *if* $a_{Y_n} = 0$ *and then define* $J_0 = 0$ *and* $J_n := S_0 + S_1 + \cdots + S_{n-1}$ *for* $n \geq 1$.
4. *Finally let*
$$X_t := \sum_{n=0}^{\infty} Y_n 1_{J_n \leq t < J_{n+1}} \in S \text{ for } t \in \mathbb{R}_+. \tag{15.6}$$

Then for all $x_0, x_1, \ldots, x_n \in S$ *and* $0 < t_1 < t_2 < \cdots < t_n$ *(see Figure 15.6 below),*

$$\mathbb{P}_{x_0}\left(X_{t_1} = x_1, \ldots X_{t_n} = x_n\right) = \mathbf{P}_{t_1}\left(x, x_1\right) \mathbf{P}_{t_2 - t_1}\left(x_1, x_2\right) \ldots \mathbf{P}_{t_n - t_{n-1}}\left(x_{n-1}, x_n\right).$$

Remark 15.21. The expression in Eq. (15.6) is short hand for writing; $X_t = Y_n$ where n is the unique positive integer such that $t \in [J_n, J_{n+1})$. If $a_{Y_{n-1}} > 0$ while $a_{Y_n} = 0$, then $J_n < \infty$ while $J_{n+1} = \infty$ and we will have $X_t = Y_n$ for all $t \geq J_n$.

15.3 Sums of Exponentials

From Theorem 15.14, we see that the trajectories of the Markov chain described there were defined for $0 \leq t < \sum_{j=0}^{\infty} S_j$ and so we could run into trouble if $\sum_{j=0}^{\infty} S_j < \infty$ with positive probability. As we will see in the next theorem, we avoided this problem by assuming the jump rates $\{a_x\}_{x\in S}$ of our chain were bounded.

Theorem 15.22. *Let* $\{T_j\}_{j=1}^{\infty}$ *be independent random variables such that* $T_j \overset{d}{=} E\left(\lambda_j\right)$ *with* $0 < \lambda_j < \infty$ *for all* j. *Then:*

1. *If* $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ *then* $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 0$ *(i.e.* $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n < \infty\right) = 1$*).*
2. *If* $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ *then* $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 1$.

In summary, there are precisely two possibilities; 1) $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 0$ *or 2)* $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 1$ *and moreover*

$$\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = \begin{cases} 0 \text{ iff } \mathbb{E}\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty. \\ 1 \text{ iff } \mathbb{E}\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} \lambda_n^{-1} = \infty \end{cases}$$

Proof. 1. Since

$$\mathbb{E}\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[T_n\right] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} T_n < \infty$ a.s., i.e. $\mathbb{P}\left(\sum_{n=1}^{\infty} T_n = \infty\right) = 0$.

2. First observe that for $\alpha < \lambda$ if $T \overset{d}{=} \exp(\lambda)$, then

$$\mathbb{E}\left[e^{\alpha T}\right] = \int_0^{\infty} e^{\alpha\tau} \lambda e^{-\lambda\tau} d\tau = \frac{\lambda}{\lambda - \alpha} = \frac{1}{1 - \alpha\lambda^{-1}}. \tag{15.7}$$

By the DCT, independence, and Eq. (15.7) with $\alpha = -1$ we find,

$$\begin{aligned} \mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] &= \lim_{N\to\infty} \mathbb{E}\left[e^{-\sum_{n=1}^{N} T_n}\right] = \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{-T_n}\right] \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \left(\frac{1}{1+\lambda_n^{-1}}\right) = \prod_{n=1}^{\infty} (1 - a_n) \end{aligned}$$

where

$$a_n = 1 - \frac{1}{1+\lambda_n^{-1}} = \frac{1}{1+\lambda_n}.$$

Hence by Exercise 15.4 below, $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $\infty = \sum_{n=1}^{\infty} a_n$ which happens iff $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ as you should verify. This completes the proof since $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $e^{-\sum_{n=1}^{\infty} T_n} = 0$ a.s. or equivalently $\sum_{n=1}^{\infty} T_n = \infty$ a.s. ■

Lemma 15.23. *If* $0 \le x \le \frac{1}{2}$, *then*

$$e^{-2x} \le 1 - x \le e^{-x}. \tag{15.8}$$

Moreover, the upper bound in Eq. (15.8) is valid for all $x \in \mathbb{R}$.

Proof. The upper bound follows by the convexity of e^{-x}, see Figure 15.3. For the lower bound we use the convexity of $\varphi(x) = e^{-2x}$ to conclude that the line joining $(0,1) = (0, \varphi(0))$ and $\left(1/2, e^{-1}\right) = (1/2, \varphi(1/2))$ lies above $\varphi(x)$

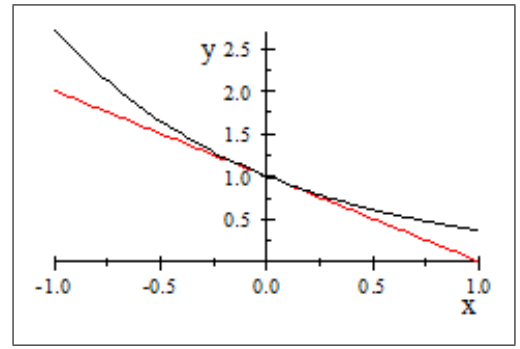


Fig. 15.3. A graph of $1 - x$ and e^{-x} showing that $1 - x \le e^{-x}$ for all x.

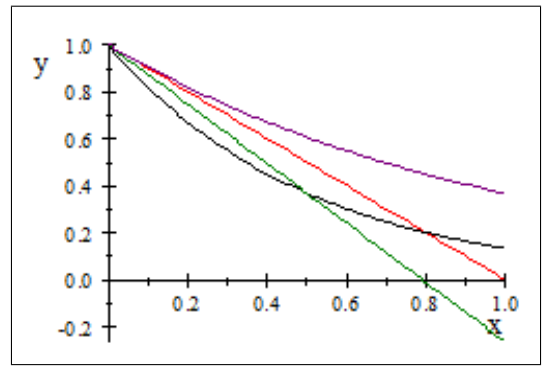


Fig. 15.4. A graph of $1-x$ (in red), the line joining $(0,1)$ and $\left(1/2, e^{-1}\right)$ (in green), e^{-x} (in purple), and e^{-2x} (in black) showing that $e^{-2x} \le 1 - x \le e^{-x}$ for all $x \in [0, 1/2]$.

for $0 \le x \le 1/2$. Then we use the fact that the line $1 - x$ lies above this line to conclude the lower bound in Eq. (15.8), see Figure 15.4. ■

For $\{a_n\}_{n=1}^{\infty} \subset [0,1]$, let

$$\prod_{n=1}^{\infty} (1 - a_n) := \lim_{N\to\infty} \prod_{n=1}^{N} (1 - a_n).$$

The limit exists since, $\prod_{n=1}^{N} (1 - a_n)$ decreases as N increases.

Exercise 15.4. Show; if $\{a_n\}_{n=1}^{\infty} \subset [0,1)$, then

$$\prod_{n=1}^{\infty} (1 - a_n) = 0 \iff \sum_{n=1}^{\infty} a_n = \infty.$$

The implication, $\Longleftarrow$, holds even if $a_n = 1$ is allowed.

15.4 Continuous Time Markov Chains (formalities)

Let S be a countable or finite state space but now let us replace discrete time, $\mathbb{N}_0$, by its continuous cousin, $\mathbf{T} := [0, \infty)$. We now formalize the notions of a continuous time homogeneous Markov chain. In more detail we will assume that $\{X_t\}_{t \geq 0}$ is a stochastic process whose sample paths are right continuous, see Figure 15.5. (These processes need not have left hand limits if there are an infinite number of jumps in a finite time interval. For the most part we will assume that this does not happen almost surely.) The fact that time is continuous adds some mathematical technicalities which typically will be suppressed in these notes unless they are too dangerous to do so.

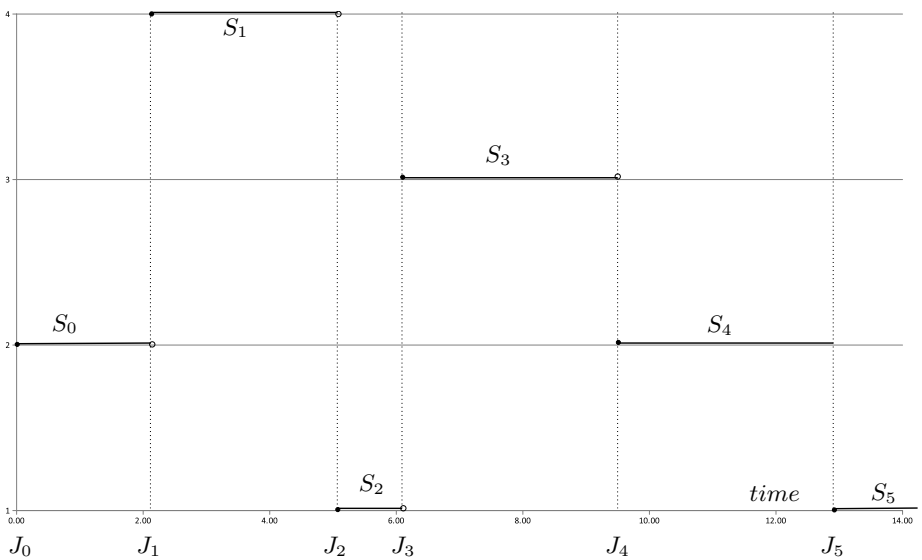


Fig. 15.5. A typical sample path for a continuous time Markov chain with $S = \{1, 2, 3, 4\}$. The picture indicates the jump and sojourn times for this random graph.

Notation 15.24 (Jump and Sojourn times) *Let $J_0 = 0$ and then define J_n inductively by*

$$J_{n+1} := \inf \{t > J_n : X_t \neq X_{J_n}\}$$

*so that J_n is the time of the nth – jump of the process $\{X_t\}$. We further define the **sojourn times** by*

$$S_n := J_{n+1} - J_n$$

so that S_n is the time spent at the nth site which has been visited by the chain. (This notation conflicts with Karlin and Taylor who label the sojourn times by the states in S.)

In this continuous time setting, if $\{X_t\}_{t \geq 0}$ is a collection of functions on Ω with values in S, we let $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$ be the continuous time **filtration** determined by $\{X_t\}_{t \geq 0}$. As usual, a function $f : \Omega \to T$ is $\mathcal{F}_t$ – measurable iff $f = F\left(\{X_s\}_{s \leq t}\right)$ is a function of the path $s \to X_s$ restricted to $[0, t]$. As in the discrete time Markov chain setting, to each $x \in S$, we will write $\mathbb{P}_x(A) := \mathbb{P}(A | X_0 = x)$. That is $\mathbb{P}_x$ is the probability associated to the scenario where the chain is forced to start at site x. We now define, for $x, y \in S$,

$$\mathbf{P}_t(x, y) := \mathbb{P}_x(X_t = y) \tag{15.9}$$

which is the probability of finding the chain at time t at site y given the chain starts at x.

Definition 15.25. *The **time homogeneous Markov property** states for every $0 \leq s < t < \infty$ and bounded functions, $f : S \to \mathbb{R}$, we require*

$$\mathbb{E}_x[f(X_t) | \mathcal{F}_s] = \mathbb{E}_x[f(X_t) | \sigma(X_s)] = (\mathbf{P}_{t-s} f)(X_s)$$

where as before,

$$(\mathbf{P}_t f)(x) = \sum_{y \in S} \mathbf{P}_t(x, y) f(y).$$

A more down to earth statement of this property is that for any choices of $0 = t_0 < t_1 < \cdots < t_n = s < t$ and $x_1, \ldots, x_n \in S$ that

$$\mathbb{P}_x(X_t = y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \mathbf{P}_{t-s}(x_n, y), \tag{15.10}$$

whenever $\mathbb{P}_x(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) > 0$. In particular if $\mathbb{P}(X_s = z) > 0$ then

$$\mathbb{P}_x(X_t = y | X_s = z) = \mathbf{P}_{t-s}(z, y). \tag{15.11}$$

Roughly speaking the Markov property may be stated as follows; the probability that $X_t = y$ given knowledge of the process up to time s is $\mathbf{P}_{t-s}(X_s, y)$. In symbols we might express this last sentence as

$$\mathbb{P}_x\left(X_t = y | \{X_\tau\}_{\tau \leq s}\right) = \mathbb{P}_x(X_t = y | X_s) = \mathbf{P}_{t-s}(X_s, y).$$

So again a continuous time Markov process is forgetful in the sense what the chain does for $t \geq s$ depend only on where the chain is located, X_s, at time s and not how it got there. See Fact 15.41 below for a more general statement of this property. The next theorem gives our first basic description of a continuous time Markov process on a discrete state space in terms of "finite dimensional distributions," see Figure 15.6.

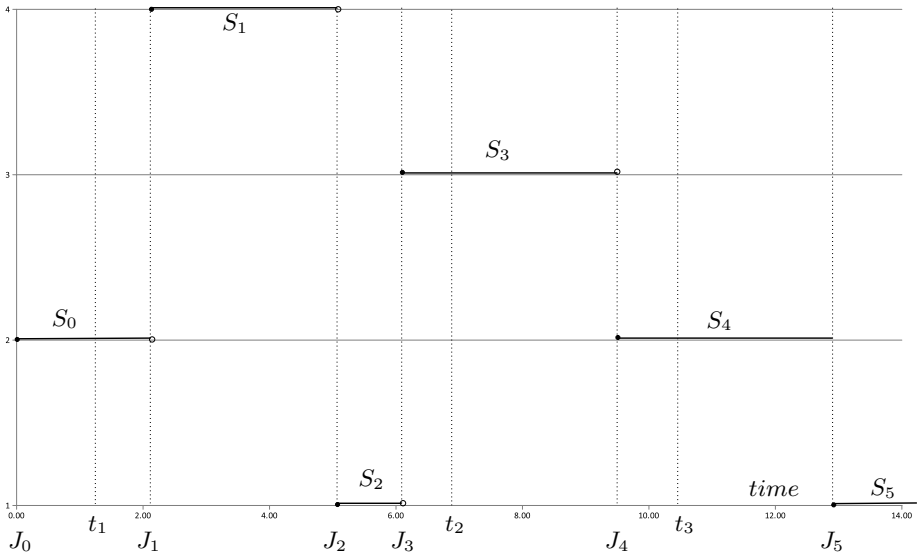


Fig. 15.6. In this figure we fix three time windows, t_1, t_2, and t_3 where we are going to probe the process, $\{X_t\}_{t\geq 0}$. In general, this can be done for any finite number of times and this is what we mean by finite dimensional distributions.

Theorem 15.26 (Finite dimensional distributions). *Let $0 < t_1 < t_2 < \cdots < t_n$ and $x_0, x_1, x_2, \ldots, x_n \in S$. Then*

$$\begin{aligned}\mathbb{P}_{x_0}(X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_n} = x_n) \\ = \mathbf{P}_{t_1}(x_0, x_1)\,\mathbf{P}_{t_2-t_1}(x_1, x_2)\ldots\mathbf{P}_{t_n-t_{n-1}}(x_{n-1}, x_n)\,. \end{aligned} \tag{15.12}$$

Proof. The proof is similar to that of Theorems 5.11 and Proposition 5.13. For notational simplicity let us suppose that $n = 3$. We then have

$$\begin{aligned}&\mathbb{P}_{x_0}(X_{t_1} = x_1, X_{t_2} = x_2, X_{t_3} = x_3) \\ &\quad = \mathbb{P}_{x_0}(X_{t_3} = x_3 | X_{t_1} = x_1, X_{t_2} = x_2)\mathbb{P}_{x_0}(X_{t_1} = x_1, X_{t_2} = x_2) \\ &\quad = \mathbf{P}_{t_3-t_2}(x_2, x_3)\,\mathbb{P}_{x_0}(X_{t_1} = x_1, X_{t_2} = x_2) \\ &\quad = \mathbf{P}_{t_3-t_2}(x_2, x_3)\,\mathbb{P}_{x_0}(X_{t_2} = x_2 | X_{t_1} = x_1)\,\mathbb{P}_{x_0}(X_{t_1} = x_1) \\ &\quad = \mathbf{P}_{t_3-t_2}(x_2, x_3)\,\mathbf{P}_{t_2-t_1}(x_1, x_2)\,\mathbf{P}_{t_1}(x_0, x_1)\end{aligned}$$

wherein we have used the Markov property once in line 2 and twice in line 4. ■

Proposition 15.27 (Properties of P). *Let $\mathbf{P}_t(x, y) := \mathbb{P}_x(X_t = y)$ be as above. Then:*

1. *For each $t \geq 0$, $\mathbf{P}_t$ is a Markov matrix, i.e.*

$$\sum_{y\in S}\mathbf{P}_t(x, y) = 1 \text{ for all } x \in S \text{ and}$$
$$\mathbf{P}_t(x, y) \geq 0 \text{ for all } x, y \in S.$$

2. *$\lim_{t\downarrow 0}\mathbf{P}_t(x, y) = \delta_{xy}$ for all $x, y \in S$.*
3. *The **Chapman − Kolmogorov equation** holds:*

$$\mathbf{P}_{t+s} = \mathbf{P}_s\mathbf{P}_t \text{ for all } s, t \geq 0, \tag{15.13}$$

i.e.

$$\mathbf{P}_{t+s}(x, z) = \sum_{y\in S}\mathbf{P}_s(x, y)\,\mathbf{P}_t(y, z) \text{ for all } s, t \geq 0. \tag{15.14}$$

*We will call a matrix $\{\mathbf{P}_t\}_{t\geq 0}$ satisfying items 1. − 3. a **continuous time Markov semigroup.***

Proof. Most of the assertions follow from the basic properties of conditional probabilities. The assumed right continuity of X_t implies that $\lim_{t\downarrow 0}\mathbf{P}_t = \mathbf{P}(0) = I$. From Equation (15.12) with $n = 2$ we learn that

$$\begin{aligned}\mathbf{P}_{t_2}(x_0, x_2) &= \sum_{x_1\in S}\mathbb{P}_{x_0}(X_{t_1} = x_1, X_{t_2} = x_2) \\ &= \sum_{x_1\in S}\mathbf{P}_{t_1}(x_0, x_1)\,\mathbf{P}_{t_2-t_1}(x_1, x_2) \\ &= [\mathbf{P}_{t_1}\mathbf{P}_{t_2-t_1}](x_0, x_2)\,.\end{aligned}$$

■

Definition 15.28 (Infinitesimal Generator). *The infinitesimal generator, $\mathbf{A}$, of a Markov semi-group $\{\mathbf{P}_t\}_{t\geq 0}$ is the matrix,*

$$\mathbf{A} := \frac{d}{dt}|_{0+}\mathbf{P}_t \tag{15.15}$$

which we are assuming to exist here. [We will not go into the technicalities of this derivative in these notes.]

Assuming that the derivative in Eq. (15.15) exists, then

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{A}\mathbf{P}_t \text{ and (Kolmogorov's Backwards Equation)}$$
$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\mathbf{A} \qquad \text{(Kolmogorov's Forward Equation).}$$

Indeed,

$$\frac{d}{dt}\mathbf{P}_t = \frac{d}{ds}|_0 \mathbf{P}_{t+s} = \frac{d}{ds}|_0 \left[\mathbf{P}_s \mathbf{P}_t\right] = \mathbf{A}\mathbf{P}_t$$
$$\frac{d}{dt}\mathbf{P}_t = \frac{d}{ds}|_0 \mathbf{P}_{t+s} = \frac{d}{ds}|_0 \left[\mathbf{P}_t \mathbf{P}_s\right] = \mathbf{P}_t\mathbf{A}.$$

We also must have;

1. Since $\mathbf{P}_t(x,y) \geq 0$ for all $t \geq 0$ and $x, y \in S$ we have for $x \neq y$ that
$$\mathbf{A}_{x,y} = \lim_{t\downarrow 0} \frac{\mathbf{P}_t(x,y) - \mathbf{P}_0(x,y)}{t} = \lim_{t\downarrow 0} \frac{\mathbf{P}_t(x,y)}{t} \geq 0.$$
2. Since $\mathbf{P}_t\mathbf{1} = \mathbf{1}$ we also have
$$\mathbf{0} = \frac{d}{dt}|_{0^+}\mathbf{1} = \frac{d}{dt}|_{0^+}\mathbf{P}_t\mathbf{1} = \mathbf{A}\mathbf{1},$$
i.e. $\mathbf{A}\mathbf{1} = \mathbf{0}$ and thus
$$a_x := \sum_{y\neq x} \mathbf{A}_{x,y} = -\mathbf{A}_{x,x} \geq 0.$$

Example 15.29. Suppose that $S = \{1, 2, \ldots, n\}$ and $\mathbf{P}_t$ is a Markov-semi-group with infinitesimal generator, $\mathbf{A}$, so that $\frac{d}{dt}\mathbf{P}_t = \mathbf{A}\mathbf{P}_t = \mathbf{P}_t\mathbf{A}$. By assumption $\mathbf{P}_t(x,y) \geq 0$ for all $x, y \in S$ and $\sum_{y=1}^n \mathbf{P}_t(x,y) = 1$ for all $x \in S$. We may write this last condition as $\mathbf{P}_t\mathbf{1} = \mathbf{1}$ for all $t \geq 0$ where $\mathbf{1}$ denotes the vector in $\mathbb{R}^n$ with all entries being 1. Differentiating $\mathbf{P}_t\mathbf{1} = \mathbf{1}$ at $t = 0$ shows that $\mathbf{A}\mathbf{1} = 0$, i.e. $\sum_{y=1}^n \mathbf{A}_{x,y} = 0$ for all $x \in S$. Since
$$\mathbf{A}_{x,y} = \lim_{t\downarrow 0} \frac{\mathbf{P}_t(x,y) - \delta_{xy}}{t}$$
if $x \neq y$ we will have,
$$\mathbf{A}_{x,y} = \lim_{t\downarrow 0} \frac{\mathbf{P}_t(x,y)}{t} \geq 0.$$
Thus we have shown the infinitesimal generator, $\mathbf{A}$, of $\mathbf{P}_t$ must satisfy $\mathbf{A}_{x,y} \geq 0$ for all $x \neq y$ and $\sum_{y=1}^n \mathbf{A}_{x,y} = 0$ for all $x \in S$. In words, $\mathbf{A}$ 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 15.5. So an explicit example of an infinitesimal generator when $S = \{1, 2, 3\}$ is
$$\mathbf{A} = \begin{pmatrix} -3 & 1 & 2 \\ 4 & -6 & 2 \\ 7 & 1 & -8 \end{pmatrix}.$$
In this case my computer finds,
$$\mathbf{P}_t = \begin{pmatrix} \frac{1}{7}e^{-7t} + \frac{1}{5}e^{-10t} + \frac{23}{35} & \frac{1}{7} - \frac{1}{7}e^{-7t} & \frac{1}{5} - \frac{1}{5}e^{-10t} \\ \frac{1}{5}e^{-10t} - \frac{6}{7}e^{-7t} + \frac{23}{35} & \frac{6}{7}e^{-7t} + \frac{1}{7} & \frac{1}{5} - \frac{1}{5}e^{-10t} \\ \frac{1}{7}e^{-7t} - \frac{4}{5}e^{-10t} + \frac{23}{35} & \frac{1}{7} - \frac{1}{7}e^{-7t} & \frac{4}{5}e^{-10t} + \frac{1}{5} \end{pmatrix}.$$

Fact 15.30 *Given a discrete state space S, the the collection of homogeneous Markov probabilities on trajectories, $\{X_t\}_{t\geq 0} \subset S$ as in Figure 15.5 are in one to one correspondence with continuous time Markov semigroups $\mathbf{P}(\cdot)$ which are in one to one correspondence with Markov infinitesimal generators A where $\mathbf{A}$ is a matrix indexed by S with the properties stated after Definition 15.28. The correspondences are given (when no explosions occur) by*
$$\mathbf{A} = \frac{d}{dt}|_{0^+}\mathbf{P}_t, \quad \mathbf{P}_t = e^{t\mathbf{A}}, \quad \mathbf{P}_t(x,y) := \mathbb{P}_x(X_t = y),$$
where the probabilities $\mathbb{P}_x$ are determined by Theorem 15.26. In more detail,
$$\{X_t\}_{t\geq 0} \to \mathbf{P}_t(x,y) := \mathbb{P}(X_{t+s} = y | X_s = x) \text{ for all } x, y \in S \text{ and } s, t \geq 0,$$
$$\mathbf{P}_t(x,y) \to \{X_t\}_{t\geq 0} \text{ with } P_x(X_{t_k} = x_k : 1 \leq k \leq n) = \prod_{k=1}^n \mathbf{P}_{t_k - t_{k-1}}(x_{k-1}, x_k),$$
$$\mathbf{P}_t \to \mathbf{A} := \frac{d}{dt}|_{0^+}\mathbf{P}_t \text{ and}$$
$$\mathbf{A} \to \mathbf{P}_t := e^{t\mathbf{A}} = \sum_{n=0}^\infty \frac{t^n}{n!}\mathbf{A}^n.$$

Proposition 15.31 (Poisson process generators). *If $S = \mathbb{N}_0$ and*
$$\mathbf{A} = \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & \ldots \end{array} \\ \hline \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & \ldots \\ 0 & -\lambda & \lambda & 0 & 0 & 0 & \ldots \\ 0 & 0 & -\lambda & \lambda & 0 & 0 & \ldots \\ 0 & 0 & 0 & -\lambda & \lambda & 0 & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}. \tag{15.16}$$
then
$$\mathbf{P}_t = e^{t\mathbf{A}} = e^{-t\lambda} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & \ldots \end{array} \\ \hline \begin{bmatrix} 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \frac{(\lambda t)5}{5!} & \ldots \\ 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \ldots \\ 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \ldots \\ 0 & 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}. \tag{15.17}$$
In other words,
$$\mathbf{P}_t(m,n) = 1_{n\geq m} e^{-t\lambda} \frac{(\lambda t)^{n-m}}{(n-m)!} = \mathbf{P}_t(0, n-m)\, 1_{n\geq m}. \tag{15.18}$$

We will summarize the matrix $\mathbf{A}$ *and hence the Markov chain via the* ***rate diagram****;*

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} \dots \xrightarrow{\lambda} (n-1) \xrightarrow{\lambda} n \xrightarrow{\lambda} \dots.$$

Proof. First proof. Let $e_0(m) = \delta_{0,m}$ and $\pi_t = e_0 \mathbf{P}_t$ be the top row of $\mathbf{P}_t$. Then

$$\begin{aligned} \frac{d}{dt}\pi_t(0) &= -\lambda \pi_t(0) \text{ and} \\ \frac{d}{dt}\pi_t(n) &= -\lambda \pi_t(n) + \lambda \pi_t(n-1). \end{aligned}$$

Thus if $w_t(n) := e^{\lambda t}\pi_t(n)$ we have with the convention that $w_t(-1) = 0$ that

$$\frac{d}{dt} w_t(n) = \lambda w_t(n-1) \text{ for all } n \text{ with } w_0(n) = \delta_{0,n}.$$

The solution to this system of equations is $w_t(n) = \frac{(\lambda t)^n}{n!}$ as can be found by induction on n. This justifies the first row $\mathbf{P}_t$ given in Eq. (15.17). The other rows are found similarly or using the above results and a little thought. The point is that the chain starting at $m \in \mathbb{N}_0$ behaves just as it has started at 0 except we have to shift the origin to n. To be more precise we should expect that if $\{N_t\}_{t\ge 0}$ is the Poisson process started at 0 then $N_t + m$ is the Poisson process started at n and hence

$$\begin{aligned} \mathbf{P}_t(m,n) &= \mathbb{P}_0(N_t + m = n) \\ &= \mathbb{P}_0(N_t = n - m) = 1_{n\ge m} e^{-t\lambda} \frac{(\lambda t)^{n-m}}{(n-m)!}. \end{aligned}$$

Second proof. Suppose that $\mathbf{P}_t$ is given as in Eq. (15.17) or equivalently Eq. (15.18). The row sums are clearly equal to one and we can check directly that $\mathbf{P}_t \mathbf{P}_s = \mathbf{P}_{s+t}$. To verify this we have,

$$\begin{aligned} \sum_{k\in S} \mathbf{P}_t(i,k)\mathbf{P}_s(k,j) &= \sum_{k\in\mathbb{N}_0} 1_{k\ge i} e^{-\lambda t}\frac{(\lambda t)^{k-i}}{(k-i)!} 1_{j\ge k} e^{-\lambda s}\frac{(\lambda s)^{j-k}}{(j-k)!} \\ &= 1_{i\le j} e^{-\lambda(t+s)} \sum_{i\le k\le j} \frac{(\lambda t)^{k-i}}{(k-i)!}\frac{(\lambda s)^{j-k}}{(j-k)!}. \end{aligned} \tag{15.19}$$

Letting $k = i + m$ with $0 \le m \le j - i$, then the above sum may be written as

$$\sum_{m=0}^{j-i} \frac{(\lambda t)^m}{m!}\frac{(\lambda s)^{j-i-m}}{(j-i-m)!} = \frac{1}{(j-i)!}\sum_{m=0}^{j-i}\binom{j-i}{m}(\lambda t)^m(\lambda s)^{j-i-m}$$

and hence by the Binomial formula we find,

$$\sum_{i\le k\le j} \frac{(\lambda t)^{k-i}}{(k-i)!}\frac{(\lambda s)^{j-k}}{(j-k)!} = \frac{1}{(j-i)!}(\lambda t + \lambda s)^{j-i}.$$

Combining this with Eq. (15.19) shows that

$$\sum_{k\in S} \mathbf{P}_t(i,k)\mathbf{P}_s(k,j) = \mathbf{P}_{s+t}(i,j)$$

as desired. To finish the proof we now need only observe that $\frac{d}{dt}|_0 \mathbf{P}_t = \mathbf{A}$ which follows from the facts that

$$\begin{aligned} \frac{d}{dt}e^{-\lambda t}|_{t=0} &= -\lambda, \ \frac{d}{dt}(\lambda t)e^{-\lambda t}|_{t=0} = \lambda, \text{ and} \\ \frac{d}{dt}\frac{(\lambda t)^k}{k!}e^{-\lambda t}|_{t=0} &= 0 \text{ for all } k \ge 2. \end{aligned}$$

■

Definition 15.32. *The continuous time Markov chain* $\{N_t\}_{t\ge 0}$ *associated to* $\mathbf{A}$ *in Eq. (15.16) when started at* $N_0 = 0$ *is called the* ***Poisson process with intensity*** λ.

Definition 15.33. *A stochastic process* $\{X_t\}_{t\ge 0}$ *is said to have independent increments if for any* $n \in \mathbb{N}$ *and* $0 = t_0 < t_1 < \dots < t_n < \infty$ *the collection of random functions,* $\{X_{t_i} - X_{t_{i-1}}\}_{i=1}^n$ *are independent.*

Proposition 15.34. *The Poisson process*$\{N_t\}_{t\ge 0}$ *has independent increments and moreover for each* $0 \le s \le t < \infty$,

$$N_t - N_s \overset{d}{=} \mathrm{Poi}(\lambda(t-s)).$$

Proof. Let us give the proof when $n = 3$ as the general case is similar. Then given $k_1, k_2, k_3 \in \mathbb{N}_0$ and $0 = t_0 < t_1 < t_2 < t_3 < \infty$ we find,

$$\begin{aligned} &\mathbb{P}_0(N_{t_1} - N_{t_0} = k_1, N_{t_2} - N_{t_1} = k_2, N_{t_3} - N_{t_2} = k_3) \\ &= \mathbb{P}_0(N_{t_1} = k_1, N_{t_2} = k_1 + k_2, N_{t_3} = k_1 + k_2 + k_3) \\ &= \mathbf{P}_{t_1}(0,k_1)\mathbf{P}_{t_2-t_1}(k_1, k_1+k_2)\mathbf{P}_{t_3-t_3}(k_1+k_2, k_1+k_2+k_3) \\ &= \mathbf{P}_{t_1}(0,k_1)\mathbf{P}_{t_2-t_1}(0,k_2)\mathbf{P}_{t_3-t_3}(0,k_3) \\ &= e^{-\lambda t_1}\frac{(\lambda t_1)^{k_1}}{k_1!}e^{-\lambda t_2}\frac{(\lambda(t_2-t_1))^{k_2}}{k_2!}e^{-\lambda t_3}\frac{(\lambda(t_3-t_2))^{k_3}}{k_3!}, \end{aligned}$$

which suffices to verify all the assertions in the proposition. ■

Corollary 15.35. *Let $\{S_n\}_{n=0}^{\infty}$ be i.i.d. $E(\lambda)$ – random times, $J_0 = 0$ and $J_k = S_0 + S_1 + \cdots + S_{k-1}$ for $k \geq 1$. If we now define, for $0 \leq t < \infty$,*

$$N_t := \#\{k \in \mathbb{N}_0 : J_k \leq t\},$$

then $\{N_t\}_{t\geq 0}$ is the Poisson Markov process with parameter λ and in particular,

$$\mathbb{P}(J_k \leq t < J_{k+1}) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

[You are asked to verify these assertions more directly in Section 15.6 below.]

Proof. We simply need to observe that we have really just given the jump-hold description of the Poisson process. In more detail, for the Poisson process generator ($\mathbf{A}$) in Eq. (15.16) the associated non-lazy Markov matrix is

$$\mathbf{Q} = \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & \dots \end{array} \\ \hline \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \end{array} \begin{array}{c} \\ \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}.$$

So assuming we start the non-lazy chain $\{Y_n\}_{n=0}^{\infty}$ associated to $\mathbf{Q}$ at $Y_0 = 0$, then we will simply have $Y_n = n$ for all n. According to Theorem 15.14, the holding times, S_n, at the $Y_n = n$ is $S_n \overset{d}{=} E(\lambda)$ and moreover the $\{S_n\}_{n=0}^{\infty}$ are independent. Thus the jump hold description of the sample paths of the Poisson process is;

1. remain at 0 for S_0 seconds then jump to 1,
2. remain at 1 for S_1 – seconds and then jump to 2,
3. remain at 2 for S_2 seconds and then jump to 3,
4. etc. etc.

In terms of formulas we have

$$N_t = \sum_{k=0}^{\infty} k \cdot 1_{J_k \leq t < J_{k+1}} = \#\{k \in \mathbb{N}_0 : J_k \leq t\}$$

and in particular,

$$\mathbb{P}(J_k \leq t < J_{k+1}) = \mathbb{P}(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

■

Exercise 15.5. Suppose that $S = \{1, 2, \dots, n\}$ and $\mathbf{A}$ is a matrix such that $\mathbf{A}_{i,j} \geq 0$ for $i \neq j$ and $\sum_{j=1}^{n} \mathbf{A}_{i,j} = 0$ for all i. Show

$$\mathbf{P}_t = e^{t\mathbf{A}} := \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{A}^n \tag{15.20}$$

is a time homogeneous Markov kernel.

Hints: 1. To show $\mathbf{P}_t(i,j) \geq 0$ for all $t \geq 0$ and $i, j \in S$, write $\mathbf{P}_t = e^{-t\lambda}e^{t(\lambda I + \mathbf{A})}$ where $\lambda > 0$ is chosen so that $\lambda I + \mathbf{A}$ has only non-negative entries. 2. To show $\sum_{j\in S} P_t(i,j) = 1$, compute $\frac{d}{dt}P_t\mathbf{1}$ where $\mathbf{1}$ is the column vector of all 1's.

Theorem 15.36 (Feynmann-Kac Formula). *Continue the notation in Exercise 15.5 and let $\left(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}_x, \{X_t\}_{t\geq 0}\right)$ be a time homogeneous Markov process (assumed to be right continuous) with transition kernels, $\{\mathbf{P}_t\}_{t\geq 0}$. Given $V : S \to \mathbb{R}$, let $T_t := T_t^V$ be defined by*

$$(T_t g)(x) = \mathbb{E}_x\left[\exp\left(\int_0^t V(X_s)\,ds\right) g(X_t)\right] \tag{15.21}$$

for all $g : S \to \mathbb{R}$. Then T_t satisfies,

$$\frac{d}{dt}T_t = T_t(\mathbf{A} + M_V) \text{ with } T_0 = I \tag{15.22}$$

where $M_V g := Vg$ for all $g : S \to \mathbb{R}$, i.e. M_V is the diagonal matrix with $V(1), \dots, V(n)$ being placed in the diagonal entries. We may summarize this result as,

$$\mathbb{E}_x\left[\exp\left(\int_0^t V(X_s)\,ds\right) g(X_t)\right] = \left(e^{t(\mathbf{A}+M_V)}g\right)(x). \tag{15.23}$$

Proof. To see what is going on let us first assume that $\frac{d}{dt}T_t g$ exists in which case we may compute it as, $\frac{d}{dt}(T_t g)(x) = \frac{d}{dh}|_{0^+}(T_{t+h}g)(x)$. Then by the chain rule, the fundamental theorem of calculus, and the Markov property, we find

$$\begin{aligned}
\frac{d}{dt}\left(T_t g\right)(x) &= \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t+h} V\left(X_s\right)ds\right)g\left(X_t\right)\right] \\
&+ \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t} V\left(X_s\right)ds\right)g\left(X_{t+h}\right)\right] \\
&= \mathbb{E}_x\left[\frac{d}{dh}|_{0^+}\exp\left(\int_0^{t+h} V\left(X_s\right)ds\right)g\left(X_t\right)\right] \\
&+ \frac{d}{dh}|_{0^+}\mathbb{E}_x\left[\exp\left(\int_0^{t} V\left(X_s\right)ds\right)\left(e^{h\mathbf{A}}g\right)\left(X_t\right)\right] \\
&= \mathbb{E}_x\left[\exp\left(\int_0^{t} V\left(X_s\right)ds\right)V\left(X_t\right)g\left(X_t\right)\right] \\
&+ \mathbb{E}_x\left[\exp\left(\int_0^{t} V\left(X_s\right)ds\right)\left(\mathbf{A}g\right)\left(X_t\right)\right] \\
&= T_t\left(Vg\right)(x) + T_t\left(\mathbf{A}g\right)(x)
\end{aligned}$$

which gives Eq. (15.22). [It should be clear that $\left(T_0 g\right)(x) = g\left(x\right)$.]

We now give a rigorous proof. For $0 \le \tau \le t < \infty$, let $Z_{\tau,t} := \exp\left(\int_\tau^t V\left(X_s\right)ds\right)$ and let $Z_t := Z_{0,t}$. For $h > 0$,

$$\begin{aligned}
\left(T_{t+h}g\right)(x) &= \mathbb{E}_x\left[Z_{t+h}g\left(X_{t+h}\right)\right] = \mathbb{E}_x\left[Z_t Z_{t,t+h}g\left(X_{t+h}\right)\right] \\
&= \mathbb{E}_x\left[Z_t g\left(X_{t+h}\right)\right] + \mathbb{E}_x\left[Z_t\left[Z_{t,t+h}-1\right]g\left(X_{t+h}\right)\right] \\
&= \mathbb{E}_x\left[Z_t\left(e^{h\mathbf{A}}g\right)\left(X_t\right)\right] + \mathbb{E}_x\left[Z_t\left[Z_{t,t+h}-1\right]g\left(X_{t+h}\right)\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\left(T_{t+h}g\right)(x) - \left(T_t g\right)(x)}{h} &= \mathbb{E}_x\left[Z_t\frac{\left(e^{h\mathbf{A}}g\right)\left(X_t\right) - g\left(X_t\right)}{h}\right] \\
&+ \mathbb{E}_x\left[Z_t\left[\frac{Z_{t,t+h}-1}{h}\right]g\left(X_{t+h}\right)\right]
\end{aligned}$$

and then letting $h \downarrow 0$ in this equation implies,

$$\frac{d}{dh}|_{0^+}\left(T_{t+h}g\right)(x) = \mathbb{E}_x\left[Z_t\left(\mathbf{A}g\right)\left(X_t\right)\right] + \mathbb{E}_x\left[Z_t V\left(X_t\right)g\left(X_{t+h}\right)\right].$$

This shows that T_t is one sided differentiable and this one sided derivatives is given as in Eq. (15.22).

On the other hand for $s, t > 0$, using the Markov property

$$\begin{aligned}
T_{t+s}g\left(x\right) &= \mathbb{E}_x\left[Z_{t+s}g\left(X_{t+s}\right)\right] = \mathbb{E}_x\left[Z_t Z_{t,t+s}g\left(X_{t+s}\right)\right] \\
&= \mathbb{E}_x\left[Z_t\mathbb{E}_{\mathcal{F}_t}\left(Z_{t,t+s}g\left(X_{t+s}\right)\right)\right] = \mathbb{E}_x\left[Z_t\mathbb{E}_{X_t}\left(Z_{0,s}g\left(X_s\right)\right)\right] \\
&= \left(T_t T_s g\right)(x),
\end{aligned}$$

i.e. $\left\{T_t\right\}_{t>0}$ still has the semi-group property. So for $h > 0$,

$$T_{t-h} - T_t = T_{t-h} - T_{t-h}T_h = T_{t-h}\left(I - T_h\right)$$

and hence

$$\frac{T_{t-h} - T_t}{-h} = T_{t-h}\frac{T_h - I}{h} \to T_t\left(\mathbf{A} + V\right) \text{ as } h \downarrow 0$$

using T_t is continuous in t and the result we have already proved. This shows T_t is differentiable in t and Eq. (15.22) is valid. ■

15.5 Jump Hold Description II

In this section, we are going to start with the Markov semi-group (or equivalently the generator, $\mathbf{A}$) and show how to recover the jump hold description of the associated Markov process, see Corollary 15.42 below. The validity of this argument is a consequence of the strong Markov property (see Fact 15.41) along with Theorem 15.39 below.

Definition 15.37. *Given a Markov-semi-group infinitesimal generator,* $\mathbf{A}$, *on* S, *let* $\mathbf{Q} = \tilde{\mathbf{A}}$ *be the matrix indexed by* S *defined by*

$$\mathbf{Q}_{x,y} := \begin{cases} \frac{\mathbf{A}_{x,y}}{a_x} & \text{if } x \ne y \text{ and } a_x \ne 0 \\ 0 & \text{if } x = y \text{ and } a_x \ne 0 \\ 0 & \text{if } x \ne y \text{ and } a_x = 0 \\ 1 & \text{if } x = y \text{ and } a_x = 0 \end{cases}. \tag{15.24}$$

Recall that

$$a_x := -\mathbf{A}_{x,x} = \sum_{y:y\ne x}\mathbf{A}_{x,y}.$$

Notice that $\mathbf{Q}$ is itself a Markov matrix for a non-lazy chain provided $a_x \ne 0$ for all $x \in S$. We will typically denote the associate chain by $\left\{Y_n\right\}_{n=0}^\infty$. Perhaps it is worth remarking if $x \in S$ is a point where $a_x = 0$, then $\pi_t\left(y\right) := e^{t\mathbf{A}}\left(x,y\right) = \left[\delta_x e^{t\mathbf{A}}\right]\left(y\right)$ satisfies,

$$\dot{\pi}_t\left(y\right) = \left[\delta_x\mathbf{A}e^{t\mathbf{A}}\right]\left(y\right) = \sum_{z\in S}\mathbf{A}\left(x,z\right)e^{t\mathbf{A}}\left(z,y\right) = 0$$

so that $\pi_t\left(y\right) = \pi_0\left(y\right) = \delta_x\left(y\right)$. This shows that $e^{t\mathbf{A}}\left(x,y\right) = \delta_x\left(y\right)$ for all $t \ge 0$ and so if the chain starts at x then it stays at x when $a_x = 0$ which explains the last two rows in Eq. (15.24).

Example 15.38. If

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ \hline -3 & 1 & 2 \\ 0 & 0 & 0 \\ 2 & 7 & -9 \end{bmatrix} \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix}, \text{ then } \mathbf{Q} = \begin{bmatrix} 1 & 2 & 3 \\ \hline 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 1 & 0 \\ \frac{2}{9} & \frac{7}{9} & 0 \end{bmatrix} \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix}.$$

Theorem 15.39 (Jump distributions). *Suppose that A is a Markov infinitesimal generator, $x \in S$ is a non-absorbing state ($a_x > 0$). Then for all $y \neq x$ and $t > 0$,*

$$\mathbb{P}_x\left(X_{J_1} = y; J_1 > t\right) = e^{-ta_x} \frac{\mathbf{A}_{x,y}}{a_x} \tag{15.25}$$

In other words, given $X_0 = x$, $J_1 \overset{d}{=} E(a_x)$ and X_{J_1} (the first jump location) is independent of J_1 with

$$\mathbb{P}_x\left(X_{J_1} = y\right) = \frac{\mathbf{A}_{x,y}}{a_x}.$$

Proof. (We will give a bit of informal proof here – the reader will need to consult the references for fully rigorous proof.) Let $J_1 = \inf\{\tau > 0 : X_\tau \neq X_0\}$. Referring to Figure 15.5, we then have[5] with $\mathcal{P}_n = \left\{\frac{k}{n}\tau : 1 \le k \le n\right\}$, that

$$\begin{aligned}
&\mathbb{P}_x\left(J_1 \in [\tau, \tau + d\tau],\ X_{J_1} = y\right) \\
&\quad = \mathbb{P}_x\left(J_1 \in [\tau, \tau + d\tau],\ X_{\tau + d\tau} = y\right) + o(d\tau) \\
&\quad = \lim_{n\to\infty} \mathbb{P}_x\left(\{X_s = x\}_{s \in \mathcal{P}_n},\ X_{\tau + d\tau} = y\right) + o(d\tau) \\
&\quad = \lim_{n\to\infty} \left[\mathbf{P}_{\frac{\tau}{n}}(x,x)\right]^n \cdot \mathbf{P}_{d\tau}(x,y) + o(d\tau) \\
&\quad = \lim_{n\to\infty} \left[1 - \frac{\tau}{n} \cdot a_x + o\left(\frac{1}{n}\right)\right]^n \cdot \mathbf{A}_{x,y} d\tau + o(d\tau) \\
&\quad = e^{-\tau a_x} \mathbf{A}_{x,y} d\tau + o(d\tau).
\end{aligned}$$

Integrating this equation over $\tau > t$ then shows,

$$\mathbb{P}_x\left(J_1 > t,\ X_{J_1} = y\right) = \int_t^\infty e^{-\tau a_x} \mathbf{A}_{x,y} d\tau = \frac{\mathbf{A}_{x,y}}{a_x} e^{-ta_x}.$$

As a check, taking $t = 0$ above shows,

$$\mathbb{P}_x\left(X_{J_1} = y\right) = \frac{\mathbf{A}_{x,y}}{a_x}$$

[5] Although we do not know where in $[t, t + dt]$ that J_x is we do expect that with relatively high likelihood (relative to dt) that there will be at most one jump in this infinitesimally small interval and hence we should have $X_{t+dt} = X_{J_x}$ when $J_x \in [t, t + dt]$.

and summing this equation on $y \in S \setminus \{x\}$ then gives the following required identity,

$$\sum_{y \neq x} \mathbb{P}_x\left(X_{J_1} = y\right) = \sum_{y \neq x} \frac{\mathbf{A}_{x,y}}{a_x} = 1.$$

■

Definition 15.40 (Informal). *A stopping time, T, for $\{X_t\}$, is a random variable with the property that the event $\{T \le t\}$ is determined from the knowledge of $\{X_s : 0 \le s \le t\}$. Alternatively put, for each $t \ge 0$, there is a functional, f_t, such that*

$$1_{T \le t} = f_t\left(\{X_s : 0 \le s \le t\}\right).$$

As in the discrete state space setting, the first time the chain hits some subset of states, $A \subset S$, is a typical example of a stopping time whereas the last time the chain hits a set $A \subset S$ is typically **not** a stopping time. Similar the discrete time setting, the Markov property leads to a strong form of forgetfulness of the chain. This property is again called the **strong Markov property** which we take for granted here.

Fact 15.41 (Strong Markov Property) *If $\{X_t\}_{t \ge 0}$ is a Markov chain, T is a stopping time, and $j \in S$, then, conditioned on $\{T < \infty$ and $X_T = j\}$,*

$$\{X_s : 0 \le s \le T\} \text{ and } \{X_{t+T} : t \ge 0\} \text{ are independent}$$

and $\{X_{t+T} : t \ge 0\}$ has the same distribution as $\{X_t\}_{t \ge 0}$ under $\mathbb{P}_j$.

Using Theorem 15.39 along with the strong Markov property leads to the following jump-hold description of the continuous time Markov chain, $\{X_t\}_{t \ge 0}$, associated to A. The reader should compare this next corollary with its discrete time analogue in Exercise 15.3.

Corollary 15.42. *Let A be the infinitesimal generator of a Markov semigroup $\mathbf{P}_t$. Then the Markov chain, $\{X_t\}$, associated to $\mathbf{P}_t$ may be described as follows. Let $\{Y_k\}_{k=0}^\infty$ denote the discrete time Markov chain with Markov matrix $\tilde{A}$ as in Eq. (15.24). Let $\{S_j\}_{j=0}^\infty$ be random times such that given $\{Y_j = x_j : j \le n\}$, $S_j \overset{d}{=} \exp\left(a_{x_j}\right)$ and the $\{S_j\}_{j=0}^n$ are independent for $0 \le j \le n$.[6] Now let $N_t = \max\{j : S_0 + \cdots + S_{j-1} \le t\}$ (see Figure 15.5) and $X_t := Y_{N_t}$. Then $\{X_t\}_t \ge 0$ is the Markov process starting at x with Markov semi-group, $\mathbf{P}_t = e^{t\mathbf{A}}$.*

Put another way, if $\{T_n\}_{n=0}^\infty$ are i.i.d exponential random times with $T_n \overset{d}{=} E(1)$ for all n, then

[6] A concrete way to chose the $\{S_j\}_{j=0}^\infty$ is as follows. Given a sequence, $\{T_j\}_{j=0}^\infty$, of i.i.d. $\exp(1)$ – random variables which are independent of $\{Y_n\}_{n=0}^\infty$, define $S_j := T_j / Q_{Y_j}$.

$$\{X_{J_n}, S_n\}_{n=0}^{\infty} \stackrel{d}{=} \left\{Y_n, \frac{T_n}{a_{Y_n}}\right\} \tag{15.26}$$

provided $Y_0 \stackrel{d}{=} X_0$.

Proof. First proof. Assuming the process X starts at some $x \in S$, then 1) it stays at x for an $E(a_x)$ – amount of time, S_1, then jump to x_1 with probability $\tilde{A}_{x,x_1}$. Stay at x_1 for an $E(a_{x_1})$ amount of time, S_2, independent of S_1 and then jump to x_2 with probability $\tilde{A}_{x_1,x_2}$. Stay at x_2 for an $E(a_{x_2})$ amount of time, S_3, independent of S_1 and S_2 and then jump to x_3 with probability $\tilde{A}_{x_2,x_3}$, etc. etc. etc. etc. where one should interpret $T \stackrel{d}{=} E(0)$ to mean that $T = \infty$.

Second proof. The argument in the proof of Theorem 15.39 easily generalizes to multiple jump times. For example for $0 \le \tau < u < \infty$ and $y, z \in S$ we find,

$$\begin{aligned}
&\mathbb{P}_x\left(J_1 \in [\tau, \tau+d\tau],\ J_2 \in [u, u+du],\ X_{J_1} = y,\ X_{J_2} = z\right) \\
&\quad \cong \mathbb{P}_x\left(J_1 \in [\tau, \tau+d\tau],\ J_2 \in [u, u+du],\ X_{\tau+d\tau} = y,\ X_{u+du} = z\right) \\
&\quad \cong \lim_{n\to\infty} \mathbb{P}_x\left(\left\{X_{\frac{k}{n}\tau} = x, X_{\tau+d\tau+\frac{k}{n}(u-(\tau+d\tau))} = y\right\}_{s\in\mathcal{P}},\ X_{\tau+d\tau} = y, X_{u+du} = z\right) \\
&\quad \cong \lim_{n\to\infty} \left\{\left[\mathbf{P}_{\frac{\tau}{n}}(x,x)\right]^n \cdot \mathbf{P}_{d\tau}(x,y) \cdot \left[\mathbf{P}_{\frac{u-\tau}{n}}(y,y)\right]^n \mathbf{P}_{du}(y,z)\right\} \\
&\quad \cong \lim_{n\to\infty} \left\{\left(1 - \frac{\tau}{n}a_x + o\left(\frac{1}{n}\right)\right)^n \left(1 - \frac{u-\tau}{n}a_y + o\left(\frac{1}{n}\right)\right)^n \cdot \mathbf{P}_{d\tau}(x,y)\,\mathbf{P}_{du}(y,z)\right\} \\
&\quad = e^{-\tau a_x} e^{-(u-\tau)a_y} \mathbf{A}_{x,y} d\tau \mathbf{A}_{y,z} du + o\left((d\tau, du)\right)
\end{aligned}$$

from which we conclude that

$$\begin{aligned}
&\mathbb{E}_x\left[F\left(J_1, J_2, X_0, X_{J_1}, X_{J_2}\right)\right] \\
&\quad = \sum_{y:y\neq x, z:z\neq y} \int_0^\infty d\tau \int_0^\infty du\ F(\tau, u, x, y, z)\, e^{-\tau a_x} e^{-(u-\tau)a_y} \mathbf{A}_{x,y}\mathbf{A}_{y,z} \\
&\quad = \sum_{y:y\neq x, z:z\neq y} \int_0^\infty d\tau \int_\tau^\infty du\ F(\tau, u, x, y, z)\, a_x e^{-\tau a_x} a_y e^{-(u-\tau)a_y} \frac{\mathbf{A}_{x,y}}{a_x}\frac{\mathbf{A}_{y,z}}{a_y} \\
&\quad = \sum_{y:y\neq x, z:z\neq y} \int_0^\infty d\tau \int_0^\infty du\ F(\tau, v+\tau, x, y, z)\, a_x e^{-\tau a_x} a_y e^{-v a_y} \frac{\mathbf{A}_{x,y}}{a_x}\frac{\mathbf{A}_{y,z}}{a_y}
\end{aligned}$$

where we should take $\frac{\mathbf{A}_{x,y}}{a_x} = 0$ if $a_x = 0$. If T_1 and T_2 are independent exponential random variables with parameter 1 and $\{Y_n\}$ is the Markov chain with transition matrix $\tilde{A}$ given below then

$$\mathbb{E}_x\left[F\left(J_1, J_2, X_0, X_{J_1}, X_{J_2}\right)\right] = \mathbb{E}_x\left[F\left(\frac{T_1}{Y_0}, \frac{T_1}{Y_0} + \frac{T_2}{Y_1}, Y_0, Y_1, Y_2\right)\right].$$

In other words

$$\left(J_1, J_2, X_0, X_{J_1}, X_{J_2}\right) \stackrel{d}{=} \left(\frac{T_1}{Y_0}, \frac{T_1}{Y_0} + \frac{T_2}{Y_1}, Y_0, Y_1, Y_2\right).$$

This has a clear generalization to any number of times.

The proof of Eq. (15.26) is that on one hand,

$$\mathbb{P}\left(\{X_{J_n} = x_n, S_n > s_n\}_{n=0}^N\right) = \mathbb{P}\left(X_0 = x_0\right) e^{-a_{x_0}s_0} \prod_{n=1}^N \left[\mathbf{Q}\left(x_{n-1}, x_n\right) e^{-a_{x_n}s_n}\right]$$

while on the other hand,

$$\begin{aligned}
\mathbb{P}\left(\{Y_n = x_n, T_n/a_{Y_n} > s_n\}_{n=0}^N\right) &= \mathbb{P}\left(\{Y_n = x_n, T_n > a_{x_n}s_n\}_{n=0}^N\right) \\
&= \mathbb{P}\left(\{Y_n = x_n\}_{n=0}^N\right) \cdot \mathbb{P}\left(\{T_n > a_{x_n}s_n\}_{n=0}^N\right) \\
&= \mathbb{P}\left(Y_0 = x_0\right) \prod_{n=1}^N \mathbf{Q}\left(x_{n-1}, x_n\right) \cdot \prod_{n=0}^N e^{-a_{x_n}s_n}.
\end{aligned}$$

Thus we have shown

$$\mathbb{P}\left(\{X_{J_n} = x_n, S_n > s_n\}_{n=0}^N\right) = \mathbb{P}\left(\{Y_n = x_n, T_n/a_{Y_n} > s_n\}_{n=0}^N\right)$$

provided $\mathbb{P}\left(X_0 = x_0\right) = \mathbb{P}\left(Y_0 = x_0\right)$. ■

It is possible to show the description in Corollary 15.42 defines a Markov process with the correct semi-group, $\mathbf{P}_t$. For the details the reader is referred to Norris [14, See Theorems 2.8.2 and 2.8.4].

Here is one more interpretation of the jump hold or **sojourn** description of our continuous time Markov chains. When we arrive at a site $x \in S$ we start a collection of independent exponential random clocks, $\{T_y\}_{y\in S\setminus\{x\}}$ with $T_y \stackrel{d}{=} E(\mathbf{A}_{x,y})$ where we interpret $T_y = \infty$ if $\mathbf{A}_{x,y} = 0$. If the first clock to ring is the y clock, then we jump to y. It is fairly simple to check that $J = \min\{T_z : z \neq x\} \stackrel{d}{=} E(a_x)$ and the probability that the y clock rings first is precisely, $\mathbf{A}_{x,y}/a_x$. At the next site in the chain, we restart the process all over again.

15.6 Jump Hold Construction of the Poisson Process*

The goal of this section is to give a direct proof of Corollary 15.35. We will use the following notation here. Let $\{T_k\}_{k=1}^\infty$ be an i.i.d. sequence of random exponential times with parameter λ, i.e. $\mathbb{P}\left(T_k \in [t, t+dt]\right) = \lambda e^{-\lambda t}dt$. Let $W_0 =$

0 and for each $n \in \mathbb{N}$ let $W_n := T_1 + \cdots + T_n$ be the "**waiting time**" for the nth event to occur. Because of Theorem 15.22 we know that $\lim_{n\to\infty} W_n = \infty$ a.s. [Sorry for the change of notation here! To match with the above section you should identify T_k with S_{k-1} and W_n with J_n.]

Definition 15.43 (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** λ.*

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

$$\Delta_n(T) := \{(w_1, \ldots, w_n) \in \mathbb{R}^n : 0 < w_1 < w_2 < \cdots < w_n < T\}$$

and let

$$\Delta_n := \cup_{T>0} \Delta_n(T) = \{(w_1, \ldots, w_n) \in \mathbb{R}^n : 0 < w_1 < w_2 < \cdots < w_n < \infty\}.$$

In the exercises below it will be very useful to observe that

$$\{N_t = n\} = \{W_n \leq t < W_{n+1}\}.$$

Exercise 15.6. For $0 \leq s < t < \infty$, show by induction on n that

$$a_n(s,t) := \int_{s \leq w_1 \leq w_2 \leq \cdots \leq w_n \leq t} dw_1 \ldots dw_n = \frac{(t-s)^n}{n!}.$$

Exercise 15.7. If $n \in \mathbb{N}$ and $g : \Delta_n \to \mathbb{R}$ bounded (non-negative) function, then

$$\mathbb{E}[g(W_1, \ldots, W_n)] = \int_{\Delta_n} g(w_1, w_2, \ldots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \ldots dw_n. \tag{15.27}$$

As a simple corollary we have the following result which explains the name given to the Poisson process.

Exercise 15.8. Show $N_t \overset{d}{=} \text{Poi}(\lambda t)$ for all $t > 0$.

Exercise 15.9. Show for $0 \leq s < t < \infty$ that $N_t - N_s \overset{d}{=} \text{Poi}(\lambda(t-s))$ and $\{N_s, N_t - N_s\}$ are independent.

Hint: for $m, n \in \mathbb{N}_0$ show using the few exercises above that

$$\mathbb{P}_0(N_s = m, N_t = m+n) = e^{-\lambda s} \frac{(\lambda s)^m}{m!} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}. \tag{15.28}$$

[This result easily generalizes to show $\{N_t\}_{t\geq 0}$ has independent increments and using this one easily verifies that $\{N_t\}_{t\geq 0}$ is a Markov process with transition kernel given as in Eq. (15.18).]

15.6.1 More related results*

Corollary 15.44. *If $n \in \mathbb{N}$, then $W_n \overset{d}{=} Gamma(n, \lambda^{-1})$.*

Proof. Taking $g(w_1, w_2, \ldots, w_n) = f(w_n)$ in Eq. (15.27) we find with the aid of Exercise 15.6 that

$$\begin{aligned}\mathbb{E}[f(W_n)] &= \int_{\Delta_n} f(w_n) \lambda^n e^{-\lambda w_n} dw_1 \ldots dw_n \\ &= \int_0^\infty f(w) \lambda^n \frac{w^{n-1}}{(n-1)!} e^{-\lambda w} dw\end{aligned}$$

which shows that $W_n \overset{d}{=} \text{Gamma}(n, \lambda^{-1})$. ■

Corollary 15.45. *If $t \in \mathbb{R}_+$ and $f : \Delta_n(t) \to \mathbb{R}$ is a bounded (or non-negative) measurable function, then*

$$\begin{aligned}&\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n] \\ &= \lambda^n e^{-\lambda t} \int_{\Delta_n(t)} f(w_1, w_2, \ldots, w_n) dw_1 \ldots dw_n.\end{aligned} \tag{15.29}$$

Proof. Making use of the observation that $\{N_t = n\} = \{W_n \leq t < W_{n+1}\}$, we may apply Eq. (15.27) 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

$$\begin{aligned}&\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n] \\ &= \int_{0<w_1<\cdots<w_n<t<w_{n+1}} f(w_1, w_2, \ldots, w_n) \lambda^{n+1} e^{-\lambda w_{n+1}} dw_1 \ldots dw_n dw_{n+1} \\ &= \int_{\Delta_n(t)} f(w_1, w_2, \ldots, w_n) \lambda^n e^{-\lambda t} dw_1 \ldots dw_n.\end{aligned}$$

■

Definition 15.46 (Order Statistics). *Suppose that $X_1, \ldots, X_n$ are non-negative random variables such that $\mathbb{P}(X_i = X_j) = 0$ 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*

$$\tilde{X}_k = \min_{\#(\Lambda)=k} \max\{X_i : i \in \Lambda\} \tag{15.30}$$

where Λ always denotes a subset of $\{1, 2, \ldots, n\}$ in Eq. (15.30).

The reader should verify that $\tilde{X}_1 \le \tilde{X}_2 \le \cdots \le \tilde{X}_n$, $\{X_1, \ldots, X_n\} = \left\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\right\}$ with repetitions, and that $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n$ if $X_i \ne X_j$ for all $i \ne j$. In particular if $\mathbb{P}(X_i = X_j) = 0$ for all $i \ne j$ then $\mathbb{P}(\cup_{i \ne j}\{X_i = X_j\}) = 0$ and $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n$ a.s.

Exercise 15.10. Suppose that $X_1, \ldots, X_n$ are non-negative[7] random variables such that $\mathbb{P}(X_i = X_j) = 0$ for all $i \ne j$. Show;

1. If $f : \Delta_n \to \mathbb{R}$ is bounded (non-negative) measurable, then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = \sum_{\sigma \in S_n} \mathbb{E}\left[f(X_{\sigma 1}, \ldots, X_{\sigma n}) : X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\right], \tag{15.31}$$
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
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = n! \cdot \mathbb{E}\left[f(X_1, \ldots, X_n) : X_1 < X_2 < \cdots < X_n\right]. \tag{15.32}$$
(It is not important that $f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)$ is not defined on the null set, $\cup_{i \ne j}\{X_i = X_j\}$.)
3. $f : \mathbb{R}_+^n \to \mathbb{R}$ is a bounded (non-negative) measurable symmetric function (i.e. $f(w_{\sigma 1}, \ldots, w_{\sigma n}) = f(w_1, \ldots, w_n)$ for all $\sigma \in S_n$ and $(w_1, \ldots, w_n) \in \mathbb{R}_+^n$) then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = \mathbb{E}\left[f(X_1, \ldots, X_n)\right].$$
4. Suppose that $Y_1, \ldots, Y_n$ is another collection of non-negative random variables such that $\mathbb{P}(Y_i = Y_j) = 0$ for all $i \ne j$ such that
$$\mathbb{E}\left[f(X_1, \ldots, X_n)\right] = \mathbb{E}\left[f(Y_1, \ldots, Y_n)\right]$$
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;
$$f(y_1, \ldots, y_n) = \sum_{\sigma \in S_n} 1_{y_{\sigma 1} < y_{\sigma 2} < \cdots < y_{\sigma n}} g(y_{\sigma 1}, y_{\sigma 2}, \ldots, y_{\sigma n})$$
and then show f is symmetric.

[7] The non-negativity of the X_i are not really necessary here but this is all we need to consider.

Exercise 15.11. 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 $(U_1, \ldots, U_n)$ 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 15.47 (Joint Distributions). *If $\{A_i\}_{i=1}^k \subset \mathcal{B}_{[0,t]}$ is a partition of $[0, t]$, then $\{N(A_i)\}_{i=1}^k$ are independent random variables and $N(A) \overset{d}{=} \mathrm{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_{[0,t]}$ with $m(A) < \infty$. In particular, if $0 < t_1 < t_2 < \cdots < t_n$, then $\left\{N_{t_i} - N_{t_{i-1}}\right\}_{i=1}^n$ are independent random variables and $N_t - N_s \overset{d}{=} \mathrm{Poi}(\lambda(t - s))$ for all $0 \le s < t < \infty$. (We say that $\{N_t\}_{t \ge 0}$ is a stochastic process with **independent increments**.)*

Proof. If $z \in \mathbb{C}$ and $A \in \mathcal{B}_{[0,t]}$, then
$$z^{N(A)} = z^{\sum_{i=1}^n 1_A(W_i)} \text{ on } \{N_t = n\}.$$
Let $n \in \mathbb{N}$, $z_i \in \mathbb{C}$, and define
$$f(w_1, \ldots, w_n) = z_1^{\sum_{i=1}^n 1_{A_1}(w_i)} \ldots z_k^{\sum_{i=1}^n 1_{A_k}(w_i)}$$
which is a symmetric function. On $N_t = n$ we have,
$$z_1^{N(A_1)} \ldots z_k^{N(A_k)} = f(W_1, \ldots, W_n)$$
and therefore,
$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \ldots z_k^{N(A_k)} | N_t = n\right] &= \mathbb{E}\left[f(W_1, \ldots, W_n) | N_t = n\right] \\
&= \mathbb{E}\left[f(U_1, \ldots, U_n)\right] \\
&= \mathbb{E}\left[z_1^{\sum_{i=1}^n 1_{A_1}(U_i)} \ldots z_k^{\sum_{i=1}^n 1_{A_k}(U_i)}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\left(z_1^{1_{A_1}(U_i)} \ldots z_k^{1_{A_k}(U_i)}\right)\right] \\
&= \left(\mathbb{E}\left[\left(z_1^{1_{A_1}(U_1)} \ldots z_k^{1_{A_k}(U_1)}\right)\right]\right)^n \\
&= \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n,
\end{aligned}$$
wherein we have made use of the fact that $\{A_i\}_{i=1}^n$ is a partition of $[0, t]$ so that

$$z_1^{1_{A_1}(U_1)} \dots z_k^{1_{A_k}(U_1)} = \sum_{i=1}^k z_i 1_{A_i}(U_i).$$

Thus it follows that

$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \dots z_k^{N(A_k)}\right] &= \sum_{n=0}^\infty \mathbb{E}\left[z_1^{N(A_1)} \dots z_k^{N(A_k)} | N_t = n\right] \mathbb{P}(N_t = n) \\
&= \sum_{n=0}^\infty \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\
&= \sum_{n=0}^\infty \frac{1}{n!}\left(\lambda \sum_{i=1}^k m(A_i) \cdot z_i\right)^n e^{-\lambda t} \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i) z_i - t\right]\right) \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i)(z_i - 1)\right]\right).
\end{aligned}$$

From this result it follows that $\{N(A_i)\}_{i=1}^n$ are independent random variables and $N(A) = \text{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_\mathbb{R}$ with $m(A) < \infty$.

Alternatively; suppose that $a_i \in \mathbb{N}_0$ and $n := a_1 + \dots + a_k$, then

$$\begin{aligned}
\mathbb{P}[N(A_1) = a_1, \dots, N(A_k) = a_k | N_t = n] &= \mathbb{P}\left[\sum_{i=1}^n 1_{A_l}(U_i) = a_l \text{ for } 1 \le l \le k\right] \\
&= \frac{n!}{a_1! \dots a_k!} \prod_{l=1}^k \left[\frac{m(A_l)}{t}\right]^{a_l} \\
&= \frac{n!}{t^n} \cdot \prod_{l=1}^k \frac{[m(A_l)]^{a_l}}{a_l!}
\end{aligned}$$

and therefore,

$$\begin{aligned}
&\mathbb{P}[N(A_1) = a_1, \dots, N(A_k) = a_k] \\
&\quad = \mathbb{P}[N(A_1) = a_1, \dots, N(A_k) = a_k | N_t = n] \cdot \mathbb{P}(N_t = n) \\
&\quad = \frac{n!}{t^n} \cdot \prod_{l=1}^k \frac{[m(A_l)]^{a_l}}{a_l!} \cdot e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&\quad = \prod_{l=1}^k \frac{[m(A_l)]^{a_l}}{a_l!} \cdot e^{-\lambda t} \lambda^n \\
&\quad = \prod_{l=1}^k \frac{[m(A_l)\lambda]^{a_l}}{a_l!} e^{-\lambda a_l}
\end{aligned}$$

which shows that $\{N(A_l)\}_{l=1}^k$ are independent and that $N(A_l) \stackrel{d}{=} \text{Poi}(\lambda m(A_l))$ for each l. ■

Remark 15.48. If $A \in \mathcal{B}_{[0,\infty)}$ with $m(A) = \infty$, then $N(A) = \infty$ a.s. To prove this observe that $N(A) =\uparrow \lim_{n\to\infty} N(A \cap [0,n])$. Therefore for any $k \in \mathbb{N}$, we have

$$\begin{aligned}
\mathbb{P}(N(A) \ge k) &\ge \mathbb{P}(N(A \cap [0,n]) \ge k) \\
&= 1 - e^{-\lambda m(A\cap[0,n])} \sum_{0 \le l < k} \frac{(\lambda m(A \cap [0,n]))^l}{l!} \to 1 \text{ as } n \to \infty.
\end{aligned}$$

This shows that $N(A) \ge k$ a.s. for all $k \in \mathbb{N}$, i.e. $N(A) = \infty$ a.s.

15.7 First jump analysis

As usual let S be a finite or countable state space, $\mathbf{A}$ be the infinitesimal generator of a continuous time Markov chain $\{X_t\}_{t\ge0}$, and

$$\mathbf{Q}_{x,y} := \frac{\mathbf{A}_{x,y}}{a_x} 1_{x\ne y}$$

be Markov matrix for the underlying non-lazy Markov chain. We will also use the following notation.

Notation 15.49 *Given a right continuous "jump" path $\left(\{X_t\}_{t\ge0}\right)$ with values in S, $x \in S$, and $\tau \in (0,\infty)$, let $x1_{[0,\tau)} * X$ be the right continuos jump path in S defined by*

$$\left[x1_{[0,\tau)} * X\right]_t := \begin{cases} x & \text{if } 0 \le t < \tau \\ X_{t-\tau} & \text{if } t \ge \tau \end{cases}.$$

In words, we start at x and stay there up to time τ and then jump to X_0 at time τ and then follows the path X thereafter, see Figure 15.7.

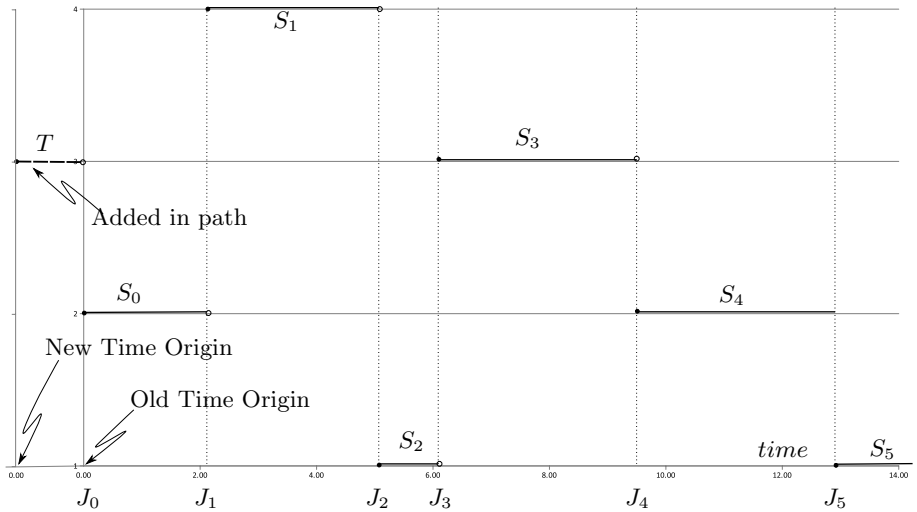


Fig. 15.7. This picture indicates the construction of the new path $3 \cdot 1_{[0,T)} * X$ where X is the path in Figure 15.5.

The next two results summarizes the "first jump" analysis in this continuous time context.

Theorem 15.50 (First jump conditioning). *If $X \to F(X)$ is a bounded or non-negative function of the paths $\{X_t\}_{t \geq 0}$ and $x, y \in S$ with $x \neq y$, then*

$$\mathbb{E}_x [F(X) | X_{S_0} = y] = \mathbb{E}_y \left[F\left(\left[x 1_{[0,T)} * X \right] \right) \right] \tag{15.33}$$

where T is a random time independent of X with $T \overset{d}{=} E(a_x)$. As usual

$$S_0 := J_1 = \inf \{ t > 0 : X_t \neq X_0 \}$$

denotes the first jump time of X.

Proof. By the jump hold description of the chain X starting at x and conditioned on $X_{S_0} = y$, we know that X_t stays at x until time $S_0 \overset{d}{=} E(a_x)$ and then jumps to y and continues on from there independent of the holding time S_0. These simple remarks constitute the proof of Eq. (15.33). ■

The following corollary is now an immediate consequence of Theorem 15.50.

Corollary 15.51 (First jump analysis). *If $X \to F(X)$ is a bounded or non-negative function of the paths $\{X_t\}_{t \geq 0}$, $x \in S$, and $T \overset{d}{=} E(a_x)$ is independent of X, then*

$$\mathbb{E}_x [F(X)] = \sum_{y \in S} \mathbf{Q}_{x,y} \mathbb{E}_x [F(X) | X_{S_0} = y] = \sum_{y \in S} \mathbf{Q}_{x,y} \mathbb{E}_y \left[F\left(\left[x 1_{[0,T)} * X \right] \right) \right]. \tag{15.34}$$

[Recall that if X has not absorbing sites that $\mathbf{Q}_{x,x} = 0$.]

Proof. Using Theorem 15.50 along with the identity, $\mathbb{P}_x (X_{S_0} = y) = \mathbf{Q}_{xy}$, shows

$$\begin{aligned} \mathbb{E}_x [F(X)] &= \sum_{y \in S} \mathbb{E}_x [F(X) | X_{S_0} = y] \mathbb{P}_x (X_{S_0} = y) \\ &= \sum_{y \in S} \mathbb{E}_x [F(X) | X_{S_0} = y] \mathbf{Q}_{x,y} \\ &= \sum_{y \in S} \mathbf{Q}_{x,y} \mathbb{E}_y \left[F\left(\left[x 1_{[0,T)} * X \right] \right) \right]. \end{aligned}$$

■

Example 15.52 (Expected Hitting Times). For example, let B be a proper subset of S, $A := S \setminus B$, and

$$w(x) := \mathbb{E}_x [H_B(X)]$$

where $H_B(X) = \inf \{ t \geq 0 : X_t \in B \}$. By Theorem 15.50 with $T \overset{d}{=} E(a_x)$ independent of X we have

$$\begin{aligned} \mathbb{E}_x [H_B(X) | X_{J_1} = y] &= \mathbb{E}_y \left[H_B \left(\left[x 1_{[0,T)} * X \right] \right) \right] = \mathbb{E}_y [T + H_B(X)] \\ &= \frac{1}{a_x} + \mathbb{E}_y [H_B(X)]. \end{aligned}$$

Therefore by Corollary 15.51,

$$\begin{aligned} \mathbb{E}_x [H_B(X)] &= \sum_{y \in S} \left(\frac{1}{a_x} + \mathbb{E}_y [H_B(X)] \right) \mathbf{Q}_{xy} \\ &= \frac{1}{a_x} + \sum_{y \in A} \mathbf{Q}_{xy} \mathbb{E}_y [H_B(X)]. \end{aligned}$$

Solving this equation for $\mathbf{w} := \{\mathbb{E}_x [H_B(X)]\}_{x \in A}$ gives, (assuming $(\mathbf{I} - \mathbf{Q}_{A \times A})^{-1}$ is invertible)

$$\mathbf{w} = (\mathbf{I} - \mathbf{Q}_{A \times A})^{-1} \frac{1}{\mathbf{a}|_A}$$

wherein we think of $\mathbf{w}$ and $\frac{1}{\mathbf{a}|_A} := \left\{ \frac{1}{a_x} \right\}_{x \in A}$ as column vectors.

Exercise 15.12 (Expected return times). Let $\mathbf{A}$ be a infinitesimal generator of a continuous time Markov chain $\{X_t\}_{t\geq 0}$ on S with $0 < a_x = -\mathbf{A}_{x,x} \leq K < \infty$ for all $x \in S$. Further let

$$R_y := \inf\{t \geq S_0 : X_t = y\}$$

be the first return time[8] to y on or after the first jump and $m_{x,y} := \mathbb{E}_x R_y$. Use the first jump analysis to show,

$$m_{xy} = \frac{1}{a_x} + \sum_{z\neq y} \mathbf{Q}_{x,z} m_{zy} = \frac{1}{a_x} + \sum_{z\notin\{x,y\}} \frac{\mathbf{A}_{xz}}{a_x} m_{zy}, \tag{15.35}$$

where $\mathbf{Q}_{x,z} := 1_{x\neq y}\mathbf{A}_{x,y}/a_x$. [You may find it useful to observe; if $H_y := \inf\{t \geq 0 : X_t = y\}$ is the first hitting time of y, then $\mathbb{E}_x H_y = \mathbb{E}_x R_y$ if $x \neq y$ and $\mathbb{E}_y H_y = 0$.]

Definition 15.53 (Invariant distributions). *A probability distribution, $\pi : S \to [0,1]$ is said to be an **invariant distribution** for $\mathbf{A}$ if $\pi\mathbf{A} = 0$, i.e.*

$$\sum_{x\in S} \pi_x \mathbf{A}_{x,y} = 0 \text{ for all } y \in S.$$

Remark 15.54. If π is an invariant distribution for $\mathbf{A}$ and $\mathbf{P}_t = e^{t\mathbf{A}}$, then $\frac{d}{dt}\pi\mathbf{P}_t = \pi\mathbf{A}\mathbf{P}_t = 0$ and hence $\pi\mathbf{P}_t = \pi\mathbf{P}_0 = \pi$. Consequently, if $\{X_t\}_{t\geq 0}$ is the continuous time Markov chain associated to $\mathbf{A}$, then $\mathbb{P}_\pi(X_t = y) = \pi_y$ for all $t > 0$ and $y \in S$.

Corollary 15.55. *Let $\{X_t\}_{t\geq 0}$ be a finite state irreducible Markov chain[9] with generator, $\mathbf{A} = (\mathbf{A}_{xy})_{x,y\in S}$. If $\pi = (\pi_x)_{x\in S}$ is an invariant distribution for $\mathbf{A}$, then*

$$\pi_x = \frac{1}{a_x m_{xx}} = \frac{1}{a_x \mathbb{E}_x R_x} = \frac{\mathbb{E}_x S_0}{\mathbb{E}_x R_x}. \tag{15.36}$$

[Intuitively (and as one would anticipate), Eq. (15.36) indicates that π_x is the relative fraction of the time the chain spends at site x.]

Proof. Since the chain is irreducible we will definitely return to x and in fact one can show $m_{x,x} < \infty$ – we omit this detail here. As π is an invariant distribution for $\mathbf{A}$, we are given

$$\sum_{x:x:\neq z} \pi_x \mathbf{A}_{xz} = -\pi_z \mathbf{A}_{zz} = \pi_z a_z \;\forall\; z \in S.$$

Using this identity along with Eq. (15.35) allows us to conclude;

$$\begin{aligned}
\sum_{x\in S} \pi_x a_x m_{xy} &= \sum_{x\in S} \pi_x a_x \left[\frac{1}{a_x} + \sum_{z\neq y} \mathbf{Q}_{x,z} m_{zy}\right] \\
&= 1 + \sum_{x,z} \pi_x 1_{x\neq z, z\neq y} \mathbf{A}_{xz} m_{zy} \\
&= 1 + \sum_{z} 1_{z\neq y} \pi_z a_z m_{zy} \\
&= 1 + \sum_{x\in S} 1_{x\neq y} \pi_x a_x m_{xy}.
\end{aligned}$$

Hence it follows that $\pi_x a_x m_{xx} = 1$ which proves Eq. (15.36). ■

Example 15.56. Consider the two state Markov chain with rate diagram being $0 \underset{\beta}{\overset{\alpha}{\rightleftarrows}} 1$ and let $m_0 = \mathbb{E}_0 R_0$ and $m_1 = \mathbb{E}_1 R_0$. Then the first jump analysis gives,

$$\mathbb{E}_1 R_0 = m_1 = \mathbb{E}_1[R_0 | X_{S_0} = 0]\, P(X_{S_0} = 0) = \mathbb{E}_1[S_0] = \frac{1}{\beta} \text{ and}$$

$$m_0 = \mathbb{E}_0[R_0 | X_{S_0} = 1]\, P(X_{S_0} = 1) = (\mathbb{E}_0[S_0] + \mathbb{E}_1[R_0]) \cdot 1 = \frac{1}{\alpha} + m_1.$$

Therefore $\mathbb{E}_0 R_0 = m_0 = \frac{1}{\alpha} + \frac{1}{\beta}$ and

$$\pi_0 = \frac{\mathbb{E}_0 S_0}{\mathbb{E}_0 R_0} = \frac{1}{\alpha} \cdot \frac{1}{\frac{1}{\alpha} + \frac{1}{\beta}} = \frac{1}{1 + \alpha/\beta}$$

and by symmetry,

$$\pi_1 = \frac{1}{1 + \beta/\alpha}.$$

Observe that

$$\pi_0 + \pi_1 = \frac{1}{1+\alpha/\beta} + \frac{1}{1+\beta/\alpha} = \frac{1 + \beta/\alpha + 1 + \alpha/\beta}{(1+\alpha/\beta)(1+\beta/\alpha)} = 1$$

and using

$$-\alpha\frac{1}{1+\alpha/\beta} + \beta\frac{1}{1+\beta/\alpha} \propto -\alpha(1+\beta/\alpha) + \beta(1+\alpha/\beta) = -\alpha - \beta + \beta + \alpha = 0$$

implies

$$\pi\mathbf{A} = \begin{bmatrix} \frac{1}{1+\alpha/\beta} & \frac{1}{1+\beta/\alpha} \end{bmatrix} \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

as we have shown should happen in general.

[8] We require $t \geq S_0$ so that if the chain starts at y it must first leave y before we count being at y again a return to y.

[9] In detail we are assuming that $a_x > 0$ for all $x \in S$ and $\mathbf{Q}_{x,y} = 1_{x\neq y}\mathbf{A}_{x,y}/a_x$ is an irreducible Markov matrix.

15.8 Long time behavior

In this section, suppose that $\{X_t\}_{t\geq 0}$ is a continuous time Markov chain with infinitesimal generator, A, so that

$$\mathbb{P}\left(X_{t+h}=y|X_t=x\right)=\delta_{xy}+\mathbf{A}_{x,y}h+o\left(h\right).$$

We further assume that A completely determines the chain and that $\mathbf{A}$ has no absorbing states, i.e. $a_x:=-\mathbf{A}_{x,x}>0$ for all $x\in S$.

Definition 15.57. *The chain, $\{X_t\}$, is **irreducible** iff the underlying discrete time jump chain, $\{Y_n\}$, determined by the Markov matrix, $\mathbf{Q}_{x,y}:=\frac{\mathbf{A}_{x,y}}{a_x}1_{x\neq y}$, is irreducible, where as usual $a_x:=-\mathbf{A}_{x,x}=\sum_{y:y\neq x}\mathbf{A}_{x,y}$. Put more directly, $\{X_t\}$, is irreducible iff for all $x,y\in S$ with $x\neq y$ there exists $x_0,x_1,\ldots,x_n\in S$ with such that $x_0=x$, $x_n=y$, and $x_{i+1}\neq x_i$ and $A_{x_i,x_{i+1}}>0$ for all $0\leq i<n$.*

Lemma 15.58. *Let $\mathbf{A}$ be the generator of a continuous time Markov chain $\{X_t\}_{t\geq 0}$ and $\mathbf{P}_t=e^{t\mathbf{A}}$ then the following are equivalent;*

1. *$\{X_t\}$ is irreducible,*
2. *for all $x,y\in S$, $\mathbf{P}_t\left(x,y\right)>0$ for some $t>0$, and*
3. *$\mathbf{P}_t\left(x,y\right)>0$ for all $t>0$ and $x,y\in S$.*

[In particular, all irreducible chains are "aperiodic."]

Proof. (1. $\implies$ 2.) Let $\{Y_n\}$ be the discrete time Markov chain associated to $\mathbf{Q}$. If $\mathbf{Q}$ is irreducible, $t>0$, and $x,y\in S$ with $y\neq x$, let $x_0,x_1,\ldots,x_n\in S$ be as in Definition 15.57 and $\{T_j\}_{j=0}^n$ be independent random times such that $T_j\overset{d}{=}E\left(a_{x_j}\right)$. Using the jump hold description of $\{X_t\}_{t\geq 0}$ we then have

$$\begin{aligned}\mathbf{P}_t\left(x,y\right)&=\mathbb{P}_x\left(X_t=y\right)\geq\mathbb{P}_x\left(Y_n=y,J_n\leq t<J_{n+1}\right)\\&\geq\prod_{i=0}^{n-1}\mathbf{Q}_{x_i,x_{i+1}}\cdot\mathbb{P}\left(T_0+\cdots+T_{n-1}<t<T_0+\cdots+T_n\right)>0.\end{aligned}$$

(2. $\implies$ 3.) This is obvious.

(3. $\implies$ 1.) If for all $x,y\in S$, $\mathbf{P}_t\left(x,y\right)>0$ for some $t>0$ then the embedded chain $\{Y_n\}$ as a positive probability of hitting y when started at x. As x and y were arbitrary, it follows that $\{Y_n\}$ must be irreducible and hence X_t is irreducible. ■

The next theorem gives the basic limiting behavior of irreducible Markov chains. Before stating the theorem we need to introduce a little more notation.

Notation 15.59 *Let S_0 be the time of the first jump of X_t, and*

$$R_x:=\min\left\{t\geq S_0:X_t=x\right\},$$

is the first time hitting the site x after the first jump[10], and set

$$\pi_x=\frac{\mathbb{E}_xS_0}{\mathbb{E}_xR_x}=\frac{1}{a_x\cdot\mathbb{E}_xR_x}\text{ where }a_x:=-\mathbf{A}_{x,x}.$$

[So π_x is on average the the fraction of the time the chain spends at site x.] For the sample path in Figure 15.5, $R_1=J_2$, $R_2=J_4$, $R_3=J_3$ and $R_4=J_1$.

Theorem 15.60 (Limiting behavior). *Let $\{X_t\}$ be an irreducible Markov chain. Then*

1. *for all initial staring distributions, $\nu\left(y\right):=\mathbb{P}\left(X\left(0\right)=y\right)$ for all $y\in S$, and all $y\in S$,*
$$\lim_{T\to\infty}\frac{1}{T}\int_0^T 1_{X_t=y}dt=\pi_y\qquad(\mathbb{P}_\nu\text{ a.s.}).\tag{15.37}$$
In words, the fraction of the time the chain spends at site y is π_y.
2. *$\lim_{t\to\infty}\mathbf{P}_t\left(x,y\right)=\pi_y$ independent of x.*
3. *$\pi=\left(\pi_y\right)_{y\in S}$ is an invariant (stationary) distribution for $\mathbf{A}$.*
4. *If $\pi_x>0$ for some $x\in S$, then $\pi_x>0$ for all $x\in S$ and $\sum_{x\in S}\pi_x=1$.*
5. *The π_x are all positive iff there exists a solution, $\nu_x\geq 0$ to*
$$\sum_{x\in S}\nu_x\mathbf{A}_{x,y}=0\text{ for all }y\in S\text{ with }\sum_{x\in S}\nu_x=1.$$
If such a solution exists it is unique and $\nu=\pi$.

Proof. We refer the reader to [14, Theorems 3.8.1.] for the full proof. Let us make a few comments on the proof taking for granted that $\lim_{t\to\infty}\mathbf{P}_t\left(x,y\right)=:\pi_y$ exists. (See part d) below for why the limit is independent of x if it exists.)

1. Suppose we assume that and that ν is a stationary distribution, i.e. $\nu\mathbf{P}_t=\nu$, then (by dominated convergence theorem),
$$\nu_y=\lim_{t\to\infty}\sum_x\nu_x\mathbf{P}_t\left(x,y\right)=\sum_x\lim_{t\to\infty}\nu_x\mathbf{P}_t\left(x,y\right)=\left(\sum_x\nu_x\right)\pi_y=\pi_y.$$
Thus $\nu_y=\pi_y$. If $\pi_y=0$ for all y we must conclude there is no stationary distribution.

[10] We require $t\geq S_0$ so that if the chain starts at x it must first leave x before we count being at x again a return to x. In short, you can not return to x unless you have first left x or did not start at x.

2. If we are in the finite state setting, the following computation is justified:

$$\sum_{y\in S}\pi_y\mathbf{P}_s(y,z)=\sum_{y\in S}\lim_{t\to\infty}\mathbf{P}_t(x,y)\mathbf{P}_s(y,z)=\lim_{t\to\infty}\sum_{y\in S}\mathbf{P}_t(x,y)\mathbf{P}_s(y,z)$$
$$=\lim_{t\to\infty}[\mathbf{P}_t\mathbf{P}_s]_{xz}=\lim_{t\to\infty}\mathbf{P}_{t+s}(x,z)=\pi_z.$$

This show that $\pi\mathbf{P}_s=\pi$ for all s and differentiating this equation at $s=0$ then shows, $\pi A=0$.

3. Let us now explain why

$$\frac{1}{T}\int_0^T 1_{X_t=y}dt\to\frac{\mathbb{E}_yS_0}{\mathbb{E}_yR_y}=\frac{1}{a_y\cdot\mathbb{E}_yR_y}. \tag{15.38}$$

The idea is that, because the chain is irreducible, no matter how we start the chain we will eventually hit the site y. Once we hit y, the (strong) Markov property implies the chain forgets how it got there and behaves as if it started at y. The size of J_1 (as long as it is finite) is negligible when computing the limit in Eq. (15.38) and so we may now as well assume that the chain starts at y.

Now consider one typical cycle in the chain staring at y jumping away at time S_0 and then returning to y at time R_y. The average first jump time is $\mathbb{E}_yS_0=1/a_y$ while the average length of such as cycle is $\mathbb{E}_yR_y$. As the chain repeats this procedure over and over again with the same statistics, we expect (by a law of large numbers) that the average time spent at site y is given by

$$\frac{\mathbb{E}_yS_0}{\mathbb{E}_yR_y}=\frac{1/a_y}{\mathbb{E}_yR_y}=\frac{1}{a_y\cdot\mathbb{E}_yR_y}.$$

Formal argument. To make this last argument rigorous, let $\{\rho_z\}_{z=1}^\infty$ be the successive times between returns to y and $\{\sigma_z\}_{z=1}^\infty$ be the successive sojourn times at y. Letting N_T be chosen so that

$$\rho_1+\cdots+\rho_{N_T}\le T<\rho_1+\cdots+\rho_{N_T+1},$$

then we will have,

$$\frac{\sigma_1+\cdots+\sigma_N}{\rho_1+\cdots+\rho_{N+1}}\le\frac{1}{T}\int_0^T 1_{X_t=y}dt\le\frac{\sigma_1+\cdots+\sigma_N+\sigma_{N+1}}{\rho_1+\cdots+\rho_N}$$

where by the strong law of large numbers, $N=N_T\to\infty$ as $T\to\infty$,

$$\frac{\sigma_1+\cdots+\sigma_N}{\rho_1+\cdots+\rho_{N+1}}=\frac{\frac{\sigma_1+\cdots+\sigma_N}{N}}{\frac{\rho_1+\cdots+\rho_{N+1}}{N}}\to\frac{\mathbb{E}S_0}{\mathbb{E}R_y}=\frac{1}{a_y\mathbb{E}R_y}$$

and

$$\frac{\sigma_1+\cdots+\sigma_{N+1}}{\rho_1+\cdots+\rho_N}=\frac{\frac{\sigma_1+\cdots+\sigma_{N+1}}{N}}{\frac{\rho_1+\cdots+\rho_N}{N}}\to\frac{\mathbb{E}S_0}{\mathbb{E}R_y}=\frac{1}{a_y\mathbb{E}R_y}$$

as $T\to\infty$.

4. Taking expectations of Eq. (15.37) shows (at least when $\#(S)<\infty$)

$$\pi_y=\lim_{T\to\infty}\frac{1}{T}\int_0^T\mathbb{P}_\nu(X_t=y)\,dt=\lim_{T\to\infty}\frac{1}{T}\int_0^T\sum_x\nu_x\mathbf{P}_{x,y}(t)\,dt$$
$$=\sum_x\nu_x\lim_{T\to\infty}\frac{1}{T}\int_0^T\mathbf{P}_t(x,y)\,dt=\sum_x\nu_x\lim_{t\to\infty}\mathbf{P}_t(x,y)$$

for any initial distribution ν. Taking $\nu_x=\delta_{z,x}$ for any $z\in S$ we choose shows $\lim_{t\to\infty}\mathbf{P}_t(z,y)=\pi_y=\frac{1}{a_y\cdot\mathbb{E}_yR_y}$ independent of z.

■

15.9 Some Associated Martingales

This section may be considered to be a warm-up of the martingales appearing relative to Brownian motions in the next chapter. As above let S be a finite or countable state space and $\mathbf{A}$ be the infinitesimal generator of a continuous time Markov chain $\{X_t\}_{t\ge0}$ on S such that $a_x:=-\mathbf{A}_{x,x}$ is bounded for $x\in S$. For a function $[0,T]\times S\ni(t,x)\to h(t,x)\in\mathbb{R}$ which is continuously differentiable function in t such that $\sum_y|\mathbf{A}_{xy}||h(t,y)|<\infty$ for all $x\in S$ we let

$$\mathcal{L}h(t,x):=\dot h(t,x)+(\mathbf{A}h)(t,x)=\frac{\partial}{\partial t}h(t,x)+\sum_y\mathbf{A}_{x,y}h(t,y).$$

Theorem 15.61 (Martingale problem). *Let us continue the above notation and assumptions above and let $\nu:S\to[0,1]$ be a given initial distribution. If we further assume that*

$$\sup_{0\le t\le T}\mathbb{E}_\nu\left[|h(t,X_t)|+\left|\dot h(t,X_t)\right|+|(\mathbf{A}h)(t,X_t)|\right]<\infty,$$

then

$$M_t:=h(t,X_t)-\int_0^t\mathcal{L}h(\tau,X_\tau)\,d\tau$$

is an $\left\{\mathcal{F}_t^X\right\}_{t\in\mathbb{R}_+}$ – martingale. [Here M_t is a right continuous martingale with left hand limits having jumps at (at most) the jump times of $\{X_t\}$.] In particular, if h also satisfies the "heat equation" in reverse time, $\mathcal{L}h(t,x)=0$, then $M_t=h(t,X_t)$ is a martingale.

Proof. Working formally for the moment,

$$\begin{aligned}\frac{d}{d\tau}\mathbb{E}_\nu\left[h\left(\tau,X_\tau\right)|\mathcal{F}_s\right] &= \frac{d}{d\tau}\left[\mathbf{P}_{\tau-s}h\left(\tau,X_s\right)\right] \\ &= \mathbf{P}_{\tau-s}\dot{h}\left(\tau,X_s\right)+\mathbf{P}_{\tau-s}\mathbf{A}h\left(\tau,X_s\right) \\ &= \mathbb{E}_\nu\left[\dot{h}\left(\tau,X_\tau\right)+\mathbf{A}h\left(\tau,X_\tau\right)|\mathcal{F}_s\right] \\ &= \mathbb{E}_\nu\left[\mathcal{L}h\left(\tau,X_\tau\right)|\mathcal{F}_s\right]\end{aligned}$$

wherein we have used the Markov property in the first and third lines and the chain rule in the second line. Integrating this equation on $\tau\in[s,t]$ then shows

$$\mathbb{E}_\nu\left[h\left(t,X_t\right)-h\left(s,X_s\right)|\mathcal{F}_s\right]=\mathbb{E}_\nu\left[\int_s^t\mathcal{L}h\left(\tau,X_\tau\right)d\tau|\mathcal{F}_s\right]. \tag{15.39}$$

Using

$$\mathbb{E}_\nu\left[M_t-M_s|\mathcal{F}_s\right]=\mathbb{E}_\nu\left[h\left(t,X_t\right)-h\left(s,X_s\right)-\int_s^t\mathcal{L}h\left(\tau,X_\tau\right)d\tau|\mathcal{F}_s\right]$$

we see that Eq. (15.39) is precisely the statement that $\mathbb{E}_\nu\left[M_t-M_s|\mathcal{F}_s\right]=0$, i.e. to the assertion that $\{M_t\}_{t\geq 0}$ is a martingale.

The formal proof requires some more justification. Rather than do this here the reader is referred to the analogous Theorem 16.31 in the Brownian motion case for the sorts of arguments needed to fill in the missing details. ■

Corollary 15.62 (Representation formula). *Suppose that $B\subset S$ is a set and $\tau_B=\inf\{t\geq 0:X_t\in B\}$ and assume that $\mathbb{E}_x\tau_B<\infty$ for all $x\in A:=S\setminus B$. If $u:S\to\mathbb{R}$ is a bounded function, then*

$$u\left(x\right)=\mathbb{E}_x\left[u\left(X_{\tau_B}\right)-\int_0^{\tau_B}\left(\mathbf{A}u\right)\left(X_s\right)ds\right]. \tag{15.40}$$

Proof. By Theorem 15.61,

$$M_t:=u\left(X_t\right)-\int_0^t\left(\mathbf{A}u\right)\left(X_s\right)ds$$

is a martingale relative to any starting distribution ν which we now take to be δ_x. By a continuous time variant of the optional sampling theorem, one shows $t\to M_{t\wedge\tau_B}$ is still a martingale and therefore,

$$u\left(x\right)=\mathbb{E}_xM_0=\mathbb{E}_xM_{t\wedge\tau_B}=\mathbb{E}_x\left[u\left(X_{t\wedge\tau_B}\right)-\int_0^{t\wedge\tau_B}\left(\mathbf{A}u\right)\left(X_s\right)ds\right].$$

We may now use DCT to pass to the limit as $t\uparrow\infty$ in the above identity to arrive at Eq. (15.40). ■

16

Brownian Motion

This chapter is devoted to an introduction to Brownian motion on the state space S which is take to be $S = \mathbb{R}$ or more generally $\mathbb{R}^d$ for some d. The first thing we need to do is recall the basic properties of Gaussian random variables and vectors.

16.1 Normal/Gaussian Random Vectors

A random variable, Y, **is normal with mean μ standard deviation σ^2** iff

$$\mathbb{P}\left(Y \in (y, y+dy]\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy. \tag{16.1}$$

Notation 16.1 *Suppose that Y is a random-variable. We write $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$ to mean that Y is a normal random variable with mean μ and variance σ^2. In particular, $Y \overset{d}{=} N(0,1)$ is used to indicate that Y is a **standard normal** random variable.*

It turns out that rather than working with densities of normal random variables it is typically easier to make use of the Laplace (or better Fourier) transform description of the distribution.

Definition 16.2. *A random variable, $Y : \Omega \to \mathbb{R}$, is said to be **exponentially integrable** if $\mathbb{E}\left[e^{\lambda Y}\right] < \infty$ for all $\lambda \in \mathbb{R}$. This is equivalent to requiring $\mathbb{E}\left[e^{\lambda |Y|}\right] < \infty$ for all $\lambda > 0$.*

Lemma 16.3. *If Y is an exponentially integrable random variable, then $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$ iff*

$$\mathbb{E}\left[e^{\lambda Y}\right] = e^{\frac{1}{2}\sigma^2\lambda^2 + \lambda\mu}. \tag{16.2}$$

Proof. If $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$ and $\lambda \in \mathbb{R}$, then using Eq. (16.1) we find,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda Y}\right] &:= \int_{-\infty}^{\infty} e^{\lambda y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy \\
&= \int_{-\infty}^{\infty} e^{\lambda(\sigma x + \mu)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
&= e^{\lambda\mu} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\sigma\lambda)^2} e^{\frac{1}{2}\sigma^2\lambda^2} dx \\
&= e^{\frac{1}{2}\sigma^2\lambda^2 + \lambda\mu} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= e^{\frac{1}{2}\sigma^2\lambda^2 + \lambda\mu},
\end{aligned} \tag{16.3}$$

wherein the second line we made the change of variables, $y = \sigma x + \mu$, completed the squares in the third line, made the change of variables $z = x - \sigma\lambda$ in the fourth line, and used the well known fact that

$$\sqrt{2\pi} = \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

in the last. The converse result relies on the fact that distribution of an exponentially random variables is uniquely determined by its Laplace transform, i.e. by the function $\lambda \to \mathbb{E}\left[e^{\lambda Y}\right]$. ■

Corollary 16.4. *If $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$, then $\mu = \mathbb{E}Y$ and $\sigma^2 = \mathrm{Var}(Y)$.*

Proof. Differentiating Eq. (16.2) (which is justified by DCT) we find,

$$\begin{aligned}
\mathbb{E}\left[Ye^{\lambda Y}\right] &= \left(\sigma^2\lambda + \mu\right) e^{\frac{1}{2}\sigma^2\lambda^2 + \lambda\mu} \text{ and} \\
\mathbb{E}\left[Y^2 e^{\lambda Y}\right] &= \left(\sigma^2 + \left(\sigma^2\lambda + \mu\right)^2\right) e^{\frac{1}{2}\sigma^2\lambda^2 + \lambda\mu}.
\end{aligned}$$

Taking $\lambda = 0$ in these equations then implies,

$$\mu = \mathbb{E}Y \text{ and } \sigma^2 + \mu^2 = \mathbb{E}\left[Y^2\right]$$

and therefore $\sigma^2 = \mathrm{Var}(Y)$. ■

The last two results serve as motivation for the following alternative definition of Gaussian random variables and more generally random vectors.

Definition 16.5 (Normal / Gaussian Random Variable). *An $\mathbb{R}^d$ –valued random vector, $\mathbf{Z}$, is said to be **normal or Gaussian**[1] if for all $\lambda \in \mathbb{R}^d$, $\lambda \cdot \mathbf{Z}$*

[1] I will use the terms, normal and Gaussian, interchangeably.

is a Gaussian random variable. Alternatively stated, $\mathbf{Z}$ *is a* ***Gaussian random vector*** *iff*

$$\mathbb{E}\left[e^{\lambda\cdot\mathbf{Z}}\right] = \exp\left(\frac{1}{2}\operatorname{Var}(\lambda\cdot\mathbf{Z}) + \mathbb{E}(\lambda\cdot\mathbf{Z})\right)$$

for all $\lambda \in \mathbb{R}^d$. *We say that* $\mathbf{Z}$ *is a* ***standard normal random vector*** *if* $\mathbb{E}[\lambda\cdot\mathbf{Z}] = 0$ *and* $\operatorname{Var}(\lambda\cdot\mathbf{Z}) = \lambda\cdot\lambda$ *for all* $\lambda \in \mathbb{R}^d$.

If we let $\mu = (\mu_i := \mathbb{E}\mathbf{Z}_i)_{i=1}^d$ be the **mean vector** and $\mathbf{C}_{ij} := \operatorname{Cov}(\mathbf{Z}_i, \mathbf{Z}_j)$ be the **covariance matrix** of $\mathbf{Z}$ then

$$\begin{aligned}\mathbb{E}[\lambda\cdot\mathbf{Z}] &= \mu\cdot\lambda \text{ and}\\ \operatorname{Var}(\lambda\cdot\mathbf{Z}) &= \mathbb{E}\left[(\lambda\cdot\mathbf{Z})^2\right] - (\mathbb{E}[\lambda\cdot\mathbf{Z}])^2\\ &= \sum_{i,j=1}^d \operatorname{Cov}(\lambda_i Z_i, \lambda_j Z_j) = \mathbf{C}\lambda\cdot\lambda.\end{aligned}$$

With this notation, a Gaussian random vector, $\mathbf{Z}$, is standard normal iff $\mu = 0$ and $\mathbf{C} = I$.

Lemma 16.6. *If* $\mathbf{Z} \in \mathbb{R}^d$ *is a Gaussian random vector, then for all* $(k_1, \dots, k_d) \in \mathbb{N}^d$ *we have*

$$\mathbb{E}\left[\prod_{j=1}^d \left|Z_j^{k_j}\right|\right] < \infty.$$

Proof. A simple calculus exercise shows for each $k \geq 0$ that there exists $C_k < \infty$ such that

$$|x|^k \leq C_k e^{|x|} \leq C_k\left[e^x + e^{-x}\right].$$

Hence we know,

$$\mathbb{E}\prod_{j=1}^d \left|Z_j^{k_j}\right| \leq \mathbb{E}\left(\prod_{j=1}^d C_{k_j}\left[e^{Z_j} + e^{-Z_j}\right]\right)$$

and the latter integrand is a linear combination of random variables of the form $e^{a\cdot\mathbf{Z}}$ with $a \in \{\pm 1\}^d$. By assumption $\mathbb{E}\left[e^{a\cdot\mathbf{Z}}\right] < \infty$ for all $a \in \mathbb{R}^d$ and hence we learn that

$$\mathbb{E}\left(\prod_{j=1}^d C_{k_j}\left[e^{Z_j} + e^{-Z_j}\right]\right) < \infty.$$

■

Fact 16.7 *If* $\mathbf{W}$ *and* $\mathbf{Z}$ *are any two* $\mathbb{R}^d$ *– valued random vectors such that* $\mathbb{E}\left[e^{\lambda\cdot\mathbf{Z}}\right] < \infty$ *and* $\mathbb{E}\left[e^{\lambda\cdot\mathbf{W}}\right] < \infty$ *for all* $\lambda \in \mathbb{R}^d$, *then*

$$\mathbb{E}\left[e^{\lambda\cdot\mathbf{Z}}\right] = \mathbb{E}\left[e^{\lambda\cdot\mathbf{W}}\right] \text{ for all } \lambda \in \mathbb{R}^d$$

iff $\mathbf{W} \overset{d}{=} \mathbf{Z}$ *where* $\mathbf{W} \overset{d}{=} \mathbf{Z}$ *by definition means*

$$\mathbb{E}[f(\mathbf{W})] = \mathbb{E}[f(\mathbf{Z})]$$

for all functions, $f : \mathbb{R}^d \to \mathbb{R}$, *such that the expectations makes sense.*

Fact 16.8 *If* $\mathbf{W} \in \mathbb{R}^k$ *and* $\mathbf{Z} \in \mathbb{R}^l$ *are any two random vectors such that* $\mathbb{E}\left[e^{a\cdot\mathbf{W}}\right] < \infty$ *and* $\mathbb{E}\left[e^{b\cdot\mathbf{Z}}\right] < \infty$ *for all* $a \in \mathbb{R}^k$ *and* $b \in \mathbb{R}^l$, *then* $\mathbf{W}$ *and* $\mathbf{Z}$ *are independent iff*

$$\mathbb{E}\left[e^{a\cdot\mathbf{W}}e^{b\cdot\mathbf{Z}}\right] = \mathbb{E}\left[e^{a\cdot\mathbf{W}}\right]\cdot\mathbb{E}\left[e^{b\cdot\mathbf{Z}}\right] \;\; \forall\, a \in \mathbb{R}^k \;\&\; b \in \mathbb{R}^l. \tag{16.4}$$

Example 16.9. Suppose $\mathbf{W} \in \mathbb{R}^k$ and $\mathbf{Z} \in \mathbb{R}^l$ are two random vectors such that $(\mathbf{W}, \mathbf{Z}) \in \mathbb{R}^k \times \mathbb{R}^l$ is a Gaussian random vector. Then $\mathbf{W}$ and $\mathbf{Z}$ are independent iff $\operatorname{Cov}(W_i, Z_j) = 0$ for all $1 \leq i \leq k$ and $1 \leq j \leq l$. To keep the notation to a minimum let us verify this fact when $k = l = 1$. Then according to Fact 16.8 we need only verify Eq. (16.4) for $a, b \in \mathbb{R}$. However,

$$\begin{aligned}\operatorname{Var}((a,b)\cdot(W,Z)) &= \operatorname{Var}(aW + bZ)\\ &= \operatorname{Var}(aW) + \operatorname{Var}(bZ) + 2ab\operatorname{Cov}(W,Z)\\ &= \operatorname{Var}(aW) + \operatorname{Var}(bZ)\end{aligned}$$

and hence

$$\begin{aligned}\mathbb{E}\left[e^{a\cdot W}e^{b\cdot Z}\right] &= \mathbb{E}\left[e^{(a,b)\cdot(W,Z)}\right]\\ &= \exp\left(\frac{1}{2}\operatorname{Var}((a,b)\cdot(W,Z)) + \mathbb{E}[(a,b)\cdot(W,Z)]\right)\\ &= \exp\left(\frac{1}{2}\operatorname{Var}(aW) + \mathbb{E}[aW] + \frac{1}{2}\operatorname{Var}(bZ) + \mathbb{E}[bZ]\right)\\ &= \mathbb{E}\left[e^{a\cdot W}\right]\mathbb{E}\left[e^{b\cdot Z}\right].\end{aligned}$$

Remark 16.10. In general it is not true that two uncorrelated random variables are independent. For example, suppose that X is any random variable with an even distribution, i.e. $X \overset{d}{=} -X$. Let $Y := |X|$ which is typically not independent of X.[2] Nevertheless X and Y are uncorrelated. Indeed ,

[2] For example if $X : \Omega \to \{\pm 1, \pm 2\}$ all with probability 1/4, then

$$\mathbb{P}(X = 1, |X| = 1) = \mathbb{P}(X = 1) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(X = 1)\,\mathbb{P}(|X| = 1).$$

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y = \mathbb{E}[X|X|] - \mathbb{E}X \cdot \mathbb{E}|X| = 0$$

wherein we have used both X and $|X|X$ have even distributions and therefore have zero expectations. Indeed, if $Z \overset{d}{=} -Z$ then

$$\mathbb{E}Z = \mathbb{E}[-Z] = -\mathbb{E}Z \implies \mathbb{E}Z = 0.$$

Exercise 16.1. Suppose that X and Y are **independent** normal random variables. Show;

1. $\mathbf{Z} = (X,Y)$ is a normal random vector, and
2. $W = X + Y$ is a normal random variable.
3. If N is a standard normal random variable and X is any normal random variable, show $X \overset{d}{=} \sigma N + \mu$ where $\mu = \mathbb{E}X$ and $\sigma = \sqrt{\operatorname{Var}(X)}$.

Remark 16.11. 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 $X \overset{d}{=} N(0,1)$ and $\xi \overset{d}{=} Bern(1/2)$ with X and ξ being independent and then let $Y = \xi X$. Then $Y \overset{d}{=} N(0,1)$ so that both X and Y are Gaussian random variables yet (X,Y) is not a Gaussian random vector.

Exercise 16.2. Prove the assertion made in Remark 16.11. **Hints:** compute $\mathbb{E}\left[e^{\lambda Y}\right]$ in order to show $Y \overset{d}{=} N(0,1)$, then show $\operatorname{Cov}(X,Y) = 0$ even though X and Y are not independent.

Definition 16.12 (Gaussian Process). *We say a real valued stochastic process, $\{X_t\}_{t\geq 0}$, is a **Gaussian process** if for all $n \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_n < \infty$, $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian random vector.*

16.2 Stationary and Independent Increment Processes

Rather than considering all continuous time Markov processes with values in $\mathbb{R}$ or $(\mathbb{R}^d)$ we are going to focus out attention on one extremely important example – namely Brownian motion. However, before restricting to the Brownian case let us first introduce a more general class of $\mathbb{R}^d$ – valued stochastic processes.

Definition 16.13. *A stochastic process, $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$, has **independent increments** if for all finite subsets, $\Lambda = \{0 \leq t_0 < t_1 < \cdots < t_n\} \subset T$ the random variables $\{X_0\} \cup \left\{X_{t_k} - X_{t_{k-1}}\right\}_{k=1}^n$ are independent. We refer to $X_t - X_s$ for $s < t$ as an **increment** of X.*

Definition 16.14. *A stochastic process, $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$, has **stationary increments** if for all $0 \leq s < t < \infty$, the law of $X_t - X_s$ depends only on $t - s$.*

Example 16.15. The Poisson process of rate λ, $\{X_t\}_{t\geq 0}$, has stationary independent increments with $X_t - X_s \overset{d}{=} \operatorname{Poi}(\lambda(t-s))$. Also if we let

$$\hat{X}_t := X_t - \mathbb{E}X_t = X_t - \lambda t$$

be the "centering" of X_t, then $\left\{\hat{X}_t\right\}_{t\geq 0}$ is a mean zero process with independent increments. In this case, $\mathbb{E}\hat{X}_t^2 = \operatorname{Var}(X_t) = \lambda t$.

Lemma 16.16. *Suppose that $\{X_t\}_{t\geq 0}$ is a mean zero square-integrable stochastic process with stationary independent increments. If we further assume that $X_0 = 0$ a.s. and $t \to \mathbb{E}X_t^2$ is continuous, then $\mathbb{E}X_t^2 = \lambda t$ for some $\lambda \geq 0$.*

Proof. Let $\varphi(t) := \mathbb{E}X_t^2$ and let $s, t \geq 0$. Then

$$\begin{aligned}
\varphi(t+s) &= \mathbb{E}X_{t+s}^2 = \mathbb{E}\left[(X_{t+s} - X_t) + X_t\right]^2 \\
&= \mathbb{E}(X_{t+s} - X_t)^2 + \mathbb{E}X_t^2 + 2\mathbb{E}\left[(X_{t+s} - X_t)X_t\right] \\
&= \mathbb{E}X_s^2 + \mathbb{E}X_t^2 + 2\mathbb{E}\left[X_{t+s} - X_t\right] \cdot \mathbb{E}X_t \\
&= \varphi(t) + \varphi(s) + 0.
\end{aligned}$$

Now let $\lambda := \varphi(1)$ and use the previous equation to learn, for $n \in \mathbb{N}_0$, that

$$\lambda = \varphi(1) = \varphi\left(\overbrace{\frac{1}{n} + \cdots + \frac{1}{n}}^{n\text{ -times}}\right) = n \cdot \varphi\left(\frac{1}{n}\right),$$

i.e. $\varphi\left(\frac{1}{n}\right) = \frac{1}{n}\lambda$. Similarly if $k \in \mathbb{N}$ we may conclude,

$$\varphi\left(\frac{k}{n}\right) = \varphi\left(\overbrace{\frac{1}{n} + \cdots + \frac{1}{n}}^{k\text{ -times}}\right) = k\varphi\left(\frac{1}{n}\right) = k\frac{1}{n}\lambda = \frac{k}{n}\lambda.$$

Thus $\varphi(t) = \lambda t$ for all $t \in \mathbb{Q}_+$ and then by continuity we must have $\varphi(t) = \lambda t$ for all $t \in \mathbb{R}_+$. ∎

Proposition 16.17. *Suppose that $\left\{X_t : \Omega \to S := \mathbb{R}^d\right\}_{t\in T}$ is a stochastic process with stationary and independent increments and let $\mathcal{F}_t := \mathcal{F}_t^X$ for all $t \in T$. Then*

$$\mathbb{E}\left[f\left(X_t\right)|\mathcal{F}_s\right] = \left(\mathbf{P}_{t-s}f\right)\left(X_s\right) \text{ where} \tag{16.5}$$
$$\left(\mathbf{P}_t f\right)(x) = \mathbb{E}\left[f\left(x + X_t - X_0\right)\right]. \tag{16.6}$$

Moreover, $\{\mathbf{P}_t\}_{t\geq 0}$ satisfies the Chapman – Kolmogorov equation, i.e. $\mathbf{P}_t\mathbf{P}_s = \mathbf{P}_{t+s}$ for all $s, t \geq 0$.

Proof. By assumption we have $X_t - X_s$ is independent of $\mathcal{F}_s$ and therefore by a jazzed up version of Proposition 3.18,

$$\mathbb{E}\left[f\left(X_t\right)|\mathcal{F}_s\right] = \mathbb{E}\left[f\left(X_s + X_t - X_s\right)|\mathcal{F}_s\right] = G\left(X_s\right)$$

where

$$G(x) = \mathbb{E}\left[f\left(x + X_t - X_s\right)\right].$$

Because the increments are stationary we may replace $X_t - X_s$ in the definition of G above by $X_{(t-s)} - X_0$ which then proves Eqs. (16.5) with $\mathbf{P}_t$ as in Eq. (16.6).

To verify the Chapman – Kolmogorov equation we have,

$$\begin{aligned}
\left(\mathbf{P}_t\mathbf{P}_s f\right)(x) &= \mathbb{E}\left[\left(\mathbf{P}_s f\right)\left(x + X_t - X_0\right)\right] \\
&= \mathbb{E}\left[\tilde{\mathbb{E}} f\left(x + X_t - X_0 + \tilde{X}_s - \tilde{X}_0\right)\right] \\
&= \mathbb{E}\left[\tilde{\mathbb{E}} f\left(x + X_t - X_0 + \tilde{X}_{t+s} - \tilde{X}_t\right)\right] \\
&= \mathbb{E}\left[f\left(x + X_t - X_0 + X_{t+s} - X_t\right)\right] \\
&= \mathbb{E}\left[f\left(x + X_{t+s} - X_0\right)\right] = \left(\mathbf{P}_{t+s} f\right)(x).
\end{aligned}$$

■

Example 16.18. The Poisson process (see Definition 15.32) of rate λ, $\{N_t\}_{t\geq 0}$, has stationary independent increments with $N_t - N_s \overset{d}{=} \mathrm{Poi}\left(\lambda\left(t-s\right)\right)$ as was shown in Proposition 15.34. In this case,

$$\left(\mathbf{P}_t f\right)(x) = \mathbb{E} f\left(x + N_t - N_0\right) = \sum_{k=0}^{\infty} f\left(x+k\right)\frac{\left(\lambda t\right)^k}{k!}e^{-\lambda t}$$

and we find

$$\begin{aligned}
\left(\mathbf{A} f\right)(x) &= \frac{d}{dt}|_{0^+}\left(\mathbf{P}_t f\right)(x) = \sum_{k=0}^{\infty} f\left(x+k\right)\frac{d}{dt}|_{0^+}\frac{\left(\lambda t\right)^k}{k!}e^{-\lambda t} \\
&= \sum_{k=0}^{1} f\left(x+k\right)\frac{d}{dt}|_{0^+}\frac{\left(\lambda t\right)^k}{k!}e^{-\lambda t} = -\lambda f(x) + \lambda f\left(x+1\right).
\end{aligned}$$

Proposition 16.19 (Brownian Motion I). *Suppose that $\{B_t\}_{t\geq 0}$ is a real valued stochastic process on some probability space, $(\Omega, \mathcal{B}, P)$ with right continuous sample paths such that;*

1. *$B_0 = 0$ a.s., $\mathbb{E}B_t = 0$ for all $t \geq 0$,*
2. *$\mathbb{E}\left(B_t - B_s\right)^2 = t - s$ 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 $\left\{B_{t_j} - B_{t_{j-1}}\right\}_{j=1}^n$ are independent random variables.*
4. *(Moment Condition) There exists $p > 2$, $q > 1$ and $c < \infty$ such that $\mathbb{E}\left|B_t - B_s\right|^p \leq c\left|t-s\right|^q$ for all $s, t \in \mathbb{R}_+$.*

Then $B_t - B_s \overset{d}{=} N\left(0, t-s\right)$ for all $0 \leq s < t < \infty$ and in fact $\{B_t\}_{t\geq 0}$ is a "Brownian motion" as to be described in Section 16.3 below. [You are asked to show in Exercise 16.3 below that the infinitesimal generator of $\{B_t\}_{t\geq 0}$ is "$\mathbf{A} = \frac{1}{2}\frac{d^2}{dx^2}$."]

The previous result is a consequence of the central limit theorem. As the next example illustrates, the moment condition above is needed in order to show $\{B_t\}_{t\geq 0}$ is a Gaussian process as defined in Definition 16.12.

Example 16.20 (Normalized Poisson Process). There certainly are other right continuous processes satisfying items 1.-3. but **not** the moment condition 4. that are not Gaussian. Indeed, if $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ (see Definition 15.32), then $B_t := \lambda^{-1}N_t - t$ satisfies the items 1.–3. above. The independence of the increments follows from Proposition 15.34 and from Exercise 15.1, we know that $\mathrm{Var}\left(N_t - N_s\right) = \left(\lambda t - \lambda s\right)^2$ and $\mathbb{E}\left(N_t - N_s\right) = \lambda\left(t-s\right)$ so $\mathbb{E}B_t = 0$ and

$$\begin{aligned}
\mathbb{E}\left(B_t - B_s\right)^2 &= \mathrm{Var}\left(B_t - B_s\right) = \mathrm{Var}\left(\lambda^{-1}\left(N_t - N_s\right)\right) \\
&= \lambda^{-2}\,\mathrm{Var}\left(N_t - N_s\right) = \lambda^{-2}\left(\lambda t - \lambda s\right)^2 = t - s.
\end{aligned}$$

In this case one can show that $\mathbb{E}\left[\left|B_t - B_s\right|^p\right] \sim \left|t-s\right|$ for all $1 \leq p < \infty$ and the moment condition in item 4. fails. This is good as $\{B_t\}_{t\geq 0}$ is not Gaussian!

16.3 Brownian motion defined

Definition 16.21 (Brownian Motion). *Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{R}_+}, \mathbb{P}\right)$ be a filtered probability space. A real valued adapted process, $\{B_t : \Omega \to \mathbb{R}\}_{t\in\mathbb{R}_+}$, 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 $\mathcal{F}_s$ for all $0 \leq s < t < \infty$.*

2. *for* $0 \le s < t$, $B_t - B_s \overset{d}{=} N(0, t-s)$, *i.e.* $B_t - B_s$ *is a normal mean zero random variable with variance* $(t-s)$,
3. $t \to B_t(\omega)$ *is continuous for all* $\omega \in \Omega$.

Remark 16.22. In light of Lemma 16.16, we could have described a Brownian motion (up to a scale) as a continuous in time stochastic process $\{B_t\}_{t\ge 0}$ with stationary independent mean zero Gaussian increments such that $B_0 = 0$ and $t \to \mathbb{E}B_t^2$ is continuous.

Exercise 16.3 (Brownian Motion). Let $\{B_t\}_{t\ge 0}$ be a Brownian motion as in Definition 16.21.

1. Explain why $\{B_t\}_{t\ge 0}$ is a time homogeneous Markov process with transition operator,
$$(\mathbf{P}_t f)(x) = \int_{\mathbb{R}} p_t(x,y) f(y)\, dy \tag{16.7}$$
where
$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|y-x|^2}. \tag{16.8}$$
In more detail use Proposition 16.17 and Eq. (16.10) below to argue,
$$\mathbb{E}[f(B_t)|\mathcal{F}_s] = (\mathbf{P}_{t-s} f)(B_s). \tag{16.9}$$
2. Show by direct computation that $\mathbf{P}_t \mathbf{P}_s = \mathbf{P}_{t+s}$ for all $s, t > 0$. **Hint:** probably the easiest way to do this is to make use of Exercise 16.1 along with the identity,
$$(\mathbf{P}_t f)(x) = \mathbb{E}\left[f\left(x + \sqrt{t}Z\right)\right], \tag{16.10}$$
where $Z \overset{d}{=} N(0,1)$.
3. Show by direct computation that $p_t(x,y)$ of Eq. (16.8) satisfies the **heat equation**,
$$\frac{d}{dt} p_t(x,y) = \frac{1}{2}\frac{d^2}{dx^2} p_t(x,y) = \frac{1}{2}\frac{d^2}{dy^2} p_t(x,y) \text{ for } t > 0.$$
4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function with compact support. Show
$$\frac{d}{dt}\mathbf{P}_t f = \mathbf{A}\mathbf{P}_t f = \mathbf{P}_t \mathbf{A} f \text{ for all } t > 0,$$
where
$$\mathbf{A} f(x) = \frac{1}{2} f''(x).$$
Note: you may use (under the above assumptions) without proof the fact that it permissible to interchange the $\frac{d}{dt}$ and $\frac{d}{dx}$ derivatives with the integral in Eq. (16.7).

Modulo technical details, Exercise 16.3 shows that $\mathbf{A} = \frac{1}{2}\frac{d^2}{dx^2}$ is the infinitesimal generator of Brownian motion, i.e. the infinitesimal generator of $\mathbf{P}_t$ in Eq. (16.7). 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 omit these delicate issues here. However, let me at least point out that it is no longer necessarily a good idea to try to recover $\mathbf{P}_t$ as $\sum_{n=0}^\infty \frac{t^n}{n!} A^n$ in the Brownian motion setting. Indeed, $\sum_{n=0}^\infty \frac{t^n}{n!} A^n f$ to make sense one needs to assume that, at very least, f is C^∞ and even this will not guarantee convergence of the sum! Nevertheless, when $\mathbf{A} = \frac{1}{2}\frac{d^2}{dx^2}$ it is true that

$$\mathbf{P}_t f = \sum_{n=0}^\infty \frac{t^n}{n!} \mathbf{A}^n f = \sum_{n=0}^\infty \frac{t^n}{2^n n!} f^{(2n)}$$

when if f is a polynomial function, see Exercise 19.2. The infinite series expansion also gives the correct answer when applied to exponential functions, $f(x) = e^{\lambda x}$, where $\lambda \in \mathbb{C}$.

Corollary 16.23. *If* $0 = t_0 < t_1 < \cdots < t_n < \infty$, $\Delta_i t := t_i - t_{i-1}$, *then for*

$$\mathbb{E}F(B_{t_1}, \dots, B_{t_n}) = \int F(x_1, \dots, x_n)\, p_{\Delta_1 t}(x_0, x_1) \dots p_{\Delta_n t}(x_{n-1}, x_n)\, dx_1 \dots dx_n \tag{16.11}$$

for all bounded or non-negative functions $F : \mathbb{R}^n \to \mathbb{R}$. *In particular if* $J_i = (a_i, b_i) \subset \mathbb{R}$ *are given bounded intervals, then*

$$\begin{aligned}&\mathbb{P}(B(t_i) \in J_i \text{ for } i = 1, 2, \dots, n)\\ &= \int \dots \int_{J_1 \times \cdots \times J_n} p_{\Delta_1 t}(0, x_1)\, p_{\Delta_2 t}(x_1, x_2) \dots p_{\Delta_n t}(x_{n-1}, x_n)\, dx_1 \dots dx_n.\end{aligned} \tag{16.12}$$

as follows from Eq. (16.11) by taking $F(x_1, \dots, x_n) := 1_{J_1}(x_1) \dots 1_{J_n}(x_n)$.

Proof. This result is a formal consequence of the Markov property (Proposition 16.17) similar to the discrete space case in Theorem 5.11. Rather than carry out this style of proof here I will give a proof based directly on the independent increments of the Brownian motion. Let $x_0 := 0$. We are going to prove Eq. (16.11) by induction on n.

For $n = 1$, we have

$$\mathbb{E}F(B_{t_1}) = \mathbb{E}F\left(\sqrt{t_1}N\right) = \int_{\mathbb{R}} p_{t_1}(0, y) f(y)\, dy$$

which is Eq. (16.11) with $n = 1$. For the induction step, let N be a standard normal random variable independent of $\left(B_{t_1}, \dots, B_{t_{n-1}}\right)$, then

$$\begin{aligned}\mathbb{E}F\left(B_{t_1},\ldots,B_{t_n}\right) &= \mathbb{E}F\left(B_{t_1},\ldots,B_{t_{n-1}},B_{t_{n-1}}+B_{t_n}-B_{t_{n-1}}\right)\\ &= \mathbb{E}F\left(B_{t_1},\ldots,B_{t_{n-1}},B_{t_{n-1}}+\sqrt{\Delta_n t}N\right)\\ &= \mathbb{E}\int_{\mathbb{R}} F\left(B_{t_1},\ldots,B_{t_{n-1}},y\right)p_{\Delta_n t}\left(B_{t_{n-1}},y\right)dy\\ &= \int_{\mathbb{R}}\mathbb{E}\left[F\left(B_{t_1},\ldots,B_{t_{n-1}},y\right)p_{\Delta_n t}\left(B_{t_{n-1}},y\right)\right]dy. \quad (16.13)\end{aligned}$$

By the induction hypothesis,

$$\begin{aligned}&\mathbb{E}\left[F\left(B_{t_1},\ldots,B_{t_{n-1}},y\right)p_{\Delta_n t}\left(B_{t_{n-1}},y\right)\right]\\ &= \int F\left(x_1,\ldots,x_{n-1},y\right)\left[\begin{array}{c} p_{\Delta_1 t}\left(x_0,x_1\right)\cdot\ldots\\ \cdot p_{\Delta_{n-1}t}\left(x_{n-2},x_{n-1}\right)p_{\Delta_n t}\left(x_{n-1},y\right)\end{array}\right]dx_1\ldots dx_{n-1}dy. \quad (16.14)\end{aligned}$$

Combining Eqs. (16.13) and (16.14) and then replacing y by x_n verifies Eq. (16.11). ■

Remark 16.24 (Gaussian aspects of Brownian Motion). Brownian motion is a Gaussian process (see Definition 16.12). Indeed, if $0=t_0<t_1<\cdots<t_n<\infty$, then $(B_{t_1},\ldots,B_{t_n})$ is a Gaussian random vector since it is a linear transformation of the increment vector, $(\Delta_1 B,\ldots,\Delta_n B)$ where $\Delta_i B := B_{t_i}-B_{t_{i-1}}$. The distribution of mean zero Gaussian random vectors is completely determined by its covariance matrix. In this case if $0\le s<t<\infty$, we have

$$\mathbb{E}\left[B_s B_t\right] = \mathbb{E}\left[B_s\left[B_s+B_t-B_s\right]\right] = \mathbb{E}B_s^2+\mathbb{E}B_s\cdot\mathbb{E}\left[B_t-B_s\right] = s.$$

In general we have

$$\mathbb{E}\left[B_s B_t\right] = \min\left(s,t\right)\ \forall\ s,t\ge 0. \quad (16.15)$$

Thus we could define Brownian motion $\{B_t\}_{t\ge 0}$ to be the mean-zero continuos Gaussian process such that Eq. (16.15) holds.

Indeed, if $\{B_t\}_{t\ge 0}$ to be the mean-zero continuos Gaussian process such that Eq. (16.15) holds then for $0<t_1<\cdots<t_n=s<t$, we will have

$$\mathbb{E}\left[\left(B_t-B_s\right)B_{t_i}\right] = t\wedge t_i - s\wedge t_i = t_i-t_i = 0$$

for all $1\le i\le n$. Since $(B_{t_1},\ldots,B_{t_n},B_t-B_s)$ is a Gaussian random vector we may conclude that B_t-B_s is independent of $(B_{t_1},\ldots,B_{t_n})$. As the $\{t_i\}$ as above were arbitrary we conclude that B_t-B_s is independent of $\mathcal{F}_s^B=\sigma\left(B_r: r\le s\right)$. Therefore if $f:\mathbb{R}\to\mathbb{R}$ is a bounded measurable function we have,

$$\mathbb{E}\left[f\left(B_t\right)|\mathcal{F}_s\right] = \mathbb{E}\left[f\left(B_t-B_s+B_s\right)|\mathcal{F}_s\right] = G\left(B_s\right)$$

where

$$\begin{aligned}G\left(x\right) &:= \mathbb{E}\left[f\left(B_t-B_s+x\right)\right] = \mathbb{E}\left[f\left(x+\sqrt{t-s}N\right)\right]\\ &= \int_{\mathbb{R}} p_{t-s}\left(x,y\right)G\left(y\right)dy = \left(\mathbf{P}_t f\right)\left(x\right),\end{aligned}$$

where $N\overset{d}{=}N\left(0,1\right)$. Thus we have shown

$$\mathbb{E}\left[f\left(B_t\right)|\mathcal{F}_s\right] = \left(\mathbf{P}_t f\right)\left(B_s\right).$$

It is now not to difficult to verify Eq. (16.11) holds.

Remark 16.25 (White Noise). If we **formally** differentiate Eq. (16.15) with respect to s and and t we find,

$$\mathbb{E}\left[\dot{B}_s\dot{B}_t\right] = \frac{d}{dt}\frac{d}{ds}\min\left(s,t\right) = \frac{d}{dt}\left[1_{s\le t}\right] = \delta\left(t-s\right). \quad (16.16)$$

Thus $\left\{\dot{B}_t\right\}_{t\ge 0}$ is a totally uncorrelated Gaussian process which formally implies $\left\{\dot{B}_t\right\}_{t\ge 0}$ are all independent "random variables" which is referred to as **white noise.**

Warning: white noise is not an honest real-valued random stochastic process owing to the fact that Brownian motion is very rough and in particular nowhere differentiable. If $\dot{B}_t$ were to exists it should be a Gaussian and hence satisfy $\mathbb{E}\dot{B}_t^2<\infty$. On the other hand from Eq. (16.16)with $s=t$ we are lead to believe that

$$\mathbb{E}\left[\dot{B}_t^2\right] = \delta\left(t-t\right) = \delta\left(0\right) = \infty.$$

Nevertheless, ignoring these issues, it is common to model noise in a system by such a white noise. For example we may have $y\left(t\right)\in\mathbb{R}^d$ satisfies, in a pristine environment, a differential equation of the form

$$\dot{y}\left(t\right) = f\left(t,y\left(t\right)\right). \quad (16.17)$$

However, the real world is not so pristine and the trajectory $y\left(t\right)$ is also influenced by noise in the system which is often modeled by adding a term of the form $g\left(t,y\left(t\right)\right)\dot{B}_t$ to the rights side of Eq. (16.17) to arrive at the " **stochastic differential equation,"**

$$\dot{y}\left(t\right) = f\left(t,y\left(t\right)\right)+g\left(t,y\left(t\right)\right)\dot{B}_t.$$

In the last section of these notes we will begin to explain Itô's method for making sense of such equations.

Definition 16.26 (Multi-Dimensional Brownian Motion). *For $d\in\mathbb{N}$, we say a $\mathbb{R}^d$ – valued process, $\left\{B_t=\left(B_t^1,\ldots,B_t^d\right)^{\mathrm{tr}}\right\}_{t\ge 0}$ is a d – **dimensional Brownian motion** provided $\left\{B_\cdot^i\right\}_{i=1}^d$ is an independent collection of one dimensional Brownian motions.*

We state the following multi-dimensional version of the results argued above. The proof of this theorem will be left to the reader.

Theorem 16.27. *A d – dimensional Brownian motion, $\{B_t\}_{t\geq 0}$, is a Markov process with transition semi-group, $\mathbf{P}_t$, now defined by*

$$(\mathbf{P}_t f)(x) = \mathbb{E}\left[f\left(x+\sqrt{t}Z\right)\right] = \int_{\mathbb{R}^d} p_t(x,y) f(y)\, dy$$

where now Z is a standard normal random vector in $\mathbb{R}^d$ and

$$p_t(x,y) = \left(\frac{1}{\sqrt{2\pi t}}\right)^d e^{-\frac{1}{2t}\|x-y\|^2}.$$

Moreover we have

$$\frac{d}{dt}\mathbf{P}_t = \frac{1}{2}\Delta\mathbf{P}_t = \mathbf{P}_t\frac{1}{2}\Delta$$

where now $\Delta = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$ is the d – dimensional Laplacian.

Remark 16.28. Add here comments about the heuristic path integral interpretation of Brownian motion, namely that

$$\mathbb{E}f(B) \text{ “=” } \frac{1}{Z}\int_{C(\mathbb{R}_+,\mathbb{R}^d)} f(\omega)\, e^{-\frac{1}{2}\int_0^\infty |\dot\omega(t)|^2 dt}\mathcal{D}\omega. \tag{16.18}$$

Also mention that

$$\left(\left(-\frac{d^2}{dt^2}\right)^{-1} h\right)(t) = \int_0^\infty t\wedge s\, h(s)\, ds,$$

that is $t\wedge s$ is the Green's function associate to the formal covariance matrix of the "Gaussian" measure in Eq. (16.18).

16.4 Some "Brownian" martingales

Definition 16.29 (Continuous time martnigales). *Given a filtered probability space, $\left(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq 0},\mathbb{P}\right)$, an adapted process, $X_t:\Omega\to\mathbb{R}$, is said to be a $(\{\mathcal{F}_t\})$* ***martingale*** *provided, $\mathbb{E}|X_t|<\infty$ for all t and*

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = 0 \text{ for all } 0\leq s\leq t<\infty.$$

If

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s]\geq 0 \text{ or } \mathbb{E}[X_t - X_s|\mathcal{F}_s]\leq 0 \text{ for all } 0\leq s\leq t<\infty,$$

then X is said to be ***submartingale or supermartingale*** *respectively.*

Notation 16.30 *Given a function, $h:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ such that $\dot h(t,x) = \frac{\partial}{\partial t}h(t,x)$, $\nabla_x h(t,x)$, and $\Delta_x h(t,x)$ all exist, let*

$$\mathcal{L}h(t,x) := \partial_t h(t,x) + \frac{1}{2}\Delta_x h(t,x).$$

Theorem 16.31 (Martingale problem). *Suppose that $h:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ is a continuous function such that $\dot h(t,x) = \frac{\partial}{\partial t}h(t,x)$, $\nabla_x h(t,x)$, and $\Delta_x h(t,x)$ exist and are continuous on $[0,T]\times\mathbb{R}^d$ and satisfy*

$$\sup_{0\leq t\leq T}\mathbb{E}_\nu\left[|h(t,B_t)| + \left|\dot h(t,B_t)\right| + |\nabla_x h(t,B_t)| + |\Delta_x h(t,B_t)|\right] < \infty. \tag{16.19}$$

Then the process,

$$M_t := h(t,B_t) - \int_0^t \mathcal{L}h(\tau,B_\tau)\, d\tau \tag{16.20}$$

is a $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$ – martingale. In particular, if h also satisfies the heat equation in reverse time,

$$\mathcal{L}h(t,x) := \partial_t h(t,x) + \frac{1}{2}\Delta_x h(t,x) = 0, \tag{16.21}$$

then $M_t = h(t,B_t)$ is a martingale.

Proof. Working formally for the moment,

$$\begin{aligned}\frac{d}{d\tau}\mathbb{E}_\nu[h(\tau,B_\tau)|\mathcal{F}_s] &= \frac{d}{d\tau}[\mathbf{P}_{\tau-s}h(\tau,B_s)]\\ &= \mathbf{P}_{\tau-s}\dot h(\tau,B_s) + \frac{1}{2}\mathbf{P}_{\tau-s}\Delta h(\tau,B_s)\\ &= \mathbb{E}_\nu\left[\dot h(\tau,B_\tau) + \frac{1}{2}\Delta h(\tau,B_\tau)|\mathcal{F}_s\right]\\ &= \mathbb{E}_\nu[\mathcal{L}h(\tau,B_\tau)|\mathcal{F}_s]\end{aligned}$$

wherein we have used the Markov property in the first and third lines and the chain rule in the second line. Integrating this equation on $\tau\in[s,t]$ then shows

$$\mathbb{E}_\nu[h(t,B_t) - h(s,B_s)|\mathcal{F}_s] = \mathbb{E}_\nu\left[\int_s^t \mathcal{L}h(\tau,B_\tau)\, d\tau|\mathcal{F}_s\right]. \tag{16.22}$$

Using

$$\mathbb{E}_\nu[M_t - M_s|\mathcal{F}_s] = \mathbb{E}_\nu\left[h(t,B_t) - h(s,B_s) - \int_s^t \mathcal{L}h(\tau,B_\tau)\, d\tau|\mathcal{F}_s\right]$$

we see that Eq. (16.22) is precisely the statement that $\mathbb{E}_\nu[M_t - M_s|\mathcal{F}_s] = 0$, i.e. to the assertion that $\{M_t\}_{t\geq 0}$ is a martingale. *(I would suggest omitting the more technical details to follow upon first reading.)

We now need to justify the above computations.

1. Let us first suppose there exists an $R < \infty$ such that $h(t,x) = 0$ if $|x| = \sqrt{\sum_{i=1}^d x_i^2} \geq R$. By a couple of integration by parts, we find

$$\begin{aligned}\frac{d}{d\tau}(\mathbf{P}_{\tau-s}h)(\tau,x) &= \int_{\mathbb{R}^d}\frac{d}{d\tau}\left[p_{\tau-s}(x-y)\,h(\tau,y)\right]dy \\ &= \int_{\mathbb{R}^d}\left[\frac{1}{2}(\Delta p_{\tau-s})(x-y)\,h(\tau,y) + p_{\tau-s}(x-y)\,\dot{h}(\tau,y)\right]dy \\ &= \int_{\mathbb{R}^d}\left[\frac{1}{2}p_{\tau-s}(x-y)\,\Delta_y h(\tau,y) + p_{\tau-s}(x-y)\,\dot{h}(\tau,y)\right]dy \\ &= \int_{\mathbb{R}^d} p_{\tau-s}(x-y)\,\mathcal{L}h(\tau,y)\,dy =: (\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,x), \quad (16.23)\end{aligned}$$

where

$$\mathcal{L}h(\tau,y) := \dot{h}(\tau,y) + \frac{1}{2}\Delta_y h(\tau,y). \quad (16.24)$$

Since

$$(\mathbf{P}_{\tau-s}h)(\tau,x) := \int_{\mathbb{R}^d} p_{\tau-s}(x-y)\,h(\tau,y)\,dy = \mathbb{E}_\nu\left[h\left(\tau, x+\sqrt{\tau-s}N\right)\right]$$

where $N \overset{d}{=} N(0, I_{d\times d})$, we see that $(\mathbf{P}_{\tau-s}h)(\tau,x)$ (and similarly that $\mathcal{L}h(\tau,x)$) is a continuous function in (τ,x) for $\tau \geq s$. Moreover, $(\mathbf{P}_{\tau-s}h)(\tau,x)|_{\tau=s} = h(s,x)$ and so by the fundamental theorem of calculus (using Eq. (16.23

$$(\mathbf{P}_{\tau-s}h)(\tau,x) = h(s,x) + \int_s^t (\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,x)\,d\tau. \quad (16.25)$$

Hence for $A \in \mathcal{F}_s$,

$$\begin{aligned}\mathbb{E}_\nu[h(t,B_t):A] &= \mathbb{E}_\nu[\mathbb{E}_\nu[h(t,B_t)|\mathcal{F}_s]:A] \\ &= \mathbb{E}_\nu[(\mathbf{P}_{\tau-s}h)(\tau,B_s):A] \\ &= \mathbb{E}_\nu\left[h(s,B_s) + \int_s^t (\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,B_s)\,d\tau : A\right] \\ &= \mathbb{E}_\nu[h(s,B_s):A] + \int_s^t \mathbb{E}_\nu[(\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,B_s):A]\,d\tau.\end{aligned} \quad (16.26)$$

Since

$$(\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,B_s) = \mathbb{E}_\nu[(\mathcal{L}h)(\tau,B_\tau)|\mathcal{F}_s],$$

we have

$$\mathbb{E}_\nu[(\mathbf{P}_{\tau-s}\mathcal{L}h)(\tau,B_s):A] = \mathbb{E}_\nu[(\mathcal{L}h)(\tau,B_\tau):A]$$

which combined with Eq. (16.26) shows,

$$\begin{aligned}\mathbb{E}_\nu[h(t,B_t):A] &= \mathbb{E}_\nu[h(s,B_s):A] + \int_s^t \mathbb{E}_\nu[(\mathcal{L}h)(\tau,B_\tau):A]\,d\tau \\ &= \mathbb{E}_\nu\left[h(s,B_s) + \int_s^t (\mathcal{L}h)(\tau,B_\tau)\,d\tau : A\right].\end{aligned}$$

This proves $\{M_t\}_{t\geq 0}$ is a martingale when $h(t,x) = 0$ if $|x| \geq R$.

2. For the general case, let $\varphi \in C_c^\infty(\mathbb{R}^d,[0,1])$ such that $\varphi = 1$ in a neighborhood of $0 \in \mathbb{R}^d$ and for $n \in \mathbb{N}$, let $\varphi_n(x) := \varphi(x/n)$. Observe that $\varphi_n \to 1$, $\nabla\varphi_n(x) = \frac{1}{n}(\nabla\varphi)(x/n)$, and $\Delta\varphi_n(x) = \frac{1}{n^2}(\Delta\varphi)(x/n)$ all go to zero boundedly as $n \to \infty$. Applying case 1. to $h_n(t,x) = \varphi_n(x)\,h(t,x)$ we find that

$$M_t^n := \varphi_n(B_t)\,h(t,B_t) - \int_0^t \varphi_n(B_\tau)\left[\dot{h}(\tau,B_\tau) + \frac{1}{2}\Delta h(\tau,B_\tau)\right]d\tau + \varepsilon_n(t)$$

is a martingale, where

$$\varepsilon_n(t) := \int_0^t \left[\frac{1}{2}\Delta\varphi_n(B_\tau)\,h(\tau,B_\tau) + \nabla\varphi_n(B_\tau)\cdot\nabla h(\tau,B_\tau)\right]d\tau.$$

By DCT,

$$\begin{aligned}\mathbb{E}_\nu|\varepsilon_n(t)| \leq &\int_0^t \frac{1}{2}\mathbb{E}_\nu|\Delta\varphi_n(B_\tau)\,h(\tau,B_\tau)|\,d\tau \\ &+ \int_0^t \mathbb{E}_\nu[|\nabla\varphi_n(B_\tau)|\,|\nabla h(\tau,B_\tau)|]\,d\tau \to 0 \text{ as } n\to\infty,\end{aligned}$$

and similarly,

$$\mathbb{E}_\nu|\varphi_n(B_t)\,h(t,B_t) - h(t,B_t)| \to 0,$$

$$\int_0^t \mathbb{E}_\nu\left|\varphi_n(B_\tau)\,\dot{h}(\tau,B_\tau) - \dot{h}(\tau,B_\tau)\right|d\tau \to 0, \text{ and}$$

$$\frac{1}{2}\int_0^t \mathbb{E}_\nu|\varphi_n(B_\tau)\,\Delta h(\tau,B_\tau) - \Delta h(\tau,B_\tau)|\,d\tau \to 0$$

as $n \to \infty$. From these comments one easily sees that $\mathbb{E}_\nu|M_t^n - M_t| \to 0$ as $n \to \infty$ which is sufficient to show $\{M_t\}_{t\in[0,T]}$ is still a martingale. ■

Rewriting Eq. (16.20) we have

$$h(t,B_t) = M_t + \int_0^t \left[\dot{h}(\tau,B_\tau) + \frac{1}{2}\Delta_x h(\tau,B_\tau)\right]d\tau.$$

Example 16.32. Considering the simplest case where $d = 1$ and $h(t,x) = f(x)$, Theorem 16.31 asserts that

$$f(B_t) = M_t + \frac{1}{2}\int_0^t f''(B_t)\,dt$$

where M_t is a martingale provided,

$$\sup_{0\le t\le T} \mathbb{E}_\nu\left[|f(B_t)| + |f'(B_t)| + |f''(B_t)|\right] < \infty.$$

Hence;

1. If $f'' \ge 0$ (i.e. f is subharmonic[3]) then $f(B_t)$ is a submartingale.
2. If $f'' \le 0$ (i.e. f is super harmonic) then $f(B_t)$ is a supermartingale.
3. If $f'' = 0$ (i.e. f is harmonic) then $f(B_t)$ is a martingale.

More generally, if $f : \mathbb{R}^d \to \mathbb{R}$ is a C^2 function, then $f(B_t)$ is a "local" submartingale (supermartingale) iff $\Delta f \ge 0$ ($\Delta f \le 0$).

Exercise 16.4 (h – transforms of B_t). Let $\{B_t\}_{t\in\mathbb{R}_+}$ be a d – dimensional Brownian motion. Show the following processes are martingales;

1. $M_t = u(B_t)$ where $u : \mathbb{R}^d \to \mathbb{R}$ is a Harmonic function, $\Delta u = 0$, such that $\sup_{t\le T}\mathbb{E}\left[|u(B_t)| + |\nabla u(B_t)|\right] < \infty$ for all $T < \infty$.
2. $M_t = \lambda\cdot B_t$ for all $\lambda \in \mathbb{R}^d$.
3. $M_t = e^{-\lambda Y_t}\cos(\lambda X_t)$ and $M_t = e^{-\lambda Y_t}\sin(\lambda X_t)$ where $B_t = (X_t, Y_t)$ is a two dimensional Brownian motion and $\lambda \in \mathbb{R}$.
4. $M_t = |B_t|^2 - d\cdot t$.
5. $M_t = (a\cdot B_t)(b\cdot B_t) - (a\cdot b)\,t$ for all $a, b \in \mathbb{R}^d$.
6. $M_t := e^{\lambda\cdot B_t - |\lambda|^2 t/2}$ for any $\lambda \in \mathbb{R}^d$.

Corollary 16.33 (Compensators). *Suppose $h : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ is a C^2 – function such that*

$$\mathcal{L}h = \partial_t h(t,x) + \frac{1}{2}\Delta_x h(t,x) = 0$$

and both h and h^2 satisfy the hypothesis of Theorem 16.31. If we let M_t denote the martingale, $M_t := h(t, B_t)$ and

$$A_t := \int_0^t |(\nabla_x h)(\tau, B_\tau)|^2\,d\tau,$$

then $N_t := M_t^2 - A_t$ is a martingale. Thus the submartingale has the "Doob" decomposition,

$$M_t^2 = N_t + A_t.$$

and we call the increasing process, A_t, the ***compensator to*** *M_t^2.*

Proof. We need only apply Theorem 16.31 to h^2. In order to do this we need to compute $\mathcal{L}h^2$. Since

$$\partial_i^2 h^2 = \partial_i(2h\cdot\partial_i h) = 2(\partial_i h)^2 + 2h\partial_i h$$

we see that $\frac{1}{2}\Delta h^2 = h\Delta h + |\nabla h|^2$. Therefore,

$$\mathcal{L}h^2 = 2h\dot{h} + h\Delta h + |\nabla h|^2 = |\nabla h|^2$$

and hence the proposition follows form Eq. (16.20) with h replaced by h^2. ■

Exercise 16.5 (Compensators). Let $\{B_t\}_{t\in\mathbb{R}_+}$ be a d – dimensional Brownian motion. Find the compensator, A_t, for each of the following square integrable martingales.

1. $M_t = u(B_t)$ where $u : \mathbb{R}^d \to \mathbb{R}$ is a harmonic function, $\Delta u = 0$, such that

$$\sup_{t\le T}\mathbb{E}\left[\left|u^2(B_t)\right| + |\nabla u(B_t)|^2\right] < \infty \text{ for all } T < \infty.$$

Verify the hypothesis of Corollary 16.33 to show

$$A_t = \int_0^t |\nabla u(B_\tau)|^2\,d\tau.$$

2. $M_t = \lambda\cdot B_t$ for all $\lambda \in \mathbb{R}^d$.
3. $M_t = |B_t|^2 - d\cdot t$.
4. $M_t := e^{\lambda\cdot B_t - |\lambda|^2 t/2}$ for any $\lambda \in \mathbb{R}^d$.

Fact 16.34 (Lévy's criteria for BM) *P. Lévy gives another description of Brownian motion, namely a continuous stochastic process, $\{X_t\}_{t\ge 0}$, is a Brownian motion if $X_0 = 0$, $\{X_t\}_{t\ge 0}$ is a martingale, and $\left\{X_t^2 - t\right\}_{t\ge 0}$ is a martingale.*

16.5 Optional Sampling Results

Similar to Definition 15.40, we have the informal definition of a stopping time in the Brownian motion setting.

[3] A function f is subharmonic if for any compact region, U, $f \le g$ on U where g is the harmonic function on U such that $g = f$ on ∂U. This turns out to be equivalent to assuming $\Delta f \ge 0$.

Definition 16.35 (Informal). *A stopping time, T, for $\{B_t\}_{t\geq 0}$, is a random variable with the property that the event $\{T \leq t\}$ is determined from the knowledge of $\{B_s : 0 \leq s \leq t\}$. Alternatively put, for each $t \geq 0$, there is a functional, f_t, such that*

$$1_{T\leq t} = f_t(\{B_s : 0 \leq s \leq t\}).$$

In this continuous context the optional sampling theorem is as follows.

Theorem 16.36 (Continuous time optional sampling theorem). *If $\{M_t\}_{t\geq 0}$ is a $\mathcal{F}_t^B$ – martingale and σ and τ are two stopping times such that there $\tau \leq K$ for some non-random finite constant $K < \infty$, then $M_\tau \in L^1(\Omega, \mathcal{F}_\tau, P)$, $M_{\sigma\wedge\tau} \in L^1(\Omega, \mathcal{F}_{\sigma\wedge\tau}, P)$ and*

$$M_{\sigma\wedge\tau} = \mathbb{E}[M_\tau | \mathcal{F}_\sigma]. \tag{16.27}$$

Remark 16.37. In applying the optional sampling Theorem 16.36, observe that if τ is any **optional** time and $K < \infty$ is a constant, then $\tau \wedge K$ is a bounded optional time. Indeed,

$$\{\tau \wedge K < t\} = \begin{cases} \{\tau < t\} \in \mathcal{F}_t & \text{if } t \leq K \\ \Omega \in \mathcal{F}_t & \text{if } t > K. \end{cases}$$

Therefore if σ and τ are any optional times with $\sigma \leq \tau$, we may apply Theorem 16.36 with σ and τ replaced by $\sigma \wedge K$ and $\tau \wedge K$. One may then try to pass to limit as $K \uparrow \infty$ in the resulting identity.

For the next few results let $d = 1$ and for any $y \in \mathbb{R}$ let

$$\tau_y := \inf\{t > 0 : B_t = y\}. \tag{16.28}$$

Let us now fix $a, b \in \mathbb{R}$ so that $-\infty < a < 0 < b < \infty$ and set $\tau := \tau_a \wedge \tau_b$ be the first exit time the Brownian motion from the interval, (a, b). By the law of large numbers for Brownian motions, see Corollary 16.43 below, we may deduce that $\mathbb{P}_0(\tau < \infty) = 1$. This may also be deduced from Lemma 16.56 – but this is using a rather large hammer to conclude a simple result. We will give an independent proof here as well.

Proposition 16.38. *With the notation above we have,*

$$\mathbb{E}\tau = \mathbb{E}[\tau_a \wedge \tau_b] = -ab = |a|\, b < \infty, \tag{16.29}$$

$$\mathbb{P}(\tau_a \wedge \tau_b = \infty) = 0, \text{ and} \tag{16.30}$$

$$\mathbb{E}\tau_b = \infty. \tag{16.31}$$

Proof. Let $\tau := \tau_a \wedge \tau_b$ and $u(x) = (x-a)(b-x)$, see Figure 16.1. Then $\mathcal{L}u = \frac{1}{2}u'' = -1$ and so by Theorem 16.31 (or by Itô's lemma in Theorem 17.5 below) we know that

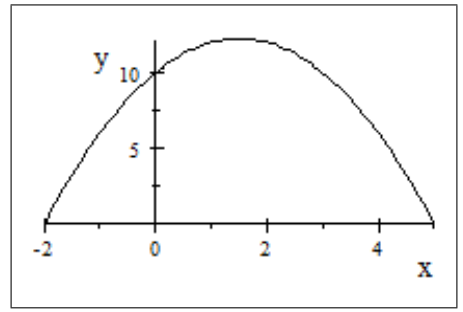


Fig. 16.1. A plot of f for $a = -2$ and $b = 5$.

$$M_t = u(B_t) - \int_0^t \mathcal{L}u(B_s)\, ds = u(B_t) + t$$

is a martingale. So by the optional sampling Theorem 16.36 we may conclude that

$$\mathbb{E}u(B_{t\wedge\tau}) + \mathbb{E}[\tau \wedge t] = \mathbb{E}M_{\tau\wedge t} = \mathbb{E}M_0 = u(B_0) = -ab. \tag{16.32}$$

As $u \geq 0$ on $[a, b]$ and $B_{t\wedge\tau} \in [a, b]$, it follows that

$$|a|\, b = \mathbb{E}[\tau \wedge t] + \mathbb{E}u(B_{t\wedge\tau}) \geq \mathbb{E}[\tau \wedge t].$$

We may now use the monotone convergence theorem (see Section 1.1) to conclude,

$$\mathbb{E}\tau = \lim_{t\uparrow\infty} \mathbb{E}[\tau \wedge t] \leq |a|\, b < \infty.$$

Next using the facts; 1) τ is integrable and $t\wedge\tau \to \tau$ as $t \uparrow \infty$ 2) u is bounded on $[a, b]$ and $u(B_{t\wedge\tau}) \to u(B_\tau) = 0$ as $t \uparrow \infty$, we may apply the dominated convergence theorem to pass to the limit in Eq. (16.32) to find,

$$-ab = \lim_{t\uparrow\infty} [\mathbb{E}[\tau \wedge t] + \mathbb{E}u(B_{t\wedge\tau})] = \mathbb{E}[\tau] + \mathbb{E}[u(B_\tau)] = \mathbb{E}\tau$$

which proves Eqs. (16.29) and (16.30). Lastly using the monotone convergence theorem we find,

$$\mathbb{E}\tau_b = \lim_{a\downarrow-\infty} \mathbb{E}[\tau_a \wedge \tau_b] = \lim_{a\downarrow-\infty} [-ab] = \infty$$

which proves Eq. (16.31). ■

Combining Proposition 16.38 with the next proposition shows that one dimensional Brownian motion hits every point in $\mathbb{R}$ but the expected hitting time of hitting any point other than 0 is infinite. As usual, using the Markov property it follows that one dimensional Brownian motion hits every point in $\mathbb{R}$ infinitely often. So in summary one dimensional Brownian motion is **null recurrent** just the same as for simple random walks on $\mathbb{Z}$.

Proposition 16.39. *Continuing the notation above we have,*

$$\mathbb{P}(\tau_b < \tau_a) = \frac{-a}{b-a} \text{ and } \mathbb{P}(\tau_b < \infty) = 1. \tag{16.33}$$

Proof. Now let $\tau := \tau_a \wedge \tau_b$ and $u(x) = x$. Since $\mathcal{L}u = \frac{1}{2}u'' \equiv 0$, by Theorem 16.31 (or by Itô's lemma in Theorem 17.5 below) we know that $M_t = u(B_t) = B_t$ is a martingale and hence, by the optional sampling Theorem 16.36,

$$0 = \mathbb{E}M_0 = \mathbb{E}M_{t\wedge\tau} = \mathbb{E}[B_{t\wedge\tau}]. \tag{16.34}$$

Since $B_{t\wedge\tau} \in [a,b]$ and $\mathbb{P}(\tau = \infty) = 0$, $B_{t\wedge\tau} \to B_\tau \in \{a,b\}$ boundedly as $t \uparrow \infty$. Therefore using the dominated convergence theorem to pass to the limit in Eq. (16.34), we find

$$\begin{aligned} 0 = \mathbb{E}[B_\tau] &= a\mathbb{P}(B_\tau = a) + b\mathbb{P}(B_\tau = b) \\ &= a \cdot \mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_b < \tau_a) \\ &= a \cdot [1 - \mathbb{P}(\tau_a < \tau_b)] + b\mathbb{P}(\tau_b < \tau_a), \end{aligned}$$

which easily proves the first equality in Eq. (16.33). Since $\{\tau_b < \tau_a\} \subset \{\tau_b < \infty\}$ for all a, it follows that

$$\mathbb{P}(\tau_b < \infty) \geq \frac{-a}{b-a} \to 1 \text{ as } a \downarrow -\infty$$

which proves the second equality in Eq. (16.33). ■

Exercise 16.6 (The Laplace transform of τ_b). Let $b > 0$ and $\tau_b := \inf\{t > 0 : B_t = b\}$. Using (with $\lambda > 0$) the martingale,

$$M_t := e^{\lambda B_t - \frac{1}{2}\lambda^2 t},$$

show using the optional sampling Theorem 16.36 that

$$\mathbb{E}_0\left[e^{-\mu\tau_b}\right] = e^{-b\sqrt{2\mu}} \text{ for all } \mu \geq 0. \tag{16.35}$$

[**Note:** for $0 \leq t \leq \tau_b$ that $e^{\lambda B_t} \leq e^{\lambda b}$ which is useful when applying DCT.]

Remark 16.40. Equation (16.35) may also be proved using Lemma 16.56 directly. Indeed, if

$$p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t},$$

then

$$\begin{aligned} \mathbb{E}_0\left[e^{-\lambda\tau_a}\right] &= \mathbb{E}_0\left[\int_{\tau_a}^\infty \lambda e^{-\lambda t} dt\right] = \mathbb{E}_0\left[\int_0^\infty 1_{\tau_a < t} \lambda e^{-\lambda t} dt\right] \\ &= \mathbb{E}_0\left[\int_0^\infty \mathbb{P}(\tau_a < t)\, \lambda e^{-\lambda t} dt\right] \\ &= -2\int_0^\infty \left[\int_a^\infty p_t(x)\, dx\right] \frac{d}{dt} e^{-\lambda t} dt. \end{aligned}$$

Integrating by parts in t, shows

$$\begin{aligned} -2\int_0^\infty \left[\int_a^\infty p_t(x)\, dx\right] \frac{d}{dt} e^{-\lambda t} dt &= 2\int_0^\infty \left[\int_a^\infty \frac{d}{dt} p_t(x)\, dx\right] e^{-\lambda t} dt \\ &= 2\int_0^\infty \left[\int_a^\infty \frac{1}{2} p_t''(x)\, dx\right] e^{-\lambda t} dt \\ &= -\int_0^\infty p_t'(a)\, e^{-\lambda t} dt \\ &= -\int_0^\infty \frac{1}{\sqrt{2\pi t}} \frac{-a}{t} e^{-a^2/2t} e^{-\lambda t} dt \\ &= \frac{a}{\sqrt{2\pi}} \int_0^\infty t^{-3/2} e^{-a^2/2t} e^{-\lambda t} dt \end{aligned}$$

where the last integral may be evaluated to be $e^{-a\sqrt{2\lambda}}$ using the Laplace transform function of Mathematica.

Alternatively, consider

$$\int_0^\infty p_t(x)\, e^{-\lambda t} dt = \int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} e^{-\lambda t} dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{t}} \exp\left(-\frac{x^2}{2t} - \lambda t\right) dt$$

which may be evaluated using the theory of the Fourier transform (or characteristic functions if you prefer) as follows,

$$\begin{aligned} \frac{\sqrt{2\pi}}{2m} e^{-m|x|} &= \int_{\mathbb{R}} \left(|\xi|^2 + m^2\right)^{-1} e^{i\xi x} \mathbf{d}\xi \\ &= \int_{\mathbb{R}} \left[\int_0^\infty e^{-\lambda(|\xi|^2 + m^2)} d\lambda\right] e^{i\xi x} \mathbf{d}\xi \\ &= \int_0^\infty d\lambda \left[e^{-\lambda m^2} \int_{\mathbb{R}} \mathbf{d}\xi e^{-\lambda|\xi|^2} e^{i\xi x}\right] \\ &= \int_0^\infty e^{-\lambda m^2} (2\lambda)^{-1/2} e^{-\frac{1}{4\lambda}x^2} d\lambda, \end{aligned}$$

where $\mathbf{d}\xi := (2\pi)^{-1/2}\, d\xi$. Now make appropriate change of variables and carry out the remaining x integral to arrive at the result.

Remark 16.41 (A representation formula). Suppose that R is a "nice" bounded sub-domain of $\mathbb{R}^d$, $u \in C^2\left(\bar{R}, \mathbb{C}\right)$, $u = g$ on ∂R and $\frac{1}{2}\Delta u = -f$ on R, then for $x \in R$ we have,

$$u(x) = \mathbb{E}\left[g(x + B_\tau) + \int_0^\tau f(x + B_s)\, ds\right], \tag{16.36}$$

where $\tau = \tau_x := \inf\{t \geq 0 : x + B_t \notin R\}$ is the first exit time from R. Here are some of the proof ideas.

1. First note that $\mathbb{E}\tau < \infty$ using the one dimensional results above.
2. Extend the functions u to $\mathbb{R}^d$ so that u has compact support in which case $M_t := u(x + B_t) - \frac{1}{2}\int_0^t \Delta u(x + B_s)\, ds$ is a martingale.
3. Apply the optional sampling theorem to learn

$$\begin{aligned} M_t^\tau &= u(x + B_{t\wedge\tau}) - \frac{1}{2}\int_0^{t\wedge\tau} \Delta u(x + B_s)\, ds \\ &= u(x + B_{t\wedge\tau}) + \int_0^{t\wedge\tau} f(x + B_s)\, ds \end{aligned}$$

is a martingale and in particular

$$\begin{aligned} u(x) &= \mathbb{E}M_0^\tau = \mathbb{E}M_t^\tau \\ &= \mathbb{E}\left[u(x + B_{\tau\wedge t}) + \int_0^{\tau\wedge t} f(x + B_s)\, ds\right]. \end{aligned}$$

4. Using the dominated convergence theorem, pass to the limit as $t \uparrow \infty$ in the previous identity to arrive at Eq. (16.36).

16.6 Scaling Properties of B. M.

Theorem 16.42 (Transformations preserving B. M.). *Let $\{B_t\}_{t\geq 0}$ be a Brownian motion and $\mathcal{B}_t := \sigma(B_s : s \leq t)$. 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. *$b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion. In particular, $\lim_{t\downarrow 0} tB_{1/t} = 0$ a.s.*
4. *for all $T \in (0, \infty)$, $b_t := B_{t+T} - B_T$ for $t \geq 0$ is again a Brownian motion which is independent of $\mathcal{B}_T$.*
5. *for all $T \in (0, \infty)$, $b_t := B_{T-t} - B_T$ for $0 \leq t \leq T$ is again a Brownian motion on $[0, T]$.*

Proof. It is clear that in each of the four cases above $\{b_t\}_{t\geq 0}$ is still a Gaussian process. Hence to finish the proof it suffices to verify, $\mathbb{E}[b_t b_s] = s \wedge t$ which is routine in all cases. Let us work out item 3. in detail to illustrate the method. For $0 < s < t$,

$$\mathbb{E}[b_s b_t] = st\mathbb{E}[B_{s^{-1}}B_{t^{-1}}] = st\left(s^{-1} \wedge t^{-1}\right) = st \cdot t^{-1} = s.$$

Notice that $t \to b_t$ is continuous for $t > 0$, so to finish the proof we must show that $\lim_{t\downarrow 0} b_t = 0$ a.s. However, this follows from "Kolmogorov's continuity criteria" which we do not cover here.

■

Corollary 16.43 (B. M. Law of Large Numbers). *Suppose $\{B_t\}_{t\geq 0}$ is a Brownian motion, then almost surely, for each $\beta > 1/2$,*

$$\limsup_{t\to\infty} \frac{|B_t|}{t^\beta} = \begin{cases} 0 & \text{if } \beta > 1/2 \\ \infty & \text{if } \beta \in (0, 1/2). \end{cases} \tag{16.37}$$

Proof. We omit the full proof here other than to show that we may replace the limiting behavior at ∞ by statements about the limiting behavior as $t \downarrow 0$ because $b_t := tB_{1/t}$ for $t > 0$ and $b_0 = 0$ is a Brownian motion.

■

16.7 Random Walks to Brownian Motion

Let $\{X_j\}_{j=1}^\infty$ be a sequence of independent Bernoulli random variables with $\mathbb{P}(X_j = \pm 1) = \frac{1}{2}$ and let $W_0 = 0$, $W_n = X_1 + \cdots + X_n$ be the random walk on $\mathbb{Z}$. For each $\varepsilon > 0$, we would like to consider W_n at $n = t/\varepsilon$. We can not expect $W_{t/\varepsilon}$ to have a limit as $\varepsilon \to 0$ without further scaling. To see what scaling is needed, recall that

$$\mathrm{Var}(X_1) = \mathbb{E}X_1^2 = \frac{1}{2}1^2 + \frac{1}{2}(-1)^2 = 1$$

and therefore, $\mathrm{Var}(W_n) = n$. Thus we have

$$\mathrm{Var}\left(W_{t/\varepsilon}\right) = t/\varepsilon$$

and hence to get a limit we should scale $W_{t/\varepsilon}$ by $\sqrt{\varepsilon}$. These considerations motivate the following theorem.

Theorem 16.44. *For all $\varepsilon > 0$, let $\{B_\varepsilon(t)\}_{t\geq 0}$ be the continuous time process, defined as follows:*

1. *If $t = n\varepsilon$ for some $n \in \mathbb{N}_0$, let $B_\varepsilon(n\varepsilon) := \sqrt{\varepsilon}W_n$ and*

2. *if* $n\varepsilon < t < (n+1)\varepsilon$, *let* $B_\varepsilon(t)$ *be given by*

$$\begin{aligned} B_\varepsilon(t) &= B_\varepsilon(n\varepsilon) + \frac{t-n\varepsilon}{\varepsilon}\left(B_\varepsilon((n+1)\varepsilon) - B_\varepsilon(n\varepsilon)\right) \\ &= \sqrt{\varepsilon}W_n + \frac{t-n\varepsilon}{\varepsilon}\left(\sqrt{\varepsilon}W_{n+1} - \sqrt{\varepsilon}W_n\right) \\ &= \sqrt{\varepsilon}W_n + \frac{t-n\varepsilon}{\varepsilon}\sqrt{\varepsilon}X_{n+1}, \end{aligned}$$

i.e. $B_\varepsilon(t)$ *is the linear interpolation between* $(n\varepsilon, \sqrt{\varepsilon}W_n)$ *and* $((n+1)\varepsilon, \sqrt{\varepsilon}W_{n+1})$, *see Figure 16.2.*

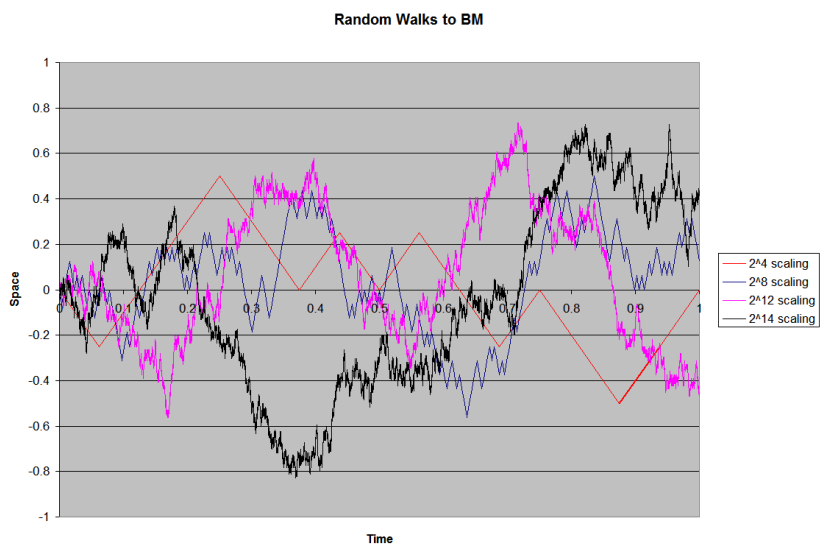


Fig. 16.2. The four graphs are constructed (in Excel) from a single realization of a random walk. Each graph corresponds to a different scaling parameter, namely, $\varepsilon \in \{2^4, 2^8, 2^{12}, 2^{14}\}$. It is clear from these pictures that $B_\varepsilon(t)$ is not converging to $B(t)$ for each realization. The convergence is only in law.

Then $B_\varepsilon \implies B$ *("weak convergence") as* $\varepsilon \downarrow 0$, *where* B *is a continuous random process which we call a* ***Brownian motion.***

Remark 16.45. Theorem 16.44 shows that Brownian motion is the result of **combining additively** the effects of many small random steps.

We may allow for more general random walk limits which is the content of Donsker's invariance principle or the so called functional central limit theorem.

The setup is to start with a "random walk," $S_n := X_1 + \cdots + X_n$ where $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with zero mean and variance one. We then define for each $n \in \mathbb{N}$ the following continuous process,

$$B_n(t) := \frac{1}{\sqrt{n}}\left(S_{[nt]} + (nt - [nt])\, X_{[nt]+1}\right) \tag{16.38}$$

where for $\tau \in \mathbb{R}_+$, $[\tau]$ is the integer part of τ, i.e. the nearest integer to τ which is no greater than τ.

Theorem 16.46 (Donsker's invariance principle). *Let* Ω, B_n, *and* B *be as above. Then for* $B_n \implies B$, *i.e.*

$$\lim_{n\to\infty} \mathbb{E}\left[F(B_n)\right] = \mathbb{E}\left[F(B)\right]$$

for all bounded continuous functions $F : \Omega \to \mathbb{R}$.

For the proof of this theorem we refer the reader to [11], [18], [10], or [15].

16.8 Path Regularity Properties of BM

For the rest of this chapter we will assume that $\{B_t\}_{t\geq 0}$ is a Brownian motion on some probability space, $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ and $\mathcal{F}_t := \sigma(B_s : s \leq t)$.

Notation 16.47 (Partitions) *Given* $\Pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$, *a partition of* $[0,T]$, *let*

$$\Delta_i B := B_{t_i} - B_{t_{i-1}}, \text{ and } \Delta_i t := t_i - t_{i-1}$$

for all $i = 1, 2, \ldots, n$. *Further let* $\operatorname{mesh}(\Pi) := \max_i |\Delta_i t|$ *denote the mesh of the partition,* Π.

Exercise 16.7 (Quadratic Variation). Let

$$\Pi_m := \left\{0 = t_0^m < t_1^m < \cdots < t_{n_m}^m = T\right\}$$

be a sequence of partitions such that $\operatorname{mesh}(\Pi_m) \to 0$ as $m \to \infty$. Further let

$$Q_m := \sum_{i=1}^{n_m} (\Delta_i B)^2 := \sum_{i=1}^{n_m} \left(B_{t_i^m} - B_{t_{i-1}^m}\right)^2. \tag{16.39}$$

Show

$$\lim_{m\to\infty} \mathbb{E}\left[(Q_m - T)^2\right] = 0.$$

Also show that

$$\sum_{m=1}^{\infty} \operatorname{mesh}(\Pi_m) < \infty \implies \mathbb{E}\left[\sum_{m=1}^{\infty} (Q_m - T)^2\right] < \infty$$
$$\implies \lim_{m\to\infty} Q_m = T \text{ a.s.}$$

[These results are often abbreviated by the writing, $dB_t^2 = dt$.]

Hints: it is useful to observe; 1)

$$Q_m - T = \sum_{i=1}^{n_m} \left[(\Delta_i B)^2 - \Delta_i t\right]$$

and 2) using $\Delta_i B \overset{d}{=} \sqrt{\Delta_i t} N(0,1)$ one easily shows there is a constant, $c < \infty$ such that

$$\mathbb{E}\left[(\Delta_i B)^2 - \Delta_i t\right]^2 = c(\Delta_i t)^2. \tag{16.40}$$

Proposition 16.48. *Suppose that $\{\Pi_m\}_{m=1}^{\infty}$ is a sequence of partitions of $[0,T]$ such that $\Pi_m \subset \Pi_{m+1}$ for all m and $\operatorname{mesh}(\Pi_m) \to 0$ as $m \to \infty$. Then $Q_m \to T$ a.s. where Q_m is defined as in Eq. (16.39).*

Corollary 16.49 (Roughness of Brownian Paths). *A Brownian motion, $\{B_t\}_{t\geq 0}$, is **not** almost surely α – Hölder continuous for any $\alpha > 1/2$.*

Proof. According to Exercise 16.7, we may choose partition, Π_m, such that $\operatorname{mesh}(\Pi_m) \to 0$ and $Q_m \to T$ a.s. If B were α – Hölder continuous for some $\alpha > 1/2$, then

$$Q_m = \sum_{i=1}^{n_m} (\Delta_i^m B)^2 \leq C \sum_{i=1}^{n_m} (\Delta_i^m t)^{2\alpha} \leq C \max\left([\Delta_i t]^{2\alpha-1}\right) \sum_{i=1}^{n_m} \Delta_i^m t$$
$$\leq C\left[\operatorname{mesh}(\Pi_m)\right]^{2\alpha-1} T \to 0 \text{ as } m \to \infty$$

which contradicts the fact that $Q_m \to T$ as $m \to \infty$. ■

16.9 The Strong Markov Property of Brownian Motion

Notation 16.50 *If ν is a probability measure on $\mathbb{R}^d$, we let $\mathbb{P}_\nu$ indicate that we are starting our Brownian motion so that B_0 distributed as ν. For Brownian motion, the expectation relative to $\mathbb{P}_\nu$ can be defined to be*

$$\mathbb{E}_\nu\left[F\left(B_{(\cdot)}\right)\right] = \int_{\mathbb{R}^d} \mathbb{E}_0\left[F\left(x + B_{(\cdot)}\right)\right] d\nu(x).$$

Theorem 16.51 (Strong-Markov Property). *Let ν be a probability measure on $\mathbb{R}^d$, τ be an optional time with $\mathbb{P}_\nu(\tau < \infty) > 0$, and let*

$$b_t := B_{t+\tau} - B_\tau \text{ on } \{\tau < \infty\}.$$

Then, conditioned on $\{\tau < \infty\}$, b is a Brownian motion starting at $0 \in \mathbb{R}^d$ which is independent of $\mathcal{F}_\tau^+$. To be more precise we are claiming, for all bounded measurable functions $F : \Omega \to \mathbb{R}^d$ and all $A \in \mathcal{F}_\tau^+$, that

$$\mathbb{E}_\nu\left[F(b) \,|\tau < \infty\right] = \mathbb{E}_0\left[F\right] \tag{16.41}$$

and

$$\mathbb{E}_\nu\left[F(b)\, 1_A | \tau < \infty\right] = \mathbb{E}_\nu\left[F(b) \,|\tau < \infty\right] \cdot \mathbb{E}_\nu\left[1_A | \tau < \infty\right]. \tag{16.42}$$

Corollary 16.52. *Let ν be a probability measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, $T > 0$,and let*

$$b_t := B_{t+T} - B_T \text{ for } t \geq 0.$$

Then b is a Brownian motion starting at $0 \in \mathbb{R}^d$ which is independent of $\mathcal{F}_T$ relative to the probability measure, $\mathbb{P}_\nu$, describing Brownian motion started with B_0 distributed as ν.

Proof. This is a direct consequence of Theorem 16.51 with $\tau = T$ or it can be proved directly using the Markov property alone. ■

Proposition 16.53 (Stitching Lemma). *Suppose that $\{B_t\}_{t\geq 0}$ and $\{X_t\}_{t\geq 0}$ are Brownian motions, τ is an optional time, and $\{X_t\}_{t\geq 0}$ is independent of $\mathcal{F}_\tau^+$. Then $\tilde{B}_t := B_{t\wedge\tau} + X_{(t-\tau)_+}$ is another Brownian motion.*

For the next couple of results we will follow [10, Chapter 13] (also see [11, Section 2.8]) where more results along this line may be found.

Theorem 16.54 (Reflection Principle). *Let τ be an optional time and $\{B_t\}_{t\geq 0}$ be a Brownian motion. Then the "reflected" process (see Figure 16.3),*

$$\tilde{B}_t := B_{t\wedge\tau} - (B_t - B_{t\wedge\tau}) = \begin{cases} B_t & \text{if } t \leq \tau \\ B_\tau - (B_t - B_\tau) & \text{if } t > \tau, \end{cases}$$

is again a Brownian motion.

Proof. Let $T > 0$ be given and set $b_t := B_{t+\tau\wedge T} - B_{\tau\wedge T}$. Then applying Proposition 16.53 with $X = b$ and τ replaced by $\tau \wedge T$ we learn that $\tilde{B}|_{[0,T]} \overset{d}{=} B|_{[0,T]}$. Again as $T > 0$ was arbitrary we may conclude that $\tilde{B} \overset{d}{=} B$. ■

Lemma 16.55. *If $d = 1$, then*

$$\mathbb{P}_0 (B_t > a) \le \min \left(\sqrt{\frac{t}{2\pi a^2}} e^{-a^2/2t}, \frac{1}{2} e^{-a^2/2t} \right)$$

for all $t > 0$ and $a > 0$.

Proof. This follows from Lemma 19.7 and the fact that $B_t \overset{d}{=} \sqrt{t} N$, where N is a standard normal random variable. ■

Lemma 16.56 (Running maximum). *If B_t is a 1 - dimensional Brownian motion and $z > 0$ and $y \ge 0$, then*

$$\mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t < z - y \right) = \mathbb{P}(B_t > z + y) \tag{16.43}$$

and

$$\mathbb{P} \left(\max_{s \le t} B_s \ge z \right) = 2P(B_t > z) = \mathbb{P} (|B_t| > z) . \tag{16.44}$$

In particular we have $\max_{s \le t} B_s \overset{d}{=} |B_t|$.

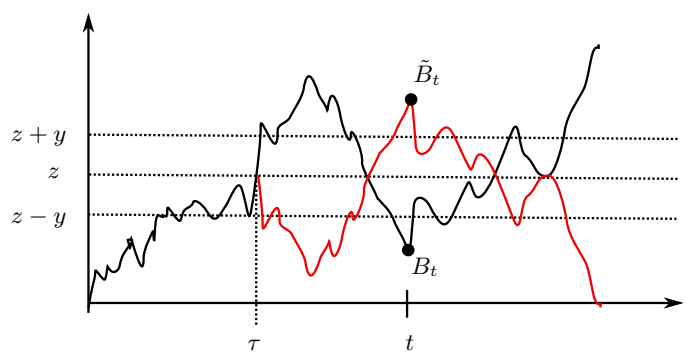


Fig. 16.3. A Brownian path B along with its reflection about the first time, τ, that B_t hits the level z.

Proof. Intuitive proof of Eq. (16.44). For $z > 0$ we have

$$\begin{aligned} \left\{ \max_{s \le t} B_s \ge z \right\} &= \{\tau_z \le t\} = \{\tau_z \le t, B_t < z\} \cup \{\tau_z \le t, B_t \ge z\} . \\ &= \{\tau_z \le t, B_t < z\} \cup \{B_t \ge z\} \end{aligned}$$

In order to compute these probabilities we make use of the intuitively "clear" fact that the events

$$\{\tau_z \le t, B_t < z\} \text{ and } \{\tau_z \le t, B_t > z\} = \{B_t > z\}$$

are equally likely[4], see Figure 16.3. Assuming this assertion and using $\mathbb{P} (B_t = z) = 0$ we find,

$$\begin{aligned} \mathbb{P} \left(\max_{s \le t} B_s \ge z \right) &= \mathbb{P} (\tau_z \le t, B_t < z) + \mathbb{P} (B_t > z) \\ &= 2\mathbb{P} (B_t > z) = 2\mathbb{P} (B_t \ge z) . \end{aligned}$$

Formal proof of Eq. (16.44). Let $z > 0$ be given. Since $\{\max_{s \le t} B_s \ge z, \ B_t \ge z\} = \{B_t \ge z\}$ and $\mathbb{P} (B_t \ge z) = \mathbb{P} (B_t > z)$ we find

$$\begin{aligned} \mathbb{P} \left(\max_{s \le t} B_s \ge z \right) &= \mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t \ge z \right) + \mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t < z \right) \\ &= \mathbb{P} (B_t \ge z) + \mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t < z \right) \\ &= \mathbb{P} (B_t > z) + \mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t < z \right) . \end{aligned}$$

So to finish the proof of Eq. (16.44) it suffices verify Eq. (16.43) for $y = 0$.

Let $\tau := \tau_z = \inf \{t > 0 : B_t = z\}$ and let $\tilde{B}_t := B_t^\tau - b_{(t-\tau)_+}$ where $b_t := (B_{t \wedge \tau} - B_\tau) 1_{\tau < \infty}$, see Figure 16.3. (By Corollary 16.43 we actually know that $\tau_z < \infty$ a.s. but we will not use this fact in the proof.) Observe from Figure 16.3 that

$$\begin{aligned} \tilde{\tau} &:= \inf \left\{ t > 0 : \tilde{B}_t = z \right\} = \tau, \\ \{B_t < z\} &= \left\{ \tilde{B}_t > z \right\} \text{ on } \{\tau \le t\} \text{ and} \\ \left\{ \max_{s \le t} B_s \ge z \right\} &= \{\tau \le t\} = \{\tilde{\tau} \le t\} . \end{aligned} \tag{16.45}$$

Hence we have

$$\begin{aligned} \mathbb{P} \left(\max_{s \le t} B_s \ge z, \ B_t < z \right) &= \mathbb{P} (\tau \le t, \ B_t < z) = \mathbb{P} \left(\tilde{\tau} \le t, \tilde{B}_t > z \right) \\ &= \mathbb{P} (\tau \le t, B_t > z) = \mathbb{P} (B_t > z) \end{aligned} \tag{16.46}$$

wherein we have used $\text{Law} (B_\cdot, \tau) = \text{Law} \left(\tilde{B}_\cdot, \tilde{\tau} \right)$ in the third equality and $\{B_t > z\} \subset \{\tau \le t\}$ for the last equality.

[4] Think in terms of a random walker starting at z.

Proof of Eq. (16.43) for general $y \geq 0$. We simply follow the above proof except we use the identity,

$$\{B_t < z - y\} = \left\{\tilde{B}_t > z + y\right\} \text{ on } \{\tau \leq t\},$$

in place of Eq. (16.45). Since $\{B_t > z + y\} \subset \{\tau \leq t\}$ and working as above we find

$$\begin{aligned}\mathbb{P}\left(\max_{s\leq t} B_s \geq z,\ B_t < z - y\right) &= \mathbb{P}\left(\tau \leq t,\ B_t < z - y\right) = \mathbb{P}\left(\tau = \tilde{\tau} \leq t, \tilde{B}_t > z + y\right) \\ &= \mathbb{P}\left(\tau \leq t, B_t > z + y\right) = \mathbb{P}\left(B_t > z + y\right).\end{aligned}$$

■

Remark 16.57. Notice that

$$\begin{aligned}\mathbb{P}\left(\max_{s\leq t} B_s \geq z\right) &= \mathbb{P}\left(|B_t| > z\right) = \mathbb{P}\left(\sqrt{t}\,|B_1| > z\right) \\ &= \mathbb{P}\left(|B_1| > \frac{z}{\sqrt{t}}\right) \to 1 \text{ as } t \to \infty\end{aligned}$$

and therefore it follows that $\sup_{s<\infty} B_s = \infty$ a.s. In particular this shows that $\tau_z < \infty$ a.s. for all $z \geq 0$ which we have already seen in Corollary 16.43.

Corollary 16.58. *Suppose now that* $T = \inf\{t > 0 : |B_t| = a\}$, *i.e. the first time* B_t *leaves the strip* $(-a, a)$. *Then*

$$\begin{aligned}\mathbb{P}_0(T < t) &\leq 4P_0(B_t > a) = \frac{4}{\sqrt{2\pi t}}\int_a^\infty e^{-x^2/2t}dx \\ &\leq \min\left(\sqrt{\frac{8t}{\pi a^2}}e^{-a^2/2t}, 1\right). \qquad (16.47)\end{aligned}$$

Notice that $\mathbb{P}_0(T < t) = \mathbb{P}_0(B_t^* \geq a)$ *where* $B_t^* = \max\{|B_\tau| : \tau \leq t\}$. *So Eq. (16.47) may be rewritten as*

$$\mathbb{P}_0(B_t^* \geq a) \leq 4P_0(B_t > a) \leq \min\left(\sqrt{\frac{8t}{\pi a^2}}e^{-a^2/2t}, 1\right) \leq 2e^{-a^2/2t}. \qquad (16.48)$$

Proof. By definition $T = T_a \wedge T_{-a}$ so that $\{T < t\} = \{T_a < t\} \cup \{T_{-a} < t\}$ and therefore

$$\begin{aligned}\mathbb{P}_0(T < t) &\leq \mathbb{P}_0\left(T_a < t\right) + \mathbb{P}_0\left(T_{-a} < t\right) \\ &= 2P_0(T_a < t) = 4P_0(B_t > a) = \frac{4}{\sqrt{2\pi t}}\int_a^\infty e^{-x^2/2t}dx \\ &\leq \frac{4}{\sqrt{2\pi t}}\int_a^\infty \frac{x}{a}e^{-x^2/2t}dx = \frac{4}{\sqrt{2\pi t}}\left(-\frac{t}{a}e^{-x^2/2t}\right)\Big|_a^\infty = \sqrt{\frac{8t}{\pi a^2}}e^{-a^2/2t}.\end{aligned}$$

This proves everything but the very last inequality in Eq. (16.48). To prove this inequality first observe the elementary calculus inequality:

$$\min\left(\frac{4}{\sqrt{2\pi}y}e^{-y^2/2}, 1\right) \leq 2e^{-y^2/2}. \qquad (16.49)$$

Indeed Eq. (16.49) holds $\frac{4}{\sqrt{2\pi}y} \leq 2$, i.e. if $y \geq y_0 := 2/\sqrt{2\pi}$. The fact that Eq. (16.49) holds for $y \leq y_0$ follows from the following trivial inequality

$$1 \leq 1.4552 \cong 2e^{-\frac{1}{\pi}} = e^{-y_0^2/2}.$$

Finally letting $y = a/\sqrt{t}$ in Eq. (16.49) gives the last inequality in Eq. (16.48). ■

Corollary 16.59 (Fernique's Theorem). *For all* $\lambda > 0$, *then*

$$\mathbb{E}\left[e^{\lambda\left[\max_{s\leq t} B_s\right]^2}\right] = \begin{cases}\frac{1}{\sqrt{1-2\lambda t}} & \text{if } \lambda < \frac{1}{2t} \\ \infty & \text{if } \lambda \geq \frac{1}{2t}\end{cases}. \qquad (16.50)$$

Moreover,

$$\mathbb{E}\left[e^{\lambda\|B_\cdot\|_{\infty,t}^2}\right] < \infty \text{ iff } \lambda < \frac{1}{2t}. \qquad (16.51)$$

Proof. From Lemma 16.56,

$$\begin{aligned}\mathbb{E}\left[e^{\lambda\left[\max_{s\leq t} B_s\right]^2}\right] &= \mathbb{E}e^{\lambda|B_t|^2} = \mathbb{E}e^{\lambda t|B_1|^2} \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\lambda t x^2}e^{-\frac{1}{2}x^2}dx \\ &= \begin{cases}\frac{1}{\sqrt{1-2\lambda t}} & \text{if } \lambda < \frac{1}{2t} \\ \infty & \text{if } \lambda \geq \frac{1}{2t}\end{cases}.\end{aligned}$$

By Corollary 16.58 we can show $\mathbb{E}\left[e^{\lambda\|B_\cdot\|_{\infty,t}^2}\right] < \infty$ if $\lambda < \frac{1}{2t}$ while if $\lambda \geq \frac{1}{2t}$ we will have,

$$\mathbb{E}\left[e^{\lambda\|B_\cdot\|_{\infty,t}^2}\right] \geq \mathbb{E}\left[e^{\lambda\left[\max_{s\leq t} B_s\right]^2}\right] = \infty.$$

Alternative weak bound. For a cruder estimate, notice that

$$\|B_\cdot\|_{\infty,t} := \max_{s\leq t}|B_s| = \left[\max_{s\leq t} B_s\right] \vee \left[\max_{s\leq t}[-B_s]\right]$$

and hence

$$\|B_\cdot\|_{\infty,t}^2 = \left[\max_{s\leq t} B_s\right]^2 \vee \left[\max_{s\leq t}[-B_s]\right]^2 \leq \left[\max_{s\leq t} B_s\right]^2 + \left[\max_{s\leq t}[-B_s]\right]^2.$$

Therefore using $\{-B_s\}_{s\geq 0}$ is still a Brownian motion along with the Cauchy-Schwarz's inequality gives,

$$\begin{aligned}\mathbb{E}\left[e^{\lambda\|B_\cdot\|^2_{\infty,t}}\right] &\leq \mathbb{E}\left[e^{\lambda\left[\max_{s\leq t} B_s\right]^2}\cdot e^{\lambda\left[\max_{s\leq t}(-B_s)\right]^2}\right]\\ &\leq \sqrt{\mathbb{E}\left[e^{2\lambda\left[\max_{s\leq t} B_s\right]^2}\right]\cdot\mathbb{E}\left[\cdot e^{\lambda\left[\max_{s\leq t}(-B_s)\right]^2}\right]}\\ &= \mathbb{E}\left[e^{2\lambda\left[\max_{s\leq t} B_s\right]^2}\right] < \infty \text{ if } \lambda < \frac{1}{4t}.\end{aligned}$$

■

16.10 Dirichelt Problem and Brownian Motion

Theorem 16.60 (The Dirichlet problem). *Suppose D is an open subset of $\mathbb{R}^d$ and $\tau := \inf\{t\geq 0 : B_t \in D^c\}$ is the first exit time form D. Given a bounded measurable function, $f : \mathrm{bd}(D)\to\mathbb{R}$, let $u : D\to\mathbb{R}$ be defined by (see Figure 16.4),*

$$u(x) := \mathbb{E}_x\left[f(B_\tau) : \tau < \infty\right] \text{ for } x\in D.$$

Then $u\in C^\infty(D)$ and $\Delta u = 0$ on D, i.e. u is a harmonic function.

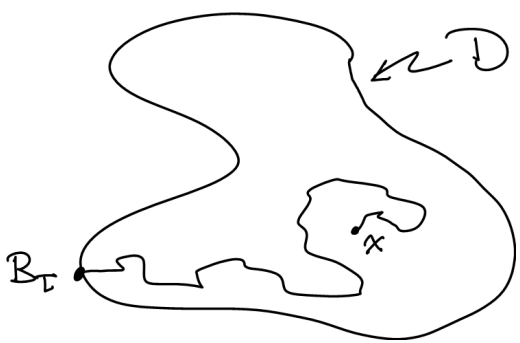


Fig. 16.4. Brownian motion starting at $x\in D$ and exiting on the boundary of D at B_τ.

Proof. (Sketch.) Let $x\in D$ and $r>0$ be such that $B(x,r)\subset D$ and let $\sigma := \inf\{t\geq 0 : B_t\notin B(x,r)\}$ as in Figure 16.5. Setting $F := f(B_\tau)1_{\{\tau<\infty\}}$, we see that $F\circ\theta_\sigma = F$ on $\{\sigma<\infty\}$ and that $\{\sigma<\infty\}$, $\mathbb{P}_x$ - a.s. on $\{\tau<\infty\}$. Moreover, by either Corollary 16.43 or by Lemma 16.56 (see Remark 16.57), we know that $\{\sigma<\infty\}$, $\mathbb{P}_x$ – a.s. Therefore by the strong Markov property,

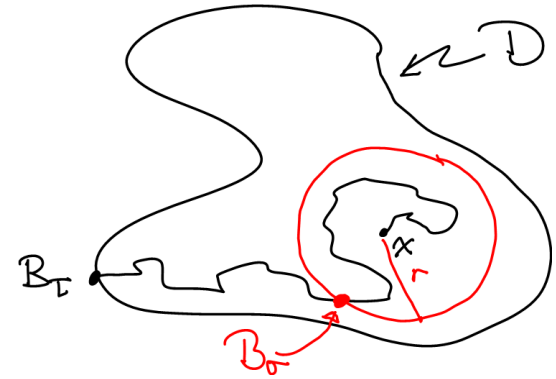


Fig. 16.5. Brownian motion starting at $x\in D$ and exiting the boundary of $B(x,r)$ at B_σ before exiting on the boundary of D at B_τ.

$$\begin{aligned}u(x) &= \mathbb{E}_x[F] = \mathbb{E}_x[F : \sigma<\infty] = \mathbb{E}_x[F\circ\theta_\sigma : \sigma<\infty]\\ &= \mathbb{E}_x[\mathbb{E}_{B_\sigma}F : \sigma<\infty] = \mathbb{E}_x[u(B_\sigma)].\end{aligned}$$

Using the rotation invariance of Brownian motion, we may conclude that

$$\mathbb{E}_x[u(B_\sigma)] = \frac{1}{\rho(\mathrm{bd}(B(x,r)))}\int_{\mathrm{bd}(B(x,r))} u(y)\,d\rho(y)$$

where ρ denotes surface measure on $\mathrm{bd}(B(x,r))$. This shows that $u(x)$ satisfies the mean value property, i.e. $u(x)$ is equal to its average about in sphere centered at x which is contained in D. It is now a well known that this property, see for example [11, Proposition 4.2.5 on p. 242], that this implies $u\in C^\infty(D)$ and that $\Delta u = 0$. ■

When the boundary of D is sufficiently regular and f is continuous on $\mathrm{bd}(D)$, it can be shown that, for $x\in\mathrm{bd}(D)$, that $u(y)\to f(x)$ as $y\in D$ tends to x. For more details in this direction, see [1], [2], [11, Section 4.2], [8], and [7].

A short introduction to Itô's calculus

17.1 A short introduction to Itô's calculus

Definition 17.1. *The Itô integral of an adapted process*[1] $\{f_t\}_{t\geq 0}$, *is defined by*

$$\int_0^T f dB = \lim_{|\Pi|\to 0} \sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right) \tag{17.1}$$

when the limit exists. Here Π denotes a partition of $[0,T]$ so that $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ and

$$|\Pi| = \max_{1\leq i\leq n} \Delta_i t \text{ where } \Delta_i t = t_i - t_{i-1}. \tag{17.2}$$

Proposition 17.2. *Keeping the notation in Definition 17.1 and further assume $\mathbb{E} f_t^2 < \infty$ for all t. Then we have,*

$$\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] = 0$$

and

$$\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 = \mathbb{E}\sum_{i=1}^n f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right).$$

Proof. Since $\left(B_{t_i} - B_{t_{i-1}}\right)$ is independent of $f_{t_{i-1}}$ we have,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] &= \sum_{i=1}^n \mathbb{E} f_{t_{i-1}} \mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right) \\ &= \sum_{i=1}^n \mathbb{E} f_{t_{i-1}} \cdot 0 = 0.\end{aligned}$$

For the second assertion, we write,

$$\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 = \sum_{i,j=1}^n f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right).$$

If $j < i$, then $f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}}$ is independent of $\left(B_{t_i} - B_{t_{i-1}}\right)$ and therefore,

$$\begin{aligned}&\mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] \\ &\qquad = \mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}}\right] \cdot \mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right) = 0.\end{aligned}$$

Similarly, if $i < j$,

$$\mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] = 0.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 &= \sum_{i,j=1}^n \mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[f_{t_{i-1}}^2 \left(B_{t_i} - B_{t_{i-1}}\right)^2\right] \\ &= \sum_{i=1}^n \mathbb{E} f_{t_{i-1}}^2 \cdot \mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right)^2 \\ &= \sum_{i=1}^n \mathbb{E} f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right) \\ &= \mathbb{E}\sum_{i=1}^n f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right),\end{aligned}$$

wherein the fourth equality we have used $B_{t_i} - B_{t_{i-1}}$ is independent of $f_{t_{i-1}}$. ■

This proposition motivates the following theorem which will not be proved in full here.

[1] To say f is adapted means that for each $t \geq 0$, f_t should only depend on $\{B_s\}_{s\leq t}$, i.e. $f_t = F_t\left(\{B_s\}_{s\leq t}\right)$.

Theorem 17.3. *If $\{f_t\}_{t\geq 0}$ is an adapted process such that $\mathbb{E}\int_0^T f_t^2 dt < \infty$, then the Itô integral, $\int_0^T f dB$, exists and satisfies,*

$$\mathbb{E}\int_0^T f dB = 0 \text{ and}$$
$$\mathbb{E}\left(\int_0^T f dB\right)^2 = \mathbb{E}\int_0^T f_t^2 dt.$$

[More generally, $M_t := \int_0^t f dB$ is a martingale such that $M_t^2 - \int_0^t f_s^2 ds$ is also a martingale.]

Proof. Let us explain the martingale property for $\int_0^t f dB$ here. First if $f_t = u \cdot 1_{[a,b)}(t)$ with $0 \leq a < b < \infty$ and u is a bounded $\mathcal{F}_a$ – measurable function, then

$$M_t = \int_0^t f dB = \int_0^\infty 1_{[0,t)} f_t dB_t = u\left[B_{t\wedge b} - B_{t\wedge a}\right].$$

If $t \leq a$ will have $M_t = 0$ and so we may as well assume that $t \geq a$. If $a \leq s \leq t$, then

$$\begin{aligned}\mathbb{E}\left[M_t|\mathcal{F}_s\right] &= \mathbb{E}\left[u\left[B_{t\wedge b} - B_{t\wedge a}\right]|\mathcal{F}_s\right] = u\mathbb{E}\left[B_{t\wedge b} - B_a|\mathcal{F}_s\right] \\ &= u\left(B_{s\wedge b} - B_a\right) = M_s\end{aligned}$$

while if $s < a \leq t$ we have

$$\mathbb{E}\left[M_t|\mathcal{F}_s\right] = \mathbb{E}\left[\mathbb{E}\left[M_t|\mathcal{F}_a\right]|\mathcal{F}_s\right] = \mathbb{E}\left[M_a|\mathcal{F}_s\right] = 0 = M_s.$$

This shows M is a martingale in this case. In general, $\int_0^t f dB$ is the L^2 – limit of linear combinations of such martingales and hence is again a martingale. Moreover, by the continuous analogue of Doob's maximal inequality, see Proposition 13.57, we may arrange it so that $t \to \int_0^t f dB$ is still continuous.

We leave it as an exercise to the reader to show; if $f_t = u \cdot 1_{[a,b)}(t)$, then

$$M_t^2 - \int_0^t f_s^2 ds = u^2\left[\left[B_{t\wedge b} - B_{t\wedge a}\right]^2 - (t\wedge b - t\wedge a)\right]$$

is a martingale. ■

Corollary 17.4. *In particular if τ is a bounded stopping time (say $\tau \leq T < \infty$) then*

$$\mathbb{E}\int_0^\tau f dB = 0 \text{ and}$$
$$\mathbb{E}\left(\int_0^\tau f dB\right)^2 = \mathbb{E}\int_0^\tau f_t^2 dt.$$

Proof. The point is that, by the definition of a stopping time, $1_{0\leq t\leq\tau} f_t$ is still an adapted process. Therefore we have,

$$\mathbb{E}\int_0^\tau f dB = \mathbb{E}\left[\int_0^T 1_{0\leq t\leq\tau} f_t dB_t\right] = 0$$

and

$$\begin{aligned}\mathbb{E}\left(\int_0^\tau f dB\right)^2 &= \mathbb{E}\left[\int_0^T 1_{0\leq t\leq\tau} f_t dB_t\right]^2 \\ &= \mathbb{E}\left[\int_0^T \left(1_{0\leq t\leq\tau} f_t\right)^2 dt\right] = \mathbb{E}\left[\int_0^\tau f_t^2 dt\right].\end{aligned}$$

■

Theorem 17.5 (Itô's Lemma). *If $f(t,x)$ is C^1 – function such that $\partial_x^2 f(t,x)$ exists and is continuous, then*

$$\begin{aligned}df(t,B_t) &= \left[\partial_t f(t,B_t) + \frac{1}{2}\partial_x^2 f(t,B_t)\right] dt + \partial_x f(t,B_t)\, dB_t \\ &= \mathcal{L}f(t,B_t)\, dt + \partial_x f(t,B_t)\, dB_t.\end{aligned}$$

More precisely,

$$\begin{aligned}f(T,B_T) &= f(0,B_0) + \int_0^T \partial_x f(t,B_t)\, dB_t \\ &\quad + \int_0^T \left[\partial_t f(t,B_t) + \frac{1}{2}\partial_x^2 f(t,B_t)\right] dt. \qquad (17.3)\end{aligned}$$

Roughly speaking, all differentials should be expanded out to second order using the Itô multiplication rules,

$$dB^2 = dt \text{ and } dBdt = 0 = dt^2.$$

Proof. The rough idea is as follows. Let $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$, so that by a telescoping series argument,

$$f(T,B_T) - f(0,B_0) = \sum_{i=1}^n \left[f(t_i,B_{t_i}) - f\left(t_{i-1},B_{t_{i-1}}\right)\right].$$

Then by Taylor's theorem,

$$\begin{aligned}f(t_i,B_{t_i}) - f\left(t_{i-1},B_{t_{i-1}}\right) &= \partial_t f\left(t_{i-1},B_{t_{i-1}}\right)\Delta_i t + \partial_x f\left(t_{i-1},B_{t_{i-1}}\right)\Delta_i B \\ &\quad + \frac{1}{2}\partial_x^2 f\left(t_{i-1},B_{t_{i-1}}\right)\left[\Delta_i B\right]^2 + O\left(\Delta_i t^2, \Delta_i B^3\right).\end{aligned}$$

The error terms are negligible, i.e.

$$\lim_{|\Pi|\to 0}\sum_{i=1}^{n} O\left(\Delta_i t^2, \Delta_i B^3\right) = 0$$

and hence

$$\begin{aligned}
&f(T,B_T) - f(0,B_0) \\
&\quad = \lim_{|\Pi|\to 0}\sum_{i=1}^{n}\left[f(t_i,B_{t_i}) - f\left(t_{i-1},B_{t_{i-1}}\right)\right] \\
&\quad = \lim_{|\Pi|\to 0}\sum_{i=1}^{n}\partial_t f\left(t_{i-1},B_{t_{i-1}}\right)\Delta_i t + \lim_{|\Pi|\to 0}\sum_{i=1}^{n}\partial_x f\left(t_{i-1},B_{t_{i-1}}\right)\Delta_i B \\
&\qquad + \frac{1}{2}\lim_{|\Pi|\to 0}\sum_{i=1}^{n}\partial_x^2 f\left(t_{i-1},B_{t_{i-1}}\right)\left[\Delta_i B\right]^2 \\
&\quad = \int_0^T \partial_t f(t,B_t)\,dt + \int_0^T \partial_x f(t,B_t)\,dB_t + \frac{1}{2}\int_0^T \partial_x^2 f(t,B_t)\,dt
\end{aligned}$$

where the last term follows by an extension of Exercise 16.7. ■

Example 17.6. Let us verify Itô's lemma in the special case where $f(x) = x^2$. For $T > 0$ and

$$\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$$

being a partition of $[0,T]$, we have, with $\Delta_i B := B_{t_i} - B_{t_{i-1}}$ that

$$\begin{aligned}
B_T^2 &= \sum_{i=1}^{n}\left(B_{t_i}^2 - B_{t_{i-1}}^2\right) \\
&= \sum_{i=1}^{n}\left(B_{t_i} + B_{t_{i-1}}\right)\left(B_{t_i} - B_{t_{i-1}}\right) \\
&= \sum_{i=1}^{n}\left(2B_{t_{i-1}} + \Delta_i B\right)\Delta_i B \\
&= 2\sum_{i=1}^{n} B_{t_{i-1}}\Delta_i B + Q_\Pi \overset{|\Pi|\to 0}{\to} 2\int_0^T B\,dB + T
\end{aligned}$$

where we have used Exercise 16.7 in finding the limit. Hence we conclude

$$B_T^2 = 2\int_0^T B\,dB + T$$

which is exactly Eq. (17.3) for $f(x) = x^2$.

17.2 Itô's formula, heat equations, and harmonic functions

Suppose D is a bounded open subset of $\mathbb{R}^d$ and $x \in D$. Let $B_t = \left(B_t^1, \ldots, B_t^d\right)$ be a standard d – dimensional Brownian motion. Let $X_t := x + B_t^\tau$ where

$$\tau = \inf\{t \geq 0 : x + B_t \notin D\}$$

be the first exit time from D. [**Warning:** I will be using the "natural" multi-dimensional and "stopped" extensions of the Itô theory developed above.] In this multi-dimensional setting Itô's formula becomes,

$$df(t,B_t) = \left[\partial_t f(t,B_t) + \frac{1}{2}\Delta f(t,B_t)\right]dt + \nabla f(t,B_t)\cdot dB_t.$$

The Itô multiplication table in this multi-dimensional setting is

$$dt\cdot dB^i = 0, \quad dt^2 = 0, \quad dB^i dB^j = \delta_{ij}dt. \tag{17.4}$$

Example 17.7 (Feynmann-Kac formula). Suppose that V is a nice function on $\mathbb{R}^d$ and $h(t,x)$ solves the partial differential equation,

$$\partial_t h = \frac{1}{2}\Delta h - Vh \text{ with } h(0,x) = f(x).$$

Given $T > 0$ let

$$M_t := h(T-t,X_t)\,e^{-\int_0^t V(X_s)ds} := h(T-t,X_t)\,Z_t.$$

Then by Itô's lemma,

$$\begin{aligned}
dM_t =& \nabla_x h(T-t,X_t)\cdot dB_t + \left(\frac{1}{2}\Delta h - \partial_t h\right)(T-t,X_t)\,Z_t dt \\
&- h(T-t,X_t)\,Z_t V(X_t)\,dt \\
=& \nabla_x h(T-t,X_t)\cdot dB_t + \left(\frac{1}{2}\Delta h - \partial_t h - Vh\right)(T-t,X_t)\,Z_t dt \\
=& \nabla_x h(T-t,X_t)\cdot dB_t.
\end{aligned}$$

From this it follows that M_t is a martingale and in particular we learn $\mathbb{E}M_0 = \mathbb{E}M_T$ from which ti follows

$$\begin{aligned}
h(T,x) &= \mathbb{E}\left[f(X_T)\,e^{-\int_0^T V(X_s)ds}\right] \\
&= \mathbb{E}\left[f(x+B_T)\,e^{-\int_0^T V(x+B_s)ds}\right].
\end{aligned}$$

Theorem 17.8. *If $u : D \to \mathbb{R}$ is a function such that $\Delta u = g$ on D and $u = f$ on ∂D, then*

$$u(x) = \mathbb{E}\left[f(x + B_\tau) - \frac{1}{2}\int_0^\tau g(x + B_s)\,ds\right]. \tag{17.5}$$

Proof. By Itô's formula,

$$\begin{aligned} d[u(X_t)] &= 1_{t\le\tau}\nabla u(X_t)\cdot dB_t + \frac{1}{2}(\Delta u)(X_t)\,1_{t\le\tau}dt \\ &= 1_{t\le\tau}\nabla u(X_t)\cdot dB_t + \frac{1}{2}g(X_t)\,1_{t\le\tau}dt, \end{aligned}$$

i.e.

$$u(X_T) = u(X_0) + \int_0^T 1_{t\le\tau}\nabla u(X_t)\cdot dB_t + \frac{1}{2}\int_0^{T\wedge\tau} g(X_t)\,dt.$$

Taking expectations of this equation and using $X_0 = x$ shows

$$\begin{aligned} \mathbb{E}[u(X_T)] &= u(x) + \mathbb{E}\int_0^T 1_{t\le\tau}\nabla u(X_t)\cdot dB_t + \frac{1}{2}\mathbb{E}\int_0^{T\wedge\tau} g(X_t)\,dt \\ &= u(x) + \frac{1}{2}\mathbb{E}\int_0^{T\wedge\tau} g(X_t)\,dt. \end{aligned}$$

Now letting $T \uparrow \infty$ gives the desired result upon observing that

$$\mathbb{E}[u(X_\infty)] = \mathbb{E}[u(x + B_\tau)] = \mathbb{E}[f(x + B_\tau)].$$

■

Notation 17.9 *In the future let us write $\mathbb{E}_x\left[F\left(B_{(\cdot)}\right)\right]$ for $\mathbb{E}\left[F\left(x + B_{(\cdot)}\right)\right]$.*

Example 17.10. Suppose $u : \bar{D} \to \mathbb{R}$ solves, $\Delta u = -2$ on D and $u = 0$ on ∂D, then Eq. (17.5) implies

$$u(x) = \mathbb{E}_x[\tau]. \tag{17.6}$$

If instead we let u solve, $\Delta u = 0$ on D and $u = f$ on ∂D, then Eq. (17.5) implies

$$u(x) = \mathbb{E}_x[f(B_\tau)].$$

Here are some technical details for Eq. (17.6). If u solves $\Delta u = -2$ on D and $u = 0$ on ∂D, then by the maximum principle applied to $-u$ we learn that $u \ge 0$ on D. So by the optional sampling theorem

$$M_t := u(B_t^\tau) + t\wedge\tau$$

is a martingale and therefore,

$$\begin{aligned} u(x) &= \mathbb{E}_x u(B_0^\tau) = \mathbb{E}_x M_0 = \mathbb{E}_x[u(B_t^\tau) + t\wedge\tau] \\ &= \mathbb{E}_x[u(B_t^\tau)] + \mathbb{E}_x[t\wedge\tau] \qquad (17.7) \\ &\ge \mathbb{E}_x[t\wedge\tau]. \qquad (17.8) \end{aligned}$$

We may now use the MCT and pass to the limit as $t \uparrow \infty$ in the inequality in Eq. (17.8) to show $\mathbb{E}_x\tau \le u(x) < \infty$. Now that we know τ is integrable and in particular $\mathbb{P}_x(\tau < \infty) = 1$, we may with the aid of DCT let $t \uparrow \infty$ in Eq. (17.7) to find,

$$u(x) = \mathbb{E}_x[u(B_\tau)] + \mathbb{E}_x\tau = \mathbb{E}_x\tau$$

wherein we have used $B_\tau \in \partial D$ where $u = 0$ in the last equality.

17.3 A Simple Option Pricing Model

In this section we are going to try to explain the Black–Scholes formula for option pricing. The following excerpt is taken from *http://en.wikipedia.org/wiki/Black-Scholes.*

> Robert C. Merton was the first to publish a paper expanding the mathematical understanding of the options pricing model and coined the term "Black-Scholes" options pricing model, by enhancing work that was published by Fischer Black and Myron Scholes. The paper was first published in 1973. The foundation for their research relied on work developed by scholars such as Louis Bachelier, A. James Boness, Sheen T. Kassouf, Edward O. Thorp, and Paul Samuelson. The fundamental insight of Black-Scholes is that the option is implicitly priced if the stock is traded.
> Merton and Scholes received the 1997 Nobel Prize in Economics for this and related work. Though ineligible for the prize because of his death in 1995, Black was mentioned as a contributor by the Swedish academy.

Definition 17.11. *A **European stock option at time T with strike price** K is a ticket that you would buy from a trader for the right to buy a particular stock at time T at a price K. If the stock price, S_T, at time T is greater that K you could then buy the stock at price K and then instantly resell it for a profit of $(S_T - K)$ dollars. If the $S_T < K$, you would not turn in your ticket but would loose whatever you paid for the ticket. So the pay off of the option at time T is $(S_T - K)_+$.*

Question: What should be the price (q) at time zero of such a stock option?

To answer this question, we will use a simplified version of a financial market which consists of only two assets; a no risk bond worth $\beta_t = \beta_0 e^{rt}$ (for some $r > 0$) dollars per share at time t and a risky stock worth S_t dollars per share. We are going to model S_t via a geometric "Brownian motion."

Definition 17.12 (Geometric Brownian Motion). *Let $\sigma > 0$, and $\mu \in \mathbb{R}$ be given parameters. We say that the solution to the "**stochastic differential equation**,"*

$$\frac{dS_t}{S_t} = \sigma dB_t + \mu dt \tag{17.9}$$

*with S_0 being non-random is a **geometric Brownian motion**. More precisely, S_t, is a solution to*

$$S_t = S_0 + \sigma \int_0^t S dB + \mu \int_0^t S_s ds. \tag{17.10}$$

*(The parameters σ and μ measure the **volatility** and **drift or trend** of the stock respectively.)*

Notice that $\frac{dS}{S}$ is the relative change of S and formally, $\mathbb{E}\left(\frac{dS}{S}\right) = \mu dt$ and $\mathrm{Var}\left(\frac{dS}{S}\right) = \sigma^2 dt$. Taking expectation of Eq. (17.10) gives,

$$\mathbb{E}S_t = S_0 + \mu \int_0^t \mathbb{E}S_s ds.$$

Differentiating this equation then implies,

$$\frac{d}{dt}\mathbb{E}S_t = \mu \mathbb{E}S_t \text{ with } \mathbb{E}S_0 = S_0,$$

which yields, $\mathbb{E}S_t = S_0 e^{\mu t}$. So on average, S_t is growing or decaying exponentially depending on the sign of μ.

Proposition 17.13 (Geometric Brownian motion). *The stochastic differential Equation (17.10) has a unique solution given by*

$$S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right).$$

Proof. We do not bother to give the proof of uniqueness here. To prove existence, let us look for a solution to Eq. (17.9) of the form;

$$S_t = S_0 \exp\left(aB_t + bt\right),$$

for some constants a and b. By Itô's lemma, using $\frac{d}{dx}e^x = \frac{d^2}{dx^2}e^x = e^x$ and the multiplication rules, $dB^2 = dt$ and $dt^2 = dB \cdot dt = 0$, we find that

$$\begin{aligned} dS &= S\left(adB + bdt\right) + \frac{1}{2}S\left(adB + bdt\right)^2 \\ &= S\left(adB + bdt\right) + \frac{1}{2}Sa^2 dt, \end{aligned}$$

i.e.

$$\frac{dS}{S} = adB + \left(b + \frac{1}{2}a^2\right) dt.$$

Comparing this with Eq. (17.9) shows that we should take $a = \sigma$ and $b = \mu - \frac{1}{2}\sigma^2$ to get a solution. ■

Definition 17.14 (Holdings and Value Processes). *Let (a_t, b_t) be the **holdings process** which denotes the number of shares of stock and bonds respectively that are held in the portfolio at time t. The **value process**, V_t, of the portfolio, is*

$$V_t = a_t S_t + b_t \beta_t. \tag{17.11}$$

Suppose time is partitioned as,

$$\Pi = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = T\}$$

for some time T in the future. Let us suppose that (a_t, b_t) is constant on the intervals, $[0, t_1]$, $(t_1, t_2]$, $\ldots, (t_{n-1}, t_n]$. Let us write $(a_t, b_t) = (a_{i-1}, b_{i-1})$ for $t_{i-1} < t \le t_i$, see Figure 17.1.

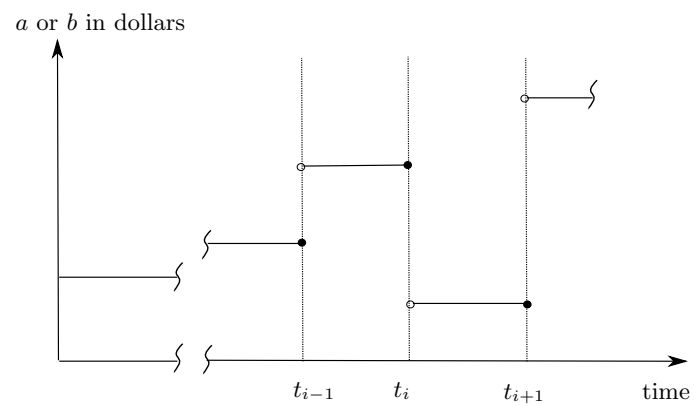


Fig. 17.1. A possible graph of either a_t or b_t.

Therefore the value of the portfolio is given by

$$V_t = a_{i-1} S_t + b_{i-1} \beta_t \text{ for } t_{i-1} < t \le t_i.$$

We now assume that our holding process is **self financing** (i.e. we do not add any external money to portfolio other than what was invested, $V_0 = a_0 S_0 + b_0 \beta_0$, at the initial time $t = 0$), then we must have[2]

[2] Equation (17.12) may be written as

$$a_{i-1}S_{t_i} + b_{i-1}\beta_{t_i} = V_{t_i} = a_i S_{t_i} + b_i \beta_{t_i} \text{ for all } i. \tag{17.12}$$

That is to say, when we rebalance our portfolio at time t_i, we are only use the money, V_{t_i}, dollars in the portfolio at time t_i. Using Eq. (17.12) at i and $i-1$ allows us to conclude,

$$\begin{aligned} V_{t_i} - V_{t_{i-1}} &= a_{i-1}S_{t_i} + b_{i-1}\beta_{t_i} - \left(a_{i-1}S_{t_{i-1}} + b_{i-1}\beta_{t_{i-1}}\right) \\ &= a_{i-1}\left(S_{t_i} - S_{t_{i-1}}\right) + b_{i-1}\left(\beta_{t_i} - \beta_{t_{i-1}}\right) \text{ for all } i, \end{aligned} \tag{17.13}$$

which states the change of the portfolio balance over the time interval, $(t_{i-1}, t_i]$ is due solely to the gain or loss made by the investments in the portfolio. (The Equations (17.12) and (17.13) are equivalent.) Summing Eq. (17.13) then gives,

$$V_{t_j} - V_0 = \sum_{i=1}^{j} a_{i-1}\left(S_{t_i} - S_{t_{i-1}}\right) + \sum_{i=1}^{j} b_{i-1}\left(\beta_{t_i} - \beta_{t_{i-1}}\right) \tag{17.14}$$

$$= \int_0^{t_j} a_r dS_r + \int_0^{t_j} b_r d\beta_r \text{ for all } j. \tag{17.15}$$

More generally, if we throw any arbitrary point, $t \in [0,T]$, into our partition we may conclude that

$$V_t = V_0 + \int_0^t a dS + \int_0^t b d\beta \text{ for all } 0 \le t \le T. \tag{17.16}$$

The interpretation of this equation is that $V_t - V_0$ is equal to the gains or losses due to trading which is given by

$$\int_0^t a dS + \int_0^t b d\beta.$$

Equation (17.16) now makes sense even if we allow for continuous trading. The previous arguments show that the integrals appearing in Eq. (17.16) should be taken to be Itô – integrals as defined in Definition 17.1. Moreover, if the investor does not have psychic abilities, we should assume that holding process is adapted.

$$(a_i - a_{i-1})S_{t_i} + (b_i - b_{i-1})\beta_{t_i} = 0.$$

This explains why the continuum limit of this equation is not $S_t da_t + \beta_t db_t = 0$ but rather must be interpreted as $S_{t+dt}da_t + \beta_{t+dt}db_t = 0$. It is also useful to observe that

$$\begin{aligned} d(XY)_t &= X_{t+dt}Y_{t+dt} - X_t Y_t \\ &= (X_{t+dt} - X_t)Y_{t+dt} + X_t(Y_{t+dt} - Y_t), \end{aligned}$$

and hence there is no quadratic differential term when $d(XY)$ is written out this way.

17.4 The Black-Scholes Formula

Now that we have set the stage we can now try to price the option. (We will closely follow [3, p. 255-264.] here.)

Fundamental Principle: The price of the option should be equal to the amount of money, V_0, that an investor would have to put into the bond-stock market at time $t = 0$ so as there exists a self-financing holding process (a_t, b_t), such that

$$V_T = a_T S_T + b_T \beta_T = (S_T - K)_+.$$

Remark 17.15 (Money for nothing). If we price the option higher than V_0, i.e. $q > V_0$, we could make risk free money by selling one of these options at q dollars, investing $V_0 < q$ of this money using the holding process (a_t, b_t) to cover the payoff at time T and then pocket the different, $q - V_0$.

If the price of the option was less than V_0, i.e. $q < V_0$, the investor should buy the option and then pursue the trading strategy, $(-a, -b)$. At time zero the investor has invested $q + (-a_0 S_0 - b_0\beta_0) = q - V_0 < 0$ dollars, i.e. he is holding $V_0 - q$ dollars in hand at time $t = 0$. The value of his portfolio at time T is now $-V_T = -(S_T - K)_+$. If $S_T > K$, the investor then exercises her option to pay off the debt she as accrued in the portfolio and if $S_T \le K$, she does nothing since his portfolio is worth zero dollars. Either way, she still has the $V_0 - q$ dollars in hand from the start of the transactions at $t = 0$.

Ansatz: We would like the price of the option $q = V_0$ to depend only on what we might know at the initial time. Thus we make the ansatz that $q := f\left(S_0, T, K, r, \sigma^2, \mu\right)$[3]. (It is part of the content of the Black-Scholes formula that this ansatz is permissible.)

If we have a self-financing holding process (a_t, b_t), then $\{(a_s, b_s)\}_{t \le s \le T}$ is also a self-financing holding process on $[t,T]$ such that $V_T = a_T S_T + b_T \beta_T = (S_T - K)_+$, therefore, given the ansatz and the fundamental principle above, if the stock price is S_t at time t, the options price at this time should be

$$V_t = f(S_t, T-t) \text{ for all } 0 \le t \le T. \tag{17.17}$$

By Itô's lemma

$$\begin{aligned} dV_t &= f_x(S_t, T-t)\,dS_t + \frac{1}{2} f_{xx}(S_t, T-t)\,dS_t^2 - f_t(S_t, T-t)\,dt \\ &= f_x(S_t, T-t)\,S_t(\sigma dB_t + \mu dt) + \left[\frac{1}{2} f_{xx}(S_t, T-t)\,S_t^2\sigma^2 - f_t(S_t, T-t)\right] dt \\ &= f_x(S_t, T-t)\,S_t\sigma dB_t \\ &\quad + \left[f_x(S_t, T-t)\,S_t\mu + \frac{1}{2} f_{xx}(S_t, T-t)\,S_t^2\sigma^2 - f_t(S_t, T-t)\right] dt \end{aligned}$$

[3] Since r, K, μ, and σ^2 are fixed, we will often drop them from the notation.

On the other hand from Eqs. (17.16) and (17.9), we know that

$$\begin{aligned} dV_t &= a_t dS_t + b_t \beta_0 r e^{rt} dt \\ &= a_t S_t (\sigma dB_t + \mu dt) + b_t \beta_0 r e^{rt} dt \\ &= a_t S_t \sigma dB_t + \left[a_t S_t \mu + b_t \beta_0 r e^{rt}\right] dt. \end{aligned}$$

Comparing these two equations implies,

$$a_t = f_x (S_t, T-t) \tag{17.18}$$

and

$$\begin{aligned} &a_t S_t \mu + b_t \beta_0 r e^{rt} \\ &\quad = f_x (S_t, T-t) S_t \mu + \frac{1}{2} f_{xx} (S_t, T-t) S_t^2 \sigma^2 - f_t (S_t, T-t). \end{aligned} \tag{17.19}$$

Using Eq. (17.18) and

$$\begin{aligned} f (S_t, T-t) = V_t &= a_t S_t + b_t \beta_0 e^{rt} \\ &= f_x (S_t, T-t) S_t + b_t \beta_0 e^{rt} \end{aligned}$$

in Eq. (17.19) allows us to conclude,

$$\begin{aligned} \frac{1}{2} f_{xx} (S_t, T-t) S_t^2 \sigma^2 - f_t (S_t, T-t) &= r b_t \beta_0 e^{rt} \\ &= r f (S_t, T-t) - r f_x (S_t, T-t) S_t. \end{aligned}$$

Thus we see that the unknown function f should solve the partial differential equation,

$$\begin{aligned} \frac{1}{2} \sigma^2 x^2 f_{xx} (x, T-t) - f_t (x, T-t) &= r f (x, T-t) - r x f_x (x, T-t) \\ \text{with } f (x, 0) &= (x-K)_+ , \end{aligned}$$

i.e.

$$f_t (x,t) = \frac{1}{2} \sigma^2 x^2 f_{xx} (x,t) + r x f_x (x,t) - r f (x,t) \tag{17.20}$$

$$\text{with } f (x,0) = (x-K)_+ . \tag{17.21}$$

Fact 17.16 *Let N be a standard normal random variable and $\Phi (x) := \mathbb{P} (N \le x)$. The solution to Eqs. (17.20) and (17.21) is given by;*

$$f (x,t) = x \Phi (g (x,t)) - K e^{-rt} \Phi (h (x,t)), \tag{17.22}$$

where,

$$\begin{aligned} g (x,t) &= \frac{\ln (x/K) + \left(r + \frac{1}{2}\sigma^2\right) t}{\sigma \sqrt{t}}, \\ h (x,t) &= g (x,t) - \sigma \sqrt{t}. \end{aligned}$$

Theorem 17.17 (Option Pricing). *Given the above setup, the "rational" price" of the European call option is*

$$\begin{aligned} q = S_0 \Phi &\left(\frac{\ln (S_0/K) + \left(r + \frac{1}{2}\sigma^2\right) T}{\sigma \sqrt{T}} \right) \\ &- K e^{-rT} \Phi \left(\frac{\ln (S_0/K) + \left(r + \frac{1}{2}\sigma^2\right) T}{\sigma \sqrt{T}} - \sigma \sqrt{T} \right) \end{aligned}$$

where $\Phi (x) := \mathbb{P} (N (0,1) \le x)$.

Part V

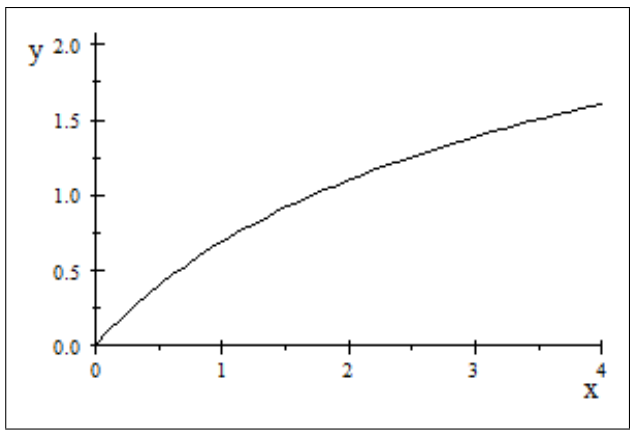
Appendix

Analytic Facts

18.1 A Stirling's Formula Like Approximation

Theorem 18.1. *Suppose that $f:(0,\infty)\to\mathbb{R}$ is an increasing concave down function (like $f(x)=\ln x$) and let $s_n:=\sum_{k=1}^n f(k)$, then*

$$\begin{aligned} s_n-\frac{1}{2}(f(n)+f(1)) &\le \int_1^n f(x)\,dx \\ &\le s_n-\frac{1}{2}[f(n+1)+2f(1)]+\frac{1}{2}f(2) \\ &\le s_n-\frac{1}{2}[f(n)+2f(1)]+\frac{1}{2}f(2). \end{aligned}$$



Proof. On the interval, $[k-1,k]$, we have that $f(x)$ is larger than the straight line segment joining $(k-1,f(k-1))$ and $(k,f(k))$ and thus

$$\frac{1}{2}(f(k)+f(k-1))\le\int_{k-1}^k f(x)\,dx.$$

Summing this equation on $k=2,\dots,n$ shows,

$$\begin{aligned} s_n-\frac{1}{2}(f(n)+f(1)) &= \sum_{k=2}^n\frac{1}{2}(f(k)+f(k-1)) \\ &\le \sum_{k=2}^n\int_{k-1}^k f(x)\,dx=\int_1^n f(x)\,dx. \end{aligned}$$

For the upper bound on the integral we observe that $f(x)\le f(k)-f'(k)(x-k)$ for all x and therefore,

$$\int_{k-1}^k f(x)\,dx\le\int_{k-1}^k[f(k)-f'(k)(x-k)]\,dx=f(k)-\frac{1}{2}f'(k).$$

Summing this equation on $k=2,\dots,n$ then implies,

$$\int_1^n f(x)\,dx\le\sum_{k=2}^n f(k)-\frac{1}{2}\sum_{k=2}^n f'(k).$$

Since $f''(x)\le 0$, $f'(x)$ is decreasing and therefore $f'(x)\le f'(k-1)$ for $x\in[k-1,k]$ and integrating this equation over $[k-1,k]$ gives

$$f(k)-f(k-1)\le f'(k-1).$$

Summing the result on $k=3,\dots,n+1$ then shows,

$$f(n+1)-f(2)\le\sum_{k=2}^n f'(k)$$

and thus ti follows that

$$\begin{aligned} \int_1^n f(x)\,dx &\le \sum_{k=2}^n f(k)-\frac{1}{2}(f(n+1)-f(2)) \\ &= s_n-\frac{1}{2}[f(n+1)+2f(1)]+\frac{1}{2}f(2) \\ &\le s_n-\frac{1}{2}[f(n)+2f(1)]+\frac{1}{2}f(2) \end{aligned}$$

■

Example 18.2 (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

$$s_n - \frac{1}{2}\ln n \le n \ln n - n + 1 \le s_n - \frac{1}{2}\ln n + \frac{1}{2}\ln 2.$$

Thus it follows that

$$\left(n + \frac{1}{2}\right)\ln n - n + 1 - \ln\sqrt{2} \le s_n \le \left(n + \frac{1}{2}\right)\ln n - n + 1.$$

Exponentiating this identity then gives the following upper and lower bounds on $n!$;

$$\frac{e}{\sqrt{2}} \cdot e^{-n} n^{n+1/2} \le n! \le e \cdot e^{-n} n^{n+1/2}.$$

These bound compare well with Strirling's formula (Theorem 18.5) which implies,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2} \overset{\text{by definition}}{\Longleftrightarrow} \lim_{n\to\infty} \frac{n!}{e^{-n} n^{n+1/2}} = \sqrt{2\pi}.$$

Observe that

$$\frac{e}{\sqrt{2}} \cong 1.922\,1 \le \sqrt{2\pi} \cong 2.506 \le e \cong 2.718\,3.$$

Definition 18.3 (Gamma Function). *The **Gamma function**, $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by*

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \tag{18.1}$$

(The reader should check that $\Gamma(x) < \infty$ for all $x > 0$.)

Here are some of the more basic properties of this function.

Example 18.4 (Γ – function properties). Let Γ 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;

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-u}\, u^{x+1}\, \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\, e^{-u}\right) du \\ &= x \int_0^\infty u^{x-1}\, e^{-u}\, du = x\, \Gamma(x).\end{aligned}$$

In particular, it follows from items 1. and 2. and induction that

$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{18.2}$$

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 19.1) that

$$\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{18.3}$$

Taking $a = 1$ and making the change of variables, $u = r^2$ below implies,

$$\sqrt{\pi} = \int_{-\infty}^\infty e^{-r^2} dr = 2\int_0^\infty u^{-1/2} e^{-u} du = \Gamma(1/2).$$

$$\begin{aligned}\Gamma(1/2) &= 2\int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr \\ &= I_1(1) = \sqrt{\pi}.\end{aligned}$$

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 $n \in \mathbb{N}$.

Theorem 18.5 (Stirling's formula). *The Gamma function (see Definition 18.3), satisfies Stirling's formula,*

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{\sqrt{2\pi} e^{-x} x^{x+1/2}} = 1. \tag{18.4}$$

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, $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

19

Multivariate Gaussians

The following basic Gaussian integration formula is needed in order to properly normalize Gaussian densities, see Definition 19.2 below.

Lemma 19.1. *Let $a > 0$ and for $d \in \mathbb{N}$ let,*

$$I_d(a) := \int_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

Then $I_d(a) = (\pi/a)^{d/2}$.

Proof. By Tonelli's theorem and induction,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \end{aligned} \tag{19.1}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)} dx_1 dx_2.$$

Using polar coordinates, we find,

$$\begin{aligned} I_2(a) &= \int_0^\infty dr\ r \int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi \int_0^\infty r e^{-ar^2} dr \\ &= 2\pi \lim_{M\to\infty} \int_0^M r e^{-ar^2} dr = 2\pi \lim_{M\to\infty} \frac{e^{-ar^2}}{-2a} \Big|_0^M = \frac{2\pi}{2a} = \pi/a. \end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (19.1). ■

19.1 Review of Gaussian Random Variables

Definition 19.2 (Normal / Gaussian Random Variable). *A random variable, Y, **is normal with mean μ standard deviation σ^2** iff*

$$\mathbb{P}(Y \in (y, y+dy]) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy. \tag{19.2}$$

*We will abbreviate this by writing $Y \overset{d}{=} N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ we say Y is a **standard normal** random variable. We will often denote standard normal random variables by Z.*

Observe that Eq. (19.2) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy$$

for all bounded functions, $f : \mathbb{R} \to \mathbb{R}$. Also observe that $Y \overset{d}{=} N(\mu, \sigma^2)$ is equivalent to $Y \overset{d}{=} \sigma Z + \mu$. Indeed, by making the change of variable, $y = \sigma x + \mu$, we find

$$\begin{aligned} \mathbb{E}[f(\sigma Z + \mu)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\sigma x + \mu) e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy. \end{aligned}$$

Lastly the constant, $\left(2\pi\sigma^2\right)^{-1/2}$ is chosen so that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy = 1.$$

Lemma 19.3 (Integration by parts). *If $X \overset{d}{=} N(0, \sigma^2)$ for some $\sigma^2 \geq 0$ and $f : \mathbb{R} \to \mathbb{R}$ is a C^1 – function such that $Xf(X)$, $f'(X)$ and $f(X)$ are all integrable random variables and*[1] $\lim_{z\to\pm\infty} \left[f(x) e^{-\frac{1}{2\sigma^2}x^2}\right] = 0$, *then*

$$\mathbb{E}[Xf(X)] = \sigma^2 \mathbb{E}[f'(X)] = \mathbb{E}\left[X^2\right] \cdot \mathbb{E}[f'(X)]. \tag{19.3}$$

Proof. If $\sigma = 0$ then $X = 0$ a.s. and both sides of Eq. (19.3) are zero. So we now suppose that $\sigma > 0$ and set $C := 1/\sqrt{2\pi\sigma^2}$. The result is a simple matter of using integration by parts;

[1] This last hypothesis is actually unnecessary!

$$\begin{aligned}\mathbb{E}\left[f'(X)\right] &= C\int_{\mathbb{R}} f'(x)\, e^{-\frac{1}{2\sigma^2}x^2}dx = C\lim_{M\to\infty}\int_{-M}^{M} f'(x)\, e^{-\frac{1}{2\sigma^2}x^2}dx \\ &= C\lim_{M\to\infty}\left[f(x)\, e^{-\frac{1}{2\sigma^2}x^2}\Big|_{-M}^{M} - \int_{-M}^{M} f(x)\frac{d}{dx}e^{-\frac{1}{2\sigma^2}x^2}dx\right] \\ &= C\lim_{M\to\infty}\int_{-M}^{M} f(x)\frac{x}{\sigma^2}e^{-\frac{1}{2}x^2}dx = \frac{1}{\sigma^2}\mathbb{E}\left[Xf(X)\right].\end{aligned}$$

■

Example 19.4. Suppose that $X \overset{d}{=} N(0,1)$ and define $\alpha_k := \mathbb{E}\left[X^{2k}\right]$ for all $k \in \mathbb{N}_0$. By Lemma 19.3,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1}\cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

Hence it follows that

$$\alpha_1 = \alpha_0 = 1,\ \alpha_2 = 3\alpha_1 = 3,\ \alpha_3 = 5\cdot 3$$

and by a simple induction argument,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!,$$

where $(-1)!! := 0$.

Actually we can use the Γ – function to say more. Namely for any $\beta > -1$,

$$\mathbb{E}\left|X\right|^{\beta} = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}|x|^{\beta}\, e^{-\frac{1}{2}x^2}dx = \sqrt{\frac{2}{\pi}}\int_0^{\infty} x^{\beta}e^{-\frac{1}{2}x^2}dx.$$

Now make the change of variables, $y = x^2/2$ (i.e. $x = \sqrt{2y}$ and $dx = \frac{1}{\sqrt{2}}y^{-1/2}dy$) to learn,

$$\begin{aligned}\mathbb{E}\left|X\right|^{\beta} &= \frac{1}{\sqrt{\pi}}\int_0^{\infty}(2y)^{\beta/2}\, e^{-y}y^{-1/2}dy \\ &= \frac{1}{\sqrt{\pi}}2^{\beta/2}\int_0^{\infty} y^{(\beta+1)/2}e^{-y}y^{-1}dy = \frac{1}{\sqrt{\pi}}2^{\beta/2}\Gamma\left(\frac{\beta+1}{2}\right).\end{aligned}$$

Exercise 19.1. Let $q(x)$ be a polynomial[2] in x, $Z \overset{d}{=} N(0,1)$, and

$$u(t,x) := \mathbb{E}\left[q\left(x+\sqrt{t}Z\right)\right] \tag{19.4}$$

$$= \int_{\mathbb{R}}\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}(y-x)^2}q(y)\,dy \tag{19.5}$$

[2] Actually, $q(x)$ can be any twice continuously differentiable function which along with its derivatives grow slower than $e^{\varepsilon x^2}$ for any $\varepsilon > 0$.

Show u satisfies the heat equation,

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x) \text{ for all } t > 0 \text{ and } x \in \mathbb{R},$$
$$\text{with } u(0,x) = q(x).$$

Hints: Make use of Lemma 19.3 along with the fact (which is easily proved here) that

$$\frac{\partial}{\partial t}u(t,x) = \mathbb{E}\left[\frac{\partial}{\partial t}q\left(x+\sqrt{t}Z\right)\right].$$

You will also have to use the corresponding fact for the x derivatives as well.

Exercise 19.2. Let $q(x)$ be a polynomial in x, $Z \overset{d}{=} N(0,1)$, and $\Delta = \frac{d^2}{dx^2}$. Show

$$\mathbb{E}\left[q(Z)\right] = \left(e^{\Delta/2}q\right)(0) := \sum_{n=0}^{\infty}\frac{1}{n!}\left(\left(\frac{\Delta}{2}\right)^n q\right)(0) = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{1}{2^n}\left(\Delta^n q\right)(0)$$

where the above sum is actually a finite sum since $\Delta^n q \equiv 0$ if $2n > \deg q$.
Hint: let $u(t) := \mathbb{E}\left[q\left(\sqrt{t}Z\right)\right]$. From your proof of Exercise 19.1 you should be able to see that $\dot{u}(t) = \frac{1}{2}\mathbb{E}\left[(\Delta q)\left(\sqrt{t}Z\right)\right]$. This latter equation may be iterated in order to find $u^{(n)}(t)$ for all $n \geq 0$. With this information in hand you should be able to finish the proof with the aid of Taylor's theorem.

Example 19.5. Suppose that $k \in \mathbb{N}$, then

$$\begin{aligned}\mathbb{E}\left[Z^{2k}\right] &= \left(e^{\frac{\Delta}{2}}x^{2k}\right)|_{x=0} = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{1}{2^n}\left(\Delta^n x^{2k}\right)|_{x=0} \\ &= \frac{1}{k!}\frac{1}{2^k}\Delta^k x^{2k} = \frac{(2k)!}{k!2^k} \\ &= \frac{2k\cdot(2k-1)\cdot 2(k-1)\cdot(2k-3)\cdot\cdots\cdot(2\cdot 2)\cdot 3\cdot 2\cdot 1}{2^k k!} \\ &= (2k-1)!!\end{aligned}$$

in agreement with Example 19.4.

Example 19.6. Let Z be a standard normal random variable and set $f(\lambda) := \mathbb{E}\left[e^{\lambda Z^2}\right]$ for $\lambda < 1/2$. Then $f(0) = 1$ and

$$\begin{aligned}f'(\lambda) &= \mathbb{E}\left[Z^2e^{\lambda Z^2}\right] = \mathbb{E}\left[\frac{\partial}{\partial Z}\left(Ze^{\lambda Z^2}\right)\right] \\ &= \mathbb{E}\left[e^{\lambda Z^2} + 2\lambda Z^2e^{\lambda Z^2}\right] \\ &= f(\lambda) + 2\lambda f'(\lambda).\end{aligned}$$

Solving for λ we find,

$$f'(\lambda) = \frac{1}{1-2\lambda} f(\lambda) \text{ with } f(0) = 1.$$

The solution to this equation is found in the usual way as,

$$\ln f(\lambda) = \int \frac{f'(\lambda)}{f(\lambda)} d\lambda = \int \frac{1}{1-2\lambda} d\lambda = -\frac{1}{2} \ln(1-2\lambda) + C.$$

By taking $\lambda = 0$ using $f(0) = 1$ we find that $C = 0$ and therefore,

$$\mathbb{E}\left[e^{\lambda Z^2}\right] = f(\lambda) = \frac{1}{\sqrt{1-2\lambda}} \text{ for } \lambda < \frac{1}{2}.$$

This can also be shown by directly evaluating the integral,

$$\mathbb{E}\left[e^{\lambda Z^2}\right] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2\lambda)z^2} dz.$$

Exercise 19.3. Suppose that $Z \overset{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda Z}\right] = \exp\left(-\lambda^2/2\right). \tag{19.6}$$

Hint: You may use without proof that $f'(\lambda) = i\mathbb{E}\left[Ze^{i\lambda Z}\right]$ (i.e. it is permissible to differentiate past the expectation.) Assuming this use Lemma 19.3 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Lemma 19.7 (Gaussian tail estimates). *Suppose that X is a standard normal random variable, i.e.*

$$P(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}},$$

then for all $x \geq 0$,

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right) \leq \frac{1}{2} e^{-x^2/2}. \tag{19.7}$$

Moreover (see [15, Lemma 2.5]),

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \tag{19.8}$$

which combined with Eq. (19.7) proves Mill's ratio (see [9]);

$$\lim_{x\to\infty} \frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}} = 1. \tag{19.9}$$

Proof. See Figure 19.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right),$$

the red is the plot of $\frac{1}{2} e^{-x^2/2}$, and the blue is the plot of

$$\max\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right).$$

The formal proof of these estimates for the reader who is not convinced by

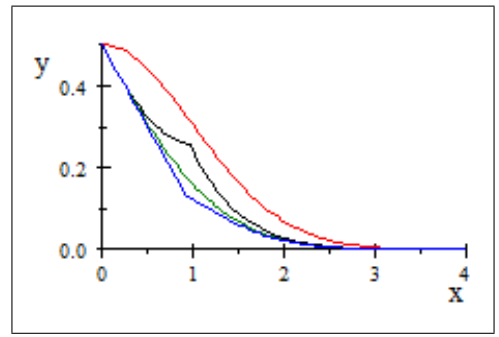


Fig. 19.1. Plots of $P(X \geq x)$ and its estimates.

Figure 19.1 is given below.

We begin by observing that

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2} dy \\ &\leq -\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-y^2/2}\Big|_x^{-\infty} = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \end{aligned} \tag{19.10}$$

If we only want to prove Mill's ratio (19.9), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \\ &\geq \frac{1}{\sqrt{2\pi}} \int_x^{\alpha x} \frac{y}{\alpha x} e^{-y^2/2} dy = -\frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-y^2/2}\Big|_{y=x}^{y=\alpha x} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-x^2/2} \left[1 - e^{-\alpha^2 x^2/2}\right] \end{aligned}$$

from which it follows,

$$\liminf_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P(X\ge x)\right]\ge 1/\alpha\uparrow 1 \text{ as } \alpha\downarrow 1.$$

The estimate in Eq. (19.10) shows $\limsup_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P(X\ge x)\right]\le 1$.

To get more precise estimates, we begin by observing,

$$\begin{aligned}P(X\ge x)&=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy \qquad (19.11)\\ &\le\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dy\le\frac{1}{2}-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}x.\end{aligned}$$

This equation along with Eq. (19.10) gives the first equality in Eq. (19.7). To prove the second equality observe that $\sqrt{2\pi}>2$, so

$$\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}\le\frac{1}{2}e^{-x^2/2} \text{ if } x\ge 1.$$

For $x\le 1$ we must show,

$$\frac{1}{2}-\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\le\frac{1}{2}e^{-x^2/2}$$

or equivalently that $f(x):=e^{x^2/2}-\sqrt{\frac{2}{\pi}}x\le 1$ for $0\le x\le 1$. Since f is convex $\left(f''(x)=\left(x^2+1\right)e^{x^2/2}>0\right)$, $f(0)=1$ and $f(1)\cong 0.85<1$, it follows that $f\le 1$ on $[0,1]$. This proves the second inequality in Eq. (19.7).

It follows from Eq. (19.11) that

$$\begin{aligned}P(X\ge x)&=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy\\ &\ge\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x 1dy=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}x \text{ for all } x\ge 0.\end{aligned}$$

So to finish the proof of Eq. (19.8) we must show,

$$\begin{aligned}f(x)&:=\frac{1}{\sqrt{2\pi}}xe^{-x^2/2}-\left(1+x^2\right)P(X\ge x)\\ &=\frac{1}{\sqrt{2\pi}}\left[xe^{-x^2/2}-\left(1+x^2\right)\int_x^\infty e^{-y^2/2}dy\right]\le 0 \text{ for all } 0\le x<\infty.\end{aligned}$$

This follows by observing that $f(0)=-1/2<0$, $\lim_{x\uparrow\infty}f(x)=0$ and

$$\begin{aligned}f'(x)&=\frac{1}{\sqrt{2\pi}}\left[e^{-x^2/2}\left(1-x^2\right)-2xP(X\ge x)+\left(1+x^2\right)e^{-x^2/2}\right]\\ &=2\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}-xP(X\ge y)\right)\ge 0,\end{aligned}$$

where the last inequality is a consequence Eq. (19.7). ■

19.2 Gaussian Random Vectors

Definition 19.8 (Gaussian Random Vectors). *A random vector, $X\in\mathbb{R}^d$, is Gaussian iff*

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right]=\exp\left(-\frac{1}{2}\operatorname{Var}(\lambda\cdot X)+i\mathbb{E}(\lambda\cdot X)\right) \text{ for all } \lambda\in\mathbb{R}^d. \qquad (19.12)$$

In short, X is a Gaussian random vector iff $\lambda\cdot X$ is a Gaussian random variable for all $\lambda\in\mathbb{R}^d$. (Implicitly in this definition we are assuming that $\mathbb{E}\left|X_j^2\right|<\infty$ for $1\le j\le d$.)

Notation 19.9 *Let X be a random vector in $\mathbb{R}^d$ with second moments, i.e. $\mathbb{E}\left[X_k^2\right]<\infty$ for $1\le k\le d$. The mean X is the vector $\mu=(\mu_1,\dots,\mu_d)^{\mathrm{tr}}\in\mathbb{R}^d$ with $\mu_k:=\mathbb{E}X_k$ for $1\le k\le d$ and the covariance matrix $C=C(X)$ is the $d\times d$ matrix with entries,*

$$C_{kl}:=\operatorname{Cov}(X_k,X_l) \text{ for } 1\le k,l\le d. \qquad (19.13)$$

Exercise 19.4. Suppose that X is a random vector in $\mathbb{R}^d$ with second moments. Show for all $\lambda=(\lambda_1,\dots,\lambda_d)^{\mathrm{tr}}\in\mathbb{R}^d$ that

$$\mathbb{E}[\lambda\cdot X]=\lambda\cdot\mu \text{ and } \operatorname{Var}(\lambda\cdot X)=\lambda\cdot C\lambda. \qquad (19.14)$$

Corollary 19.10. *If $Y\overset{d}{=}N\left(\mu,\sigma^2\right)$, then*

$$\mathbb{E}\left[e^{i\lambda Y}\right]=\exp\left(-\frac{1}{2}\lambda^2\sigma^2+i\mu\lambda\right) \text{ for all } \lambda\in\mathbb{R}. \qquad (19.15)$$

Conversely if Y is a random variable such that Eq. (19.15) holds, then $Y\overset{d}{=}N\left(\mu,\sigma^2\right)$.

Proof. ($\Longrightarrow$) From the remarks after Lemma 19.2, we know that $Y\overset{d}{=}\sigma Z+\mu$ where $Z\overset{d}{=}N(0,1)$. Therefore,

$$\mathbb{E}\left[e^{i\lambda Y}\right]=\mathbb{E}\left[e^{i\lambda(\sigma Z+\mu)}\right]=e^{i\lambda\mu}\mathbb{E}\left[e^{i\lambda\sigma Z}\right]=e^{i\lambda\mu}e^{-\frac{1}{2}(\lambda\sigma)^2}=\exp\left(-\frac{1}{2}\lambda^2\sigma^2+i\mu\lambda\right).$$

($\Longleftarrow$) This follows from the basic fact that the characteristic function or Fourier transform of a distribution uniquely determines the distribution. ■

Remark 19.11 (Alternate characterization of being Gaussian). Given Corollary 19.10, we have Y is a Gaussian random variable iff $\mathbb{E}Y^2<\infty$ and

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda Y}\right]&=\exp\left(-\frac{1}{2}\operatorname{Var}(\lambda Y)+i\lambda\mathbb{E}Y\right)\\ &=\exp\left(-\frac{\lambda^2}{2}\operatorname{Var}(Y)+i\lambda\mathbb{E}Y\right) \text{ for all } \lambda\in\mathbb{R}.\end{aligned}$$

Exercise 19.5. Suppose X_1 and X_2 are two independent Gaussian random variables with $X_i \overset{d}{=} N\left(0,\sigma_i^2\right)$ for $i=1,2$. Show X_1+X_2 is Gaussian and $X_1+X_2 \overset{d}{=} N\left(0,\sigma_1^2+\sigma_2^2\right)$. (**Hint:** use Remark 19.11.)

Exercise 19.6. Suppose that $Z \overset{d}{=} N(0,1)$ and $t\in\mathbb{R}$. Show $\mathbb{E}\left[e^{tZ}\right] = \exp\left(t^2/2\right)$. (You could follow the hint in Exercise 19.3 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 19.7. Use Exercise 19.6 to give another proof that $\mathbb{E}Z^{2k}=(2k-1)!!$ when $Z\overset{d}{=}N(0,1)$.

Exercise 19.8. Let $Z\overset{d}{=}N(0,1)$ and $\alpha\in\mathbb{R}$, find $\rho:\mathbb{R}_+\to\mathbb{R}_+ := (0,\infty)$ such that

$$\mathbb{E}\left[f\left(|Z|^\alpha\right)\right] = \int_{\mathbb{R}_+} f(x)\,\rho(x)\,dx$$

for all continuous functions, $f:\mathbb{R}_+\to\mathbb{R}$ with compact support in $\mathbb{R}_+$.

In particular a random vector (X) in $\mathbb{R}^d$ with second moments a Gaussian random vector iff

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = \exp\left(-\frac{1}{2}C\lambda\cdot\lambda + i\mu\cdot\lambda\right) \text{ for all } \lambda\in\mathbb{R}^d. \tag{19.16}$$

We abbreviate Eq. (19.16) by writing $X\overset{d}{=}N(\mu,C)$. Notice that it follows from Eq. (19.13) that $C^{\mathrm{tr}}=C$ and from Eq. (19.14) that $C\ge 0$, i.e. $\lambda\cdot C\lambda\ge 0$ for all $\lambda\in\mathbb{R}^d$.

Definition 19.12. *Given a Gaussian random vector, X, we call the pair, (C,μ) appearing in Eq. (19.16) the* ***characteristics*** *of X.*

Lemma 19.13. *Suppose that $X=\sum_{l=1}^k Z_l v_l+\mu$ where $\{Z_l\}_{l=1}^k$ are i.i.d. standard normal random variables, $\mu\in\mathbb{R}^d$ and $v_l\in\mathbb{R}^d$ for $1\le l\le k$. Then $X\overset{d}{=}N(\mu,C)$ where $C=\sum_{l=1}^k v_l v_l^{\mathrm{tr}}$.*

Proof. Using the basic properties of independence and normal random variables we find

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda\cdot X}\right] &= \mathbb{E}\left[e^{i\sum_{l=1}^k Z_l\lambda\cdot v_l + i\lambda\cdot\mu}\right] = e^{i\lambda\cdot\mu}\prod_{l=1}^k \mathbb{E}\left[e^{iZ_l\lambda\cdot v_l}\right] = e^{i\lambda\cdot\mu}\prod_{l=1}^k e^{-\frac{1}{2}(\lambda\cdot v_l)^2}\\ &= \exp\left(-\frac{1}{2}\sum_{l=1}^k(\lambda\cdot v_l)^2 + i\lambda\cdot\mu\right).\end{aligned}$$

Since

$$\sum_{l=1}^k(\lambda\cdot v_l)^2 = \sum_{l=1}^k \lambda\cdot v_l\left(v_l^{\mathrm{tr}}\lambda\right) = \lambda\cdot\left(\sum_{l=1}^k v_l v_l^{\mathrm{tr}}\right)\lambda$$

we may conclude,

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = \exp\left(-\frac{1}{2}C\lambda\cdot\lambda + i\lambda\cdot\mu\right),$$

i.e. $X\overset{d}{=}N(\mu,C)$. ■

Exercise 19.9 (Existence of Gaussian random vectors for all $C\ge 0$ and $\mu\in\mathbb{R}^d$). Suppose that $\mu\in\mathbb{R}^d$ and C is a symmetric non-negative $d\times d$ matrix. By the spectral theorem we know there is an orthonormal basis $\{u_j\}_{j=1}^d$ for $\mathbb{R}^d$ such that $Cu_j=\sigma_j^2 u_j$ for some $\sigma_j^2\ge 0$. Let $\{Z_j\}_{j=1}^d$ be i.i.d. standard normal random variables, show $X:=\sum_{j=1}^d Z_j\sigma_j u_j+\mu\overset{d}{=}N(\mu,C)$.

Theorem 19.14 (Gaussian Densities). *Suppose that $X\overset{d}{=}N(\mu,C)$ is an $\mathbb{R}^d$ – valued Gaussian random vector with $C>0$ (for simplicity). Then*

$$\mathbb{E}\left[f(X)\right] = \frac{1}{\sqrt{\det(2\pi C)}}\int_{\mathbb{R}^d} f(x)\exp\left(-\frac{1}{2}C^{-1}(x-\mu)\cdot(x-\mu)\right)dx \tag{19.17}$$

for bounded or non-negative functions, $f:\mathbb{R}^d\to\mathbb{R}$.

Proof. Let us continue the notation in Exercise 19.9 and further let

$$A := [\sigma_1 u_1|\dots|\sigma_n u_n] = U\Sigma \tag{19.18}$$

where

$$U=[u_1|\dots|u_n] \text{ and } \Sigma = \operatorname{diag}(\sigma_1,\dots,\sigma_n) = [\sigma_1 e_1|\dots|\sigma_n e_n],$$

where $\{e_i\}_{i=1}^d$ is the standard orthonormal basis for $\mathbb{R}^d$. With this notation we know that $X\overset{d}{=}AZ+\mu$ where $Z=(Z_1,\dots,Z_d)^{\mathrm{tr}}$ is a standard normal Gaussian vector. Therefore,

$$\mathbb{E}\left[f(X)\right] = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(Az+\mu)\,e^{-\frac{1}{2}\|z\|^2}dz \tag{19.19}$$

wherein we have used

$$\prod_{j=1}^d \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_i^2} = (2\pi)^{-d/2}e^{-\frac{1}{2}\|z\|^2} \text{ with } \|z\|^2 := \sum_{j=1}^d z_j^2.$$

Making the change of variables $x = Az + \mu$ in Eq. (19.19) (i.e. $z = A^{-1}(x-\mu)$ and $dz = dx/\det A$) implies

$$\mathbb{E}[f(X)] = \frac{1}{(2\pi)^{d/2}\det A}\int_{\mathbb{R}^d} f(x)\, e^{-\frac{1}{2}\left\|A^{-1}(x-\mu)\right\|^2} dx. \tag{19.20}$$

Recall from your linear algebra class (or just check) that $CU = U\Sigma^2$, i.e. $C = U\Sigma^2 U^{-1} = U\Sigma^2 U^{\mathrm{tr}}$. Therefore[3],

$$AA^{\mathrm{tr}} = U\Sigma\Sigma U^{\mathrm{tr}} = U\Sigma^2 U^{-1} = C \tag{19.21}$$

which then implies $\det A = \sqrt{\det C}$ and for all $y \in \mathbb{R}^d$

$$\left\|A^{-1}y\right\|^2 = \left(A^{-1}\right)^{\mathrm{tr}} A^{-1}y \cdot y = \left(AA^{\mathrm{tr}}\right)^{-1} y \cdot y = C^{-1}y \cdot y.$$

Equation (19.17) follows from theses observations, Eq. (19.20), and the identity;

$$(2\pi)^{d/2}\det A = (2\pi)^{d/2}\sqrt{\det C} = \sqrt{(2\pi)^d \det C} = \sqrt{\det(2\pi C)}.$$

■

Theorem 19.15 (Gaussian Integration by Parts). *Suppose that $X = (X_1, \dots, X_d)$ is a mean zero Gaussian random vector and $C_{ij} = \mathrm{Cov}(X_i, Y_j) = \mathbb{E}[X_i X_j]$. Then for any smooth function $f : \mathbb{R}^d \to \mathbb{R}$ such that f and all its derivatives grows slower that $\exp(|x|^\alpha)$ for some $\alpha < 2$, we have*

$$\begin{aligned}\mathbb{E}[X_i f(X_1, \dots, X_d)] &= \sum_{k=1}^d C_{ik}\mathbb{E}\left[\frac{\partial}{\partial X_k} f(X_1, \dots, X_d)\right] \\ &= \sum_{k=1}^d \mathbb{E}[X_i X_k] \cdot \mathbb{E}\left[\frac{\partial}{\partial X_k} f(X_1, \dots, X_d)\right].\end{aligned}$$

Here we write $\frac{\partial}{\partial X_k} f(X_1, \dots, X_d)$ for $(\partial_k f)(X_1, \dots, X_d)$ where

$$(\partial_k f)(x_1, \dots, x_d) := \frac{\partial}{\partial x_k} f(x_1, \dots, x_d) = \frac{d}{dt}|_0 f(\mathbf{x} + t\mathbf{e}_k)$$

where $\mathbf{x} := (x_1, \dots, x_d)$ and $\mathbf{e}_k$ is the kth – standard basis vector for $\mathbb{R}^d$.

[3] Alternatively,

$$\begin{aligned}C_{ik} &= \mathrm{Cov}(X_i, X_k) = \mathrm{Cov}(Y_i, Y_k) \\ &= \sum_{j,m}\mathrm{Cov}(A_{ij}Z_j, A_{km}Z_m) = \sum_{j,m} A_{ij}A_{km}\,\mathrm{Cov}(Z_j, Z_m) \\ &= \sum_{j,m} A_{ij}A_{km}\delta_{jm} = \sum_j A_{ij}A_{kj} = \left(AA^{\mathrm{tr}}\right)_{ik}.\end{aligned}$$

Proof. From Exercise 19.9 we know $X \overset{d}{=} Y$ where $Y = \sum_{j=1}^d \sigma_j Z_j u_j$ where $\{u_j\}_{j=1}^d$ is an orthonormal basis for $\mathbb{R}^d$ such that $Cu_j = \sigma_j^2 u_j$ and $\{Z_j\}_{j=1}^d$ are i.i.d. standard normal random variables. To simplify notation we define $A := [\sigma_1 u_1 | \dots | \sigma_d u_d]$ as in Eq. (19.18) so that $Y = AZ$ where $Z = (Z_1, \dots, Z_d)^{\mathrm{tr}}$ as in the proof of Theorem 19.14. From our previous observations and a simple generalization of Lemma 19.3, it follows that

$$\begin{aligned}\mathbb{E}[X_i f(X_1, \dots, X_d)] &= \mathbb{E}[Y_i f(Y_1, \dots, Y_d)] \\ &= \sum_j A_{ij}\mathbb{E}[Z_j f((AZ)_1, \dots (AZ_d))] \\ &= \sum_j A_{ij}\mathbb{E}\left[\frac{\partial}{\partial Z_j} f((AZ)_1, \dots (AZ_d))\right] \\ &= \sum_j A_{ij}\mathbb{E}\left[\sum_k (\partial_k f)((AZ)_1, \dots (AZ_d)) \cdot \frac{\partial}{\partial Z_j}(AZ)_k\right] \\ &= \sum_{j,k} A_{ij}A_{kj}\mathbb{E}[(\partial_k f)(X_1, \dots, X_d)].\end{aligned}$$

This completes the proof since, $\sum_j A_{ij}A_{kj} = (AA^{\mathrm{tr}})_{ik} = C_{ik}$ as we saw in Eq. (19.21). ■

Theorem 19.16 (Wick's Theorem). *If $X = (X_1, \dots, X_{2n})$ is a mean zero Gaussian random vector, then*

$$\mathbb{E}[X_1 \dots X_{2n}] = \sum_{pairings} C_{i_1 j_1} \dots C_{i_n j_n}$$

where the sum is over all perfect pairings of $\{1, 2, \dots, 2n\}$ and

$$C_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j].$$

Proof. From Theorem 19.15,

$$\mathbb{E}[X_1 \dots X_{2n}] = \sum_j C_{1j}\mathbb{E}\left[\frac{\partial}{\partial X_j} X_2 \dots X_{2n}\right] = \sum_{j>2} C_{1j}\mathbb{E}\left[X_2 \dots \hat{X}_j \dots X_{2n}\right]$$

where the hat indicates a term to be omitted. The result now basically follows by induction. For example,

$$\begin{aligned}\mathbb{E}[X_1X_2X_3X_4] =& C_{12}\mathbb{E}\left[\frac{\partial}{\partial X_2}(X_2X_3X_4)\right] \\ &+ C_{13}\mathbb{E}\left[\frac{\partial}{\partial X_3}(X_2X_3X_4)\right] + C_{14}\mathbb{E}\left[\frac{\partial}{\partial X_4}(X_2X_3X_4)\right] \\ =& C_{12}\mathbb{E}[X_3X_4] + C_{13}\mathbb{E}[X_2X_4] + C_{14}\mathbb{E}[X_2X_3] \\ =& C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}.\end{aligned}$$

■

Recall that if X_i and Y_j are independent, then $\operatorname{Cov}(X_i, Y_j) = 0$, i.e. independence implies uncorrelated. On the other hand, typically uncorrelated random variables are not independent. However, if the random variables involved are jointly Gaussian, then independence and uncorrelated are actually the same thing!

Lemma 19.17. *Suppose that $Z = (X, Y)^{\mathrm{tr}}$ is a Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$. Then X is independent of Y iff $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$.*

Remark 19.18. Lemma 19.17 also holds more generally. Namely if $\{X^l\}_{l=1}^n$ is a sequence of random vectors such that $(X^1, \dots, X^n)$ is a Gaussian random vector. Then $\{X^l\}_{l=1}^n$ are independent iff $\operatorname{Cov}\left(X_i^l, X_k^{l'}\right) = 0$ for all $l \neq l'$ and i and k.

Exercise 19.10. Prove Lemma 19.17. **Hint:** by basic facts about the Fourier transform, it suffices to prove

$$\mathbb{E}\left[e^{ix\cdot X} e^{iy\cdot Y}\right] = \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right] \text{ for all } x \in \mathbb{R}^k \text{ and } y \in \mathbb{R}^l.$$

If you get stuck, take a look at the proof of Corollary 19.19 below.

Corollary 19.19. *Suppose that $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$ are two independent random Gaussian vectors, then (X, Y) is also a Gaussian random vector. This corollary generalizes to multiple independent random Gaussian vectors.*

Proof. Let $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned}
\mathbb{E}\left[e^{i(x,y)\cdot(X,Y)}\right] &= \mathbb{E}\left[e^{i(x\cdot X + y\cdot Y)}\right] = \mathbb{E}\left[e^{ix\cdot X} e^{iy\cdot Y}\right] = \mathbb{E}\left[e^{ix\cdot X}\right] \cdot \mathbb{E}\left[e^{iy\cdot Y}\right] \\
&= \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X) + i\mathbb{E}(x\cdot X)\right) \\
&\quad \times \exp\left(-\frac{1}{2}\operatorname{Var}(y\cdot Y) + i\mathbb{E}(y\cdot Y)\right) \\
&= \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X) + i\mathbb{E}(x\cdot X) - \frac{1}{2}\operatorname{Var}(y\cdot Y) + i\mathbb{E}(y\cdot Y)\right) \\
&= \exp\left(-\frac{1}{2}\operatorname{Var}(x\cdot X + y\cdot Y) + i\mathbb{E}(x\cdot X + y\cdot Y)\right)
\end{aligned}$$

which shows that (X, Y) is again Gaussian. ■

Remark 19.20 (Be careful). If X_1 and X_2 are two standard normal random variables, it is **not** generally true that (X_1, X_2) is a Gaussian random vector. For example suppose $X_1 \overset{d}{=} N(0,1)$ is a standard normal random variable and ε is an independent Bernoulli random variable with $\mathbb{P}(\varepsilon = \pm 1) = \frac{1}{2}$. Then $X_2 := \varepsilon X_1 \overset{d}{=} N(0,1)$ but $X := (X_1, X_2)$ is **not** a Gaussian random vector as we now verify.

If $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$, then

$$\begin{aligned}
\mathbb{E}\left[e^{i\lambda\cdot X}\right] &= \mathbb{E}\left[e^{i(\lambda_1 X_1 + \lambda_2 X_2)}\right] = \mathbb{E}\left[e^{i(\lambda_1 X_1 + \lambda_2 \varepsilon X_1)}\right] \\
&= \frac{1}{2}\sum_{\tau=\pm 1} \mathbb{E}\left[e^{i(\lambda_1 X_1 + \lambda_2 \tau X_1)}\right] = \frac{1}{2}\sum_{\tau=\pm 1} \mathbb{E}\left[e^{i(\lambda_1 + \lambda_2\tau) X_1}\right] \\
&= \frac{1}{2}\sum_{\tau=\pm 1} \exp\left(-\frac{1}{2}(\lambda_1 + \lambda_2\tau)^2\right) \\
&= \frac{1}{2}\sum_{\tau=\pm 1} \exp\left(-\frac{1}{2}\left(\lambda_1^2 + \lambda_2^2 + 2\tau\lambda_1\lambda_2\right)\right) \\
&= \frac{1}{2} e^{-\frac{1}{2}\left(\lambda_1^2 + \lambda_2^2\right)} \cdot \left[\exp(-\lambda_1\lambda_2) + \exp(\lambda_1\lambda_2)\right] \\
&= e^{-\frac{1}{2}\left(\lambda_1^2 + \lambda_2^2\right)} \cosh(\lambda_1\lambda_2).
\end{aligned}$$

On the other hand, $\mathbb{E}\left[X_1^2\right] = \mathbb{E}\left[X_2^2\right] = 1$ and

$$\mathbb{E}\left[X_1 X_2\right] = \mathbb{E}\varepsilon \cdot \mathbb{E}\left[X_1^2\right] = 0 \cdot 1 = 0,$$

from which it follows that X_1 and X_2 are uncorrelated and $C^X = I_{2\times 2}$. Thus if X were Gaussian we would have,

$$\mathbb{E}\left[e^{i\lambda\cdot X}\right] = \exp\left(-\frac{1}{2}C^X\lambda\cdot\lambda\right) = e^{-\frac{1}{2}\left(\lambda_1^2 + \lambda_2^2\right)}$$

which is just not the case!

Incidentally, this example also shows that two uncorrelated random variables need **not** be independent. For if $\{X_1, X_2\}$ were independent, then again we would have

$$\begin{aligned}
\mathbb{E}\left[e^{i\lambda\cdot X}\right] &= \mathbb{E}\left[e^{i(\lambda_1 X_1 + \lambda_2 X_2)}\right] = \mathbb{E}\left[e^{i\lambda_1 X_1} e^{i\lambda_2 X_2}\right] \\
&= \mathbb{E}\left[e^{i\lambda_1 X_1}\right] \cdot \mathbb{E}\left[e^{i\lambda_2 X_2}\right] = e^{-\frac{1}{2}\lambda_1^2} e^{-\frac{1}{2}\lambda_2^2} = e^{-\frac{1}{2}\left(\lambda_1^2 + \lambda_2^2\right)},
\end{aligned}$$

which is not the case.

The following theorem gives another useful way of computing Gaussian integrals of polynomials and exponential functions.

Theorem 19.21. *Suppose $X \overset{d}{=} N(0,C)$ where C is a $N\times N$ symmetric positive definite matrix. Let $L = L^C := \sum_{i,j=1}^d C_{ij}\partial_i\partial_j$ (sum on repeated indices) where $\partial_i := \partial/\partial x_i$. Then for any polynomial function, $q : \mathbb{R}^N \to \mathbb{R}$,*

$$\mathbb{E}[q(X)] = \left(e^{\frac{1}{2}L}q\right)(0) := \sum_{n=0}^{\infty}\frac{1}{n!}\left(\left(\frac{L}{2}\right)^n q\right)(0) \text{ (a finite sum).} \tag{19.22}$$

Proof. This is a fairly straight forward extension of Exercise 19.2 and so I will only provide a short outline to the proof. 1) Let $u(t) := \mathbb{E}\left[q\left(\sqrt{t}X\right)\right]$. 2) Using Theorem 19.15 one shows that $\dot{u}(t) = \frac{1}{2}\mathbb{E}\left[(Lq)\left(\sqrt{t}X\right)\right]$. 3) Iterating this result and then using Taylor's theorem finishes the proof just like in Exercise 19.2.

■

Corollary 19.22. *The function $u(t,x) := \mathbb{E}\left[q\left(x+\sqrt{t}X\right)\right]$ solves the heat equation,*

$$\partial_t u(t,x) = \frac{1}{2}L^C u(t,x) \text{ with } u(0,x) = q(x).$$

If $X \overset{d}{=} N(1,0)$ we have

$$\begin{aligned} u(t,x) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} q\left(x+\sqrt{t}z\right)e^{-\frac{1}{2}z^2}dz \\ &= \int_{\mathbb{E}} p_t(x,y)\,q(y)\,dy \end{aligned}$$

where

$$p_t(x,y) := \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{1}{2t}(y-x)^2\right).$$

19.3 Gaussian Conditioning

Notation 19.23 *Let (X,Y) be a mean zero Gaussian random vector taking values in $\mathbb{R}^k\times\mathbb{R}^l$,*

$$C_X := \mathbb{E}\left[XX^{\mathrm{tr}}\right],\ C_Y := \mathbb{E}\left[YY^{\mathrm{tr}}\right],\ C_{X,Y} = \mathbb{E}\left[XY^{\mathrm{tr}}\right], \text{ and } C_{Y,X} = \mathbb{E}\left[YX^{\mathrm{tr}}\right]$$

so that C_X, C_Y, $C_{X,Y}$, and $C_{Y,X}$ are $k\times k$, $l\times l$, $k\times l$, and $l\times k$ matrices respectively. [Note that $C_{X,Y}^{\mathrm{tr}} = C_{Y,X}$ while $C_X^{\mathrm{tr}} = C_X$ and $C_Y^{\mathrm{tr}} = C_Y$.]

Definition 19.24 (Pseudo-Inverses). *If C is a symmetric $k\times k$ matrix on $\mathbb{R}^k$, let C^{-1} be the $k\times k$ matrix uniquely determined by; $C^{-1}v = 0$ if $v\in\mathrm{Nul}(C)$ while if $v\in\mathrm{Nul}(C)^\perp = \mathrm{Ran}(C)$ we let $C^{-1}v = w$ where w is the unique element of $\mathrm{Ran}(C)$ such that $Cw = v$. [If C is invertible, then the pseudo inverse is the same as the inverse matrix.]*

Lemma 19.25. *If X is a mean zero $\mathbb{R}^k$ – valued Gaussian random vector, then $X\in\mathrm{Ran}(C_X)$ a.s.*

Proof. If $\lambda\in\mathrm{Ran}(C_X)^\perp = \mathrm{Nul}(C_X^*) = \mathrm{Nul}(C_X)$, then

$$\mathbb{E}[\lambda\cdot X]^2 = C_X\lambda\cdot\lambda = 0 \implies \lambda\cdot X = 0 \text{ a.s. }.$$

Letting λ run through a basis for $\mathrm{Ran}(C_X)^\perp$, it then follows that $X\in\mathrm{Ran}(C_X)$ a.s. ■

Lemma 19.26. *If (X,Y) is a mean zero Gaussian random vector taking values in $\mathbb{R}^k\times\mathbb{R}^l$, then $Z := Y - C_{Y,X}C_X^{-1}X$ is independent of X, i.e.*

$$Y = C_{Y,X}C_X^{-1}X + Z \tag{19.23}$$

where Z is a mean zero Gaussian random vector independent of X such that

$$C_Z = C_Y - C_{Y,X}C_X^{-1}C_{X,Y}. \tag{19.24}$$

Proof. We look for a $l\times k$ matrix, A, so that $Z := Y - AX$ is independent of X. Since (X,Y) is Gaussian it suffices to find A so that

$$0 = \mathbb{E}\left[ZX^{\mathrm{tr}}\right] = \mathbb{E}\left[(Y-AX)X^{\mathrm{tr}}\right] = C_{Y,X} - AC_X,$$

i.e. we require $AC_X = C_{Y,X}$. This suggests that we let $A = C_{Y,X}C_X^{-1}$ but we must check this works even when C_X is not invertible. In this case

$$AC_X = C_{Y,X}C_X^{-1}C_X = C_{Y,X}P_X$$

where P_X is orthogonal projection onto $\mathrm{Ran}(C_X)$. The claim is that $C_{Y,X}P_X = C_{Y,X}$. The point is that if $\lambda\in\mathrm{Nul}(C_X)$, then

$$C_{Y,X}\lambda = \mathbb{E}\left[YX^{\mathrm{tr}}\lambda\right] = \mathbb{E}\left[(\lambda\cdot X)Y\right] = \mathbb{E}[0] = 0,$$

wherein we have used Lemma 19.25 to conclude $\lambda\cdot X = 0$ a.s. Consequently, $C_{Y,X}$ vanishes on $\mathrm{Ran}(C_X)^\perp$ and hence $C_{Y,X}P_X = C_{Y,X}$.

So we have shown $Z := Y - C_{Y,X}C_X^{-1}X$ is independent of X. We now compute the covariance (C_Z) of Z;

$$\begin{aligned} C_Z &= \mathbb{E}\left[ZZ^{\mathrm{tr}}\right] = \mathbb{E}\left[Z\left[Y - C_{Y,X}C_X^{-1}X\right]^{\mathrm{tr}}\right] \\ &= \mathbb{E}\left[ZY^{\mathrm{tr}}\right] = \mathbb{E}\left[\left[Y - C_{Y,X}C_X^{-1}X\right]Y^{\mathrm{tr}}\right] \\ &= C_Y - C_{Y,X}C_X^{-1}C_{X,Y}, \end{aligned}$$

wherein the third equality we made use of the fact that Z and X were independent of (X,Z) is still jointly Gaussian. ■

Theorem 19.27 (Gaussian Conditioning). *Suppose that (X, Y) is a Gaussian vector taking values in $\mathbb{R}^k \times \mathbb{R}^l$. Then*

$$\mathbb{E}\left[f\left(Y\right)|X\right] = G\left(X\right) \tag{19.25}$$

where

$$G\left(x\right) := \mathbb{E}\left[f\left(C_{Y,X}C_X^{-1}x + Z\right)\right] \tag{19.26}$$

and Z is a $\mathbb{R}^l$ – mean zero Gaussian random vector with covariance matrix (C_Z) as in Eq. (19.24). As a special case, taking $f(y) = y$ above shows

$$\mathbb{E}\left[Y|X\right] = C_{Y,X}C_X^{-1}X.$$

Proof. This follows directly from Proposition 3.18 and the decomposition in Lemma 19.26. ■

Corollary 19.28. *If (X, Y) is a Gaussian random vector taking values in $\mathbb{R}^k \times \mathbb{R}^l$ with $\mu_X = \mathbb{E}X$ and $\mu_Y = \mathbb{E}Y$, then*

$$Y = \mu_Y + C_{Y,X}C_X^{-1}\left(X - \mu_X\right) + Z$$

where now,

$$C_{Y,X} = \mathbb{E}\left[\left(Y - \mu_Y\right)\left(X - \mu_X\right)^{\mathrm{tr}}\right] = \mathbb{E}YX^{\mathrm{tr}} - \left(\mathbb{E}Y\right)\left(\mathbb{E}X\right)^{\mathrm{tr}}$$

with similar definitions for $C_{X,Y}$, C_X, and C_Y and Z is a mean zero Gaussian $\mathbb{R}^l$ – random vector independent of X with covariance matrix,

$$C_Z = C_Y - C_{Y,X}C_X^{-1}C_{X,Y}. \tag{19.27}$$

Moreover

$$\mathbb{E}\left[f\left(Y\right)|X\right] = G\left(X\right) \text{ where } G\left(x\right) := \mathbb{E}\left[f\left(C_{Y,X}C_X^{-1}x + \tilde{Z}\right)\right], \tag{19.28}$$

where $\tilde{Z}$ is a $\mathbb{R}^l$ - valued Gaussian random vector independent such that

$$\tilde{Z} = Z + \mu_Y - C_{Y,X}C_X^{-1}\mu_X \overset{d}{=} N\left(\mu_Y - C_{Y,X}C_X^{-1}\mu_X, C_Z\right). \tag{19.29}$$

Proof. Applying Lemma 19.26 to $(X - \mu_X, Y - \mu_Y)$ shows,

$$Y - \mu_Y = C_{Y,X}C_X^{-1}\left(X - \mu_X\right) + Z \tag{19.30}$$

where the matrices $C_{Y,X}$ and C_X are now the covariances as described in the statement. The $\mathbb{R}^l$ – valued random vector, Z, is still a mean zero Gaussian vector with covariance given by Eq. (19.27) as described. We may rewrite Eq. (19.30) in the form $Y = C_{Y,X}C_X^{-1}X + \tilde{Z}$ where $\tilde{Z}$ is given as in Eq. (19.29). The formula for $\mathbb{E}\left[f\left(Y\right)|X\right]$ given in Eq. (19.28) follows from this decomposition along with Proposition 3.18. ■

Solution to Selected Lawler Problems

References

1. Richard F. Bass, *Probabilistic techniques in analysis*, Probability and its Applications (New York), Springer-Verlag, New York, 1995. MR MR1329542 (96e:60001)
2. _____, *Diffusions and elliptic operators*, Probability and its Applications (New York), Springer-Verlag, New York, 1998. MR MR1483890 (99h:60136)
3. K. L. Chung and R. J. Williams, *Introduction to stochastic integration*, second ed., Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1990. MR MR1102676 (92d:60057)
4. Persi Diaconis and J. W. Neuberger, *Numerical results for the Metropolis algorithm*, Experiment. Math. **13** (2004), no. 2, 207–213. MR 2068894
5. J. L. Doob, *Stochastic processes*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR MR1038526 (91d:60002)
6. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
7. _____, *Stochastic calculus*, Probability and Stochastics Series, CRC Press, Boca Raton, FL, 1996, A practical introduction. MR MR1398879 (97k:60148)
8. Evgenii B. Dynkin and Aleksandr A. Yushkevich, *Markov processes: Theorems and problems*, Translated from the Russian by James S. Wood, Plenum Press, New York, 1969. MR MR0242252 (39 #3585a)
9. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
10. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
11. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR MR1121940 (92h:60127)
12. K. L. Mengersen and R. L. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, Ann. Statist. **24** (1996), no. 1, 101–121. MR 1389882 (98c:60081)
13. Sean Meyn and Richard L. Tweedie, *Markov chains and stochastic stability*, second ed., Cambridge University Press, Cambridge, 2009, With a prologue by Peter W. Glynn. MR 2509253 (2010h:60206)
14. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
15. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
16. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
17. Sheldon M. Ross, *Stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1983, Lectures in Mathematics, 14. MR MR683455 (84m:60001)
18. Barry Simon, *Functional integration and quantum physics*, second ed., AMS Chelsea Publishing, Providence, RI, 2005. MR MR2105995 (2005f:81003)