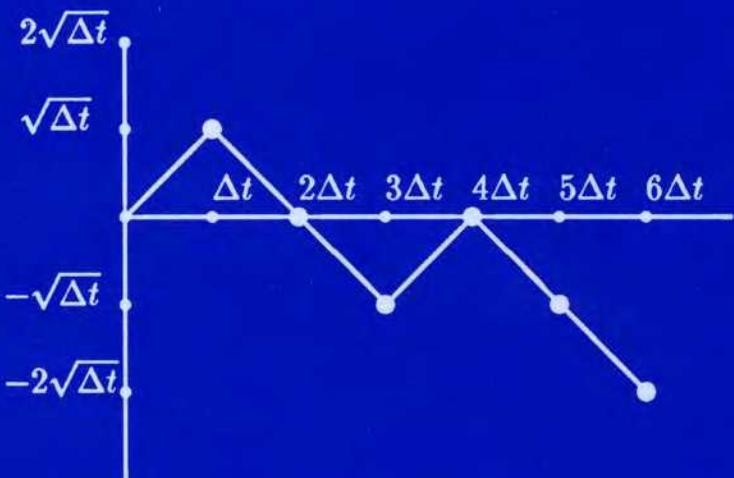


Introduction to Stochastic Processes

Second Edition



GREGORY F. LAWLER



Chapman & Hall/CRC
Taylor & Francis Group

Introduction to Stochastic Processes

Second Edition

Introduction to Stochastic Processes

Second Edition

GREGORY F. LAWLER



Chapman & Hall/CRC

Taylor & Francis Group

Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Published in 2006 by
Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2006 by Taylor & Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3

International Standard Book Number-10: 1-58488-651-X
International Standard Book Number-13: 978-1-58488-561-8

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Contents

Preface to Second Edition	ix
Preface to First Edition	xi
0 Preliminaries	1
0.1 Introduction	1
0.2 Linear Differential Equations	1
0.3 Linear Difference Equations	3
0.4 Exercises	6
1 Finite Markov Chains	9
1.1 Definitions and Examples	9
1.2 Large-Time Behavior and Invariant Probability	14
1.3 Classification of States	17
1.3.1 Reducibility	19
1.3.2 Periodicity	21
1.3.3 Irreducible, aperiodic chains	22
1.3.4 Reducible or periodic chains	22
1.4 Return Times	24
1.5 Transient States	26
1.6 Examples	31
1.7 Exercises	35
2 Countable Markov Chains	43
2.1 Introduction	43
2.2 Recurrence and Transience	45
2.3 Positive Recurrence and Null Recurrence	50
2.4 Branching Process	53
2.5 Exercises	57
3 Continuous-Time Markov Chains	65
3.1 Poisson Process	65
3.2 Finite State Space	68
3.3 Birth-and-Death Processes	74
3.4 General Case	81
3.5 Exercises	82

4 Optimal Stopping	87
4.1 Optimal Stopping of Markov Chains	87
4.2 Optimal Stopping with Cost	93
4.3 Optimal Stopping with Discounting	96
4.4 Exercises	98
5 Martingales	101
5.1 Conditional Expectation	101
5.2 Definition and Examples	106
5.3 Optional Sampling Theorem	110
5.4 Uniform Integrability	114
5.5 Martingale Convergence Theorem	116
5.6 Maximal Inequalities	122
5.7 Exercises	125
6 Renewal Processes	131
6.1 Introduction	131
6.2 Renewal Equation	136
6.3 Discrete Renewal Processes	144
6.4 M/G/1 and G/M/1 Queues	148
6.5 Exercises	151
7 Reversible Markov Chains	155
7.1 Reversible Processes	155
7.2 Convergence to Equilibrium	157
7.3 Markov Chain Algorithms	162
7.4 A Criterion for Recurrence	166
7.5 Exercises	170
8 Brownian Motion	173
8.1 Introduction	173
8.2 Markov Property	176
8.3 Zero Set of Brownian Motion	181
8.4 Brownian Motion in Several Dimensions	184
8.5 Recurrence and Transience	189
8.6 Fractal Nature of Brownian Motion	191
8.7 Scaling Rules	192
8.8 Brownian Motion with Drift	193
8.9 Exercises	195
9 Stochastic Integration	199
9.1 Integration with Respect to Random Walk	199
9.2 Integration with Respect to Brownian Motion	200
9.3 Itô's Formula	205
9.4 Extensions of Itô's Formula	209

9.5	Continuous Martingales	216
9.6	Girsanov Transformation	218
9.7	Feynman-Kac Formula	221
9.8	Black-Scholes Formula	223
9.9	Simulation	228
9.10	Exercises	228
Suggestions for Further Reading		231
Index		233

Preface to Second Edition

In the second edition we have significantly expanded the chapter on stochastic integration in order to give an introduction to modern mathematical finance. We have expanded the discussion of Itô's formula, introduced the Girsanov transformation and the Feynman-Kac formula, and derived the Black-Scholes formula for pricing options. We have tried to present this material in the same styles as other topics, that is, without complete mathematical details, but with enough ideas to explain to the reader why formulas are true.

We have added a section on maximal inequalities to the martingale section and included more material on Brownian motion. We have included a few more examples throughout the book and have increased the number of exercises at the end of the chapters. We have also made corrections and minor revisions in many places and included some recommendations for further reading.

Preface to First Edition

This book is an outgrowth of lectures in Mathematics 240, “Applied Stochastic Processes,” which I have taught a number of times at Duke University. The majority of the students in the course are graduate students from departments other than mathematics, including computer science, economics, business, biological sciences, psychology, physics, statistics, and engineering. There have also been graduate students from the mathematics department as well as some advanced undergraduates. The mathematical background of the students varies greatly, and the particular areas of stochastic processes that are relevant for their research also vary greatly.

The prerequisites for using this book are a good calculus-based undergraduate course in probability and a course in linear algebra including eigenvalues and eigenvectors. I also assume that the reader is reasonably computer literate. The exercises assume that the reader can write simple programs and has access to some software for linear algebra computations. In all of my classes, students have had sufficient computer experience to accomplish this. Most of the students have also had some exposure to differential equations and I use such ideas freely, although I have a short section on linear differential equations in the preliminary chapter.

I have tried to discuss key mathematical ideas in this book, but I have not made an attempt to put in all the mathematical details. Measure theory is not a prerequisite but I have tried to present topics in a way such that readers who have some knowledge of measure theory can fill in details. Although this is a book intended primarily for people with applications in mind, there are few real applications discussed. True applications require a good understanding of the field being studied and it is not a goal of this book to discuss the many different fields in which stochastic processes are used. I have instead chosen to stick with the very basic examples and let the experts in other fields decide when certain mathematical assumptions are appropriate for their application.

Chapter 1 covers the standard material on finite Markov chains. I have not given proofs of the convergence to equilibrium but rather have emphasized the relationship between the convergence to equilibrium and the size of the eigenvalues of the stochastic matrix. Chapter 2 deals with infinite state space. The notions of transience, null recurrence, and positive recurrence are introduced, using as the main example, a random walk on the nonnegative integers with reflecting boundary. The chapter ends with a discussion of branching processes.

Continuous-time Markov chains are discussed in Chapter 3. The discussion

centers on three main types: Poisson process, finite state space, and birth-and-death processes. For these processes I have used the forward differential equations to describe the evolution of the probabilities. This is easier and more natural than the backward equations. Unfortunately, the forward equations are not a legitimate means to analyze all continuous-time Markov chains and this fact is discussed briefly in the last section. One of the main examples of a birth-and-death process is a Markovian queue.

I have included Chapter 4 on optimal stopping of Markov chains as one example in the large area of decision theory. Optimal stopping has a nice combination of theoretical mathematics leading to an algorithm to solve a problem. The basic ideas are also similar to ideas presented in Chapter 5.

The idea of a martingale is fundamental in much of stochastic processes, and the goal of Chapter 5 is to give a solid introduction to these ideas. The modern definition of conditional expectation is first discussed and the idea of “measurable with respect to \mathcal{F}_n , the information available at time n ” is used freely without worrying about giving it a rigorous meaning in terms of σ -algebras. The major theorems of the area, optional sampling and the martingale convergence theorem, are discussed as well as their proofs. Proofs are important here since part of the theory is to understand why the theorems do not always hold. I have included a discussion of uniform integrability.

The basic ideas of renewal theory are discussed in Chapter 6. For nonlattice random variables the renewal equation is used as the main tool of analysis while for lattice random variables a Markov chain approach is used. As an application, queues with general service times are analyzed.

Chapter 7 discusses a couple of current topics in the realm of reversible Markov chains. First a more mathematical discussion about the rate of convergence to equilibrium is given, followed by a short introduction to the idea of Markov chain algorithms which are becoming very important in some areas of physics, computer science, and statistics. The final section on recurrence is a nice use of “variational” ideas to prove a result that is hard to prove directly.

Chapter 8 gives a very quick introduction to a large number of ideas in Brownian motion. It is impossible to make any attempt to put in all the mathematical details. I have discussed multidimensional as well as one-dimensional Brownian motion and have tried to show why Brownian motion and the heat equation are basically the same subject. I have also tried to discuss a little of the fractal nature of some of the sets produced by Brownian motion. In Chapter 9, a very short introduction to the idea of stochastic integration is given. This also is a very informal discussion but is intended to allow the students to at least have some ideas of what a stochastic integral is.

This book has a little more than can be covered in a one semester course. In my view the basic course consists of Chapters 1, 2, 3, 5, and 8. Which of the remaining chapters I cover depends on the particular students in the class that semester. The basic chapters should probably be done in the order listed, but the other chapters can be done at any time. Chapters 4 and 7 use the previous material on Markov chains; Chapter 6 uses Markov chains and

martingales in the last section; and Chapter 9 uses the definition of Brownian motion as well as martingales.

I would like to thank the students in Math 240 in Spring 1992 and Spring 1994 for their comments and corrections on early versions of these notes. I also thank Rick Clelland, who was my assistant when I was preparing the first version in 1992, and the reviewers, Michael Phelan and Daniel C. Wiener, for their suggestions. During the writing of this book, I was partially supported by the National Science Foundation.

Chapter 0

Preliminaries

0.1 Introduction

A *stochastic process* is a random process evolving with time. More precisely, a stochastic process is a collection of random variables X_t indexed by time. In this book, time will always be either a subset of the nonnegative integers $\{0, 1, 2, \dots\}$ or a subset of $[0, \infty)$, the nonnegative real numbers. In the first case we will call the process *discrete time*, and in the second case *continuous time*. The random variables X_t will take values in a set that we call the *state space*. We will consider cases both where the state space is *discrete*, i.e., a finite or countably infinite set, and cases where the state space is *continuous*, e.g., the real numbers \mathbb{R} or d -dimensional space \mathbb{R}^d .

The study of deterministic (nonrandom) processes changing with time leads one to the study of differential equations (if time is continuous) or difference equations (if time is discrete). A typical (first-order) differential equation is of the form

$$y'(t) = F(t, y(t)).$$

Here the change in the function $y(t)$ depends only on t and the value $y(t)$ and not on the values at times before t . A large class of stochastic processes also have the property that the change at time t is determined by the value of the process at time t and not by the values at times before t . Such processes are called *Markov processes*. The study of such processes is closely related to linear algebra, differential equations, and difference equations. We assume that the reader is familiar with linear algebra. In the next section we review some facts about linear differential equations that will be used and in the following section we discuss difference equations.

0.2 Linear Differential Equations

Here we briefly review some facts about homogeneous linear differential equations with constant coefficients. Readers who want more detail should

consult any introductory text in differential equations. Consider the homogeneous differential equation

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1y'(t) + a_0y(t) = 0, \quad (0.1)$$

where a_0, \dots, a_{n-1} are constants. For any initial conditions

$$y(0) = b_0, \quad y'(0) = b_1, \quad \dots, \quad y^{(n-1)}(0) = b_{n-1},$$

there is a unique solution to (0.1) satisfying these conditions. To obtain such a particular solution, we first find the general solution. Suppose $y_1(t), \dots, y_n(t)$ are linearly independent solutions to (0.1). Then every solution can be written in the form

$$y(t) = c_1y_1(t) + \cdots + c_ny_n(t),$$

for constants c_1, \dots, c_n . For a given set of initial conditions we can determine the appropriate constants.

The solutions y_1, \dots, y_n are found by looking for solutions of the form $y(t) = e^{\lambda t}$. Plugging in, we see that such a function $y(t)$ satisfies the equation if and only if

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0.$$

If this polynomial has n distinct roots $\lambda_1, \dots, \lambda_n$ we get n linearly independent solutions $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$. The case of repeated roots is a little trickier, but with a little calculation one can show that if λ is a root of multiplicity j , then $e^{\lambda t}, te^{\lambda t}, \dots, t^{j-1}e^{\lambda t}$ are all solutions. Hence for each root of multiplicity j , we get j linearly independent solutions, and combining them all we get n linearly independent solutions as required.

Now consider the first-order linear system of equation

$$\begin{aligned} y'_1(t) &= a_{11}y_1(t) + a_{12}y_2(t) + \cdots + a_{1n}y_n(t) \\ y'_2(t) &= a_{21}y_1(t) + a_{22}y_2(t) + \cdots + a_{2n}y_n(t) \\ &\vdots && \vdots \\ y'_n(t) &= a_{n1}y_1(t) + a_{n2}y_2(t) + \cdots + a_{nn}y_n(t). \end{aligned}$$

This can be written as a single vector valued equation:

$$\bar{y}'(t) = \mathbf{A}\bar{y}(t).$$

Here $\bar{y}(t) = [y_1(t), \dots, y_n(t)]$ (more precisely, the transpose of this vector) and \mathbf{A} is the matrix of coefficients (a_{ij}) . For any initial vector $\bar{v} = (v_1, \dots, v_n)$, there is a unique solution to this equation satisfying $\bar{y}(0) = \bar{v}$. This solution can most easily be written in terms of the exponential of the matrix,

$$\bar{y}(t) = e^{t\mathbf{A}}\bar{v}.$$

This exponential can be defined in terms of a power series:

$$e^{t\mathbf{A}} = \sum_{j=0}^{\infty} \frac{(t\mathbf{A})^j}{j!}.$$

For computational purposes one generally tries to diagonalize the matrix \mathbf{A} . Suppose that $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{D}\mathbf{Q}$ for some diagonal matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Then

$$e^{t\mathbf{A}} = \mathbf{Q}^{-1}e^{t\mathbf{D}}\mathbf{Q} = \mathbf{Q}^{-1} \begin{bmatrix} e^{td_1} & 0 & \cdots & 0 \\ 0 & e^{td_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{td_n} \end{bmatrix} \mathbf{Q}.$$

It is not true that every matrix can be diagonalized as above. However, every matrix \mathbf{A} can be written as $\mathbf{Q}^{-1}\mathbf{J}\mathbf{Q}$ where \mathbf{J} is in Jordan canonical form. Taking exponentials of matrices in Jordan form is only slightly more difficult than taking exponentials of diagonal matrices. See a text on linear algebra for more details.

0.3 Linear Difference Equations

The theory of linear difference equations is very similar to that of linear differential equations. However, since the theory is generally not studied in introductory differential equations courses and since difference equations arise naturally in discrete-time Markov chains, we will discuss their solution in more detail. First consider the equation

$$f(n) = af(n-1) + bf(n+1), \quad K < n < N. \quad (0.2)$$

Here $f(n)$ is a function defined for integers $K \leq n \leq N$ (N can be chosen to be infinity) and a, b are nonzero real numbers. If f satisfies (0.2) and the values $f(K)$ and $f(K+1)$ are known, then $f(n)$ can be determined for all $K \leq n \leq N$ recursively by the formula

$$f(n+1) = \frac{1}{b}[f(n) - af(n-1)]. \quad (0.3)$$

Conversely, if u_0, u_1 are any real numbers we can find a solution to (0.2) satisfying $f(K) = u_0, f(K + 1) = u_1$ by defining $f(n)$ recursively as in (0.3). Also, we note that the set of functions satisfying (0.2) is a vector space, i.e., if f_1, f_2 satisfy (0.2) then so does $c_1 f_1 + c_2 f_2$, where c_1, c_2 are any real numbers. This vector space has dimension 2; in fact, a basis for the vector space is given by $\{f_1, f_2\}$, where f_1 is the solution satisfying $f_1(K) = 1, f_1(K + 1) = 0$ and f_2 is the solution satisfying $f_2(K) = 0, f_2(K + 1) = 1$. If g_1 and g_2 are any two linearly independent solutions, then it is a standard fact from linear algebra that every solution is of the form

$$c_1 g_1 + c_2 g_2$$

for constants c_1, c_2 .

We now make some good guesses to find a pair of linearly independent solutions. We will try functions of the form $f(n) = \alpha^n$ for some $\alpha \neq 0$. This is a solution for a particular α if and only if

$$\alpha^n = a\alpha^{n-1} + b\alpha^{n+1}, \quad K < n < N,$$

i.e., if and only if

$$\alpha = a + b\alpha^2.$$

We can solve this with the quadratic formula, giving

$$\alpha = \frac{1 \pm \sqrt{1 - 4ab}}{2b}.$$

Case I: $1 - 4ab \neq 0$. In this case there are two distinct roots, α_1, α_2 , and hence the general solution is

$$f(n) = c_1 \alpha_1^n + c_2 \alpha_2^n. \quad (0.4)$$

Case II: $1 - 4ab = 0$. In this case we get only one solution of this type, $g_1(n) = \alpha^n = (1/2b)^n$. However, if we let $g_2(n) = n(1/2b)^n$ we see that

$$\begin{aligned} ag_2(n-1) + bg_2(n+1) &= a(n-1)(1/2b)^{n-1} + b(n+1)(1/2b)^{n+1} \\ &= (1/2b)^n [a(n-1)2b + b(n+1)/(2b)] \\ &= (1/2b)^n n = g_2(n). \end{aligned}$$

Therefore g_2 is also a solution. It is easy to check that g_1, g_2 are linearly independent, so every solution is of the form

$$f(n) = c_1 (1/2b)^n + c_2 n (1/2b)^n.$$

Example. Suppose we want to find a function f satisfying

$$f(n) = \frac{1}{6} f(n-1) + \frac{2}{3} f(n+1), \quad 0 < n < \infty,$$

with $f(0) = 4, f(1) = 3$. Plugging in we get,

$$\alpha = \frac{3 \pm \sqrt{5}}{4}.$$

The general solution is

$$f(n) = c_1 \left(\frac{3 + \sqrt{5}}{4} \right)^n + c_2 \left(\frac{3 - \sqrt{5}}{4} \right)^n.$$

If we plug in the initial conditions, we get

$$4 = f(0) = c_1 + c_2,$$

$$3 = f(1) = c_1 \frac{3 + \sqrt{5}}{4} + c_2 \frac{3 - \sqrt{5}}{4}.$$

Solving gives $c_1 = 2, c_2 = 2$, and hence

$$f(n) = 2 \left(\frac{3 + \sqrt{5}}{4} \right)^n + 2 \left(\frac{3 - \sqrt{5}}{4} \right)^n.$$

We have seen that the values of $f(K)$ and $f(K + 1)$ uniquely determine the solution to (0.2). Sometimes, one is given the boundary values $f(K)$ and $f(N)$. These boundary value problems can be solved in the same way—write down the general solution and solve for the constants. For example, suppose we want the function f which satisfies

$$f(n) = 2f(n - 1) - f(n + 1), \quad 0 < n < 10,$$

with $f(0) = 0, f(10) = 1$. We write down the general solution

$$f(n) = c_1 1^n + c_2 (-2)^n.$$

Plugging in the initial conditions gives

$$f(0) = 0 = c_1 + c_2$$

$$f(10) = 1 = c_1 + c_2 (-2)^{10},$$

and $c_1 = -c_2 = 1/(1 - 2^{10})$.

In the study of random walks, the difference equations

$$f(n) = (1 - p)f(n - 1) + pf(n + 1), \quad p \in (0, 1)$$

arise. If $p \neq 1/2$, we obtain two roots $\alpha_1 = 1, \alpha_2 = (1-p)/p$, and hence the general solution is

$$f(n) = c_1 + c_2 \left(\frac{1-p}{p} \right)^n. \quad (0.5)$$

If $p = 1/2$, $\alpha = 1$ is a repeated root so we get the general solution

$$f(n) = c_1 + c_2 n. \quad (0.6)$$

What we have analyzed are second-order linear difference equations. The general k th-order homogeneous linear difference equation is of the form

$$f(n+k) = a_0 f(n) + a_1 f(n+1) + \cdots + a_{k-1} f(n+k-1). \quad (0.7)$$

Suppose we wish to find a function satisfying (0.7) for $n \geq 0$. It suffices to give the values $f(0), \dots, f(k-1)$, for then $f(n), n \geq k$ can be determined recursively. Again we look for solutions of the form $f(n) = \alpha^n$. Such an f is a solution if and only if

$$\alpha^k = a_0 + a_1 \alpha + \cdots + a_{k-1} \alpha^{k-1}.$$

As before, if there are k distinct roots of this equation, we get k linearly independent solutions. If a certain α is a root with multiplicity j , one can check in fact that

$$\alpha^n, n\alpha^n, n^2\alpha^n, \dots, n^{j-1}\alpha^n$$

are all linearly independent solutions. In complete parallel with the case of linear differential equations, we get k linearly independent solutions to (0.7) and we can find all solutions by taking linearly combinations of these solutions.

0.4 Exercises

0.1 Find all functions $x(t), y(t)$ satisfying

$$x'(t) = y(t) - x(t),$$

$$y'(t) = 3x(t) - 3y(t).$$

Find the particular pair of functions satisfying $x(0) = y(0) = 1/2$.

0.2 Find the function $f(n), n = 0, 1, \dots, 10$ that satisfies

$$f(n) = \frac{1}{4}f(n-1) + \frac{3}{4}f(n+1), \quad n = 1, 2, \dots, 9,$$

and $f(0) = 0, f(1) = 1$.

0.3 The Fibonacci numbers F_n are defined by $F_1 = 1$, $F_2 = 1$ and for $n > 2$, $F_n = F_{n-1} + F_{n-2}$. Find a formula for F_n by solving the difference equation.

0.4 Find the function $f(n)$, $n = 0, 1, 2, \dots$ that satisfies

$$f(0) = 0,$$

$$f(n) = \frac{1}{3}f(n-1) + \frac{1}{3}f(n+1) + \frac{1}{3}f(n+2), \quad n \geq 1,$$

$$\lim_{n \rightarrow \infty} f(n) = 1.$$

0.5 Find all functions f from the integers to the real numbers satisfying

$$f(n) = \frac{1}{2}f(n+1) + \frac{1}{2}f(n-1) - 1. \quad (0.8)$$

[Hint: First show that $f(n) = n^2$ satisfies (0.8). Then suppose $f_1(n)$ and $f_2(n)$ both satisfy (0.8) and find the equation that $g(n) = f_2(n) - f_1(n)$ satisfies.]

0.6 (a) Find all functions f from the real numbers to the real numbers such that for all x ,

$$f''(x) + f'(x) + f(x) = 0.$$

(b) Find all functions f from the integers to the real numbers such that for all n ,

$$f(n+2) = -f(n) - f(n+1).$$

Chapter 1

Finite Markov Chains

1.1 Definitions and Examples

Consider a discrete-time stochastic process, $X_n, n = 0, 1, 2, \dots$, where X_n takes values in the finite set $S = \{1, \dots, N\}$ or $\{0, \dots, N-1\}$. We call the possible values for X_n the *states* of the system. To describe the probabilities for such a process we need to give the values of

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\},$$

for every n and every finite sequence of states (i_0, \dots, i_n) . Equivalently, we could give the initial probability distribution

$$\phi(i) = \mathbb{P}\{X_0 = i\}, \quad i = 1, \dots, N$$

and the “transition probabilities,”

$$q_n(i_n | i_0, \dots, i_{n-1}) = \mathbb{P}\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}, \quad (1.1)$$

for then

$$\mathbb{P}\{X_0 = i_0, \dots, X_n = i_n\} =$$

$$\phi(i_0)q_1(i_1 | i_0)q_2(i_2 | i_0, i_1) \cdots q_n(i_n | i_0, \dots, i_{n-1}). \quad (1.2)$$

In this chapter we consider a special class of such processes, those that satisfy the *Markov property*. The Markov property states that to make predictions of the behavior of a system in the future, it suffices to consider only the present state of the system and not the past history. That is to say, the state of the system is important but not how it arrived at that state. Mathematically, we can write this as

$$\mathbb{P}\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} = \mathbb{P}\{X_n = i_n | X_{n-1} = i_{n-1}\}.$$

We will also make the assumption that the transition probabilities do not depend on time. This is called time homogeneity. A *time-homogeneous Markov chain* is a process such that

$$\mathbb{P}\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} = p(i_{n-1}, i_n),$$

for some function $p : S \times S \rightarrow [0, 1]$. Unless explicitly stated otherwise in this book, when we say Markov chain we will mean time-homogeneous Markov chain. To give the probabilities for a Markov chain, we need to give an initial probability distribution $\phi(i) = \mathbb{P}\{X_0 = i\}$, and the transition probabilities $p(i, j)$, for then, by (1.2),

$$\mathbb{P}\{X_0 = i_0, \dots, X_n = i_n\} = \phi(i_0) p(i_0, i_1) p(i_1, i_2) \cdots p(i_{n-1}, i_n). \quad (1.3)$$

The *transition matrix* \mathbf{P} for the Markov chain is the $N \times N$ matrix whose (i, j) entry \mathbf{P}_{ij} is $p(i, j)$. The matrix \mathbf{P} is a *stochastic matrix*, i.e.,

$$0 \leq \mathbf{P}_{ij} \leq 1, \quad 1 \leq i, j \leq N, \quad (1.4)$$

$$\sum_{j=1}^N \mathbf{P}_{ij} = 1, \quad 1 \leq i \leq N. \quad (1.5)$$

Any matrix satisfying (1.4) and (1.5) can be the transition matrix for a Markov chain.

Example 1. Two-state Markov chain. Let us give a simple model for the state of a phone where $X_n = 0$ means that the phone is free at time n and $X_n = 1$ means that the phone is busy. We assume that during each time interval there is a probability p that a call comes in (for ease we will assume that no more than one call comes in during any particular time interval). If the phone is busy during that period, the incoming call does not get through. We also assume that if the phone is busy during a time interval, there is a probability q that it will be free during the next interval. Our model gives a Markov chain with state space $S = \{0, 1\}$ and matrix

$$\mathbf{P} = \begin{smallmatrix} & 0 & 1 \\ 0 & 1-p & p \\ 1 & q & 1-q \end{smallmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

This matrix give the general form for a transition matrix of a two-state Markov chain. In order to specify the matrix one only needs to give the values of p and q . We have written the matrix in two different ways. The first way labels the states and the latter way does not. We will use both notations in this chapter.

Example 2. Simple Queueing Model. We modify the previous example by assuming that the phone system can put one caller on hold. Hence at any time the number of callers in the system is in the set $S = \{0, 1, 2\}$. Again, any call will be completed during a time interval with probability q and a new caller will come in with probability p , unless the system is already full. To model this we set

$$p(0, 0) = 1 - p, \quad p(0, 1) = p, \quad p(0, 2) = 0,$$

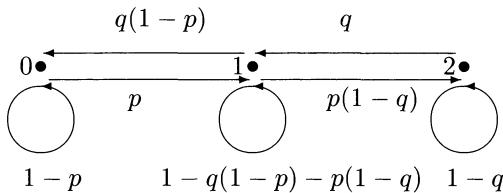
since a caller comes in with probability p (again we are assuming only one caller arrives during any time period). Also,

$$p(2, 0) = 0, \quad p(2, 1) = q, \quad p(2, 2) = 1 - q,$$

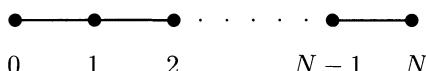
since no new callers may arrive if there are two callers in the system, and both calls may not end simultaneously. If there is exactly one caller in the system, it is a little more complicated. The state of the system goes from 1 to 0 if the current call is completed and no new callers enter the system, i.e., $p(1, 0) = q(1 - p)$. Similarly, the state goes from 1 to 2 if the current call is not completed but a new call arrives, i.e., $p(1, 2) = p(1 - q)$. Since the rows must add to 1, $p(1, 1) = 1 - q(1 - p) - p(1 - q)$ and hence

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 2 \\ 1-p & p & 0 \\ q(1-p) & 1-q(1-p)-p(1-q) & p(1-q) \\ 0 & q & 1-q \end{bmatrix}.$$

Transition probabilities are often represented by directed graphs, where the vertices of the graphs are the states and the arrows represent the transitions. The above matrix can be represented graphically as follows:



Example 3. Random Walk with Reflecting Boundary. Consider a “random walker” moving along the sites $\{0, 1, \dots, N\}$.



At each time step the walker moves one step, to the right with probability p and to the left with probability $1 - p$. If the walker is at one of the boundary points $\{0, N\}$, the walker moves with probability 1 toward the inside of the interval. The transition matrix \mathbf{P} for this Markov chain is given by

$$p(i, i+1) = p, \quad p(i, i-1) = 1 - p, \quad 0 < i < N,$$

$$p(0, 1) = 1, \quad p(N, N - 1) = 1,$$

with $p(i, j) = 0$ for other values of i, j . If $p = 1/2$, we call this *symmetric* or *unbiased* random walk with reflecting boundaries. If $p \neq 1/2$ it is called *biased* random walk. Sometimes it is more convenient to consider partially reflecting boundaries where the walker at the boundary moves the same as on the inside except that if the walker tries to leave the states $\{0, \dots, N\}$ he runs into a wall and goes nowhere. This corresponds to boundary conditions

$$p(0, 0) = 1 - p, \quad p(0, 1) = p, \quad p(N, N - 1) = 1 - p, \quad p(N, N) = p.$$

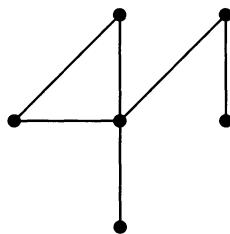
Example 4. Random Walk with Absorbing Boundaries. This chain is like the previous example except that when the walker reaches 0 or N , the walker stays there forever. The transition matrix is given by

$$p(i, i + 1) = p, \quad p(i, i - 1) = 1 - p, \quad 0 < i < N,$$

$$p(0, 0) = 1, \quad p(N, N) = 1.$$

(We adopt the convention from here on that if $p(i, j)$ is not specified for a particular i, j then it is assumed to be 0.)

Example 5. Simple Random Walk on a Graph. A (*finite, simple, undirected*) graph is a finite collection of vertices V and a collection of edges E where each edge connects two different vertices and any two vertices are connected by at most one edge. We write $v_1 \sim v_2$ if vertices v_1 and v_2 are adjacent, i.e., an edge connects the two vertices.



Consider the Markov chain whose states are the vertices of the graph. At each time interval, the chain chooses a new state randomly from among the states adjacent to the current state. The transition matrix for this chain is given by

$$p(v_i, v_j) = 1/d(v_i), \quad v_i \sim v_j,$$

where $d(v_i)$ is the number of vertices adjacent to v_i [if $d(v_i) = 0$, we let $p(v_i, v_i) = 1$]. This chain is called *simple random walk on the graph*. Symmetric random walk ($p = 1/2$) with reflecting boundaries as in Example 3 is a particular example of a simple random walk on a graph.

Given a transition matrix \mathbf{P} and an initial probability distribution ϕ , how can we determine the probability that the Markov chain will be in a certain state i at a given time n ? Define the n -step probabilities $p_n(i, j)$ by

$$p_n(i, j) = \mathbb{P}\{X_n = j \mid X_0 = i\} = \mathbb{P}\{X_{n+k} = j \mid X_k = i\}$$

(the latter equality holds because of time homogeneity). Then

$$\mathbb{P}\{X_n = j\} = \sum_{i \in S} \phi(i) \mathbb{P}\{X_n = j \mid X_0 = i\}. \quad (1.6)$$

We will now show that the n -step transition probability $p_n(i, j)$ is in fact the (i, j) entry in the matrix \mathbf{P}^n . To see this, we first note that this is trivially true for $n = 1$. Assume it is true for a given n . Then,

$$\begin{aligned} \mathbb{P}\{X_{n+1} = j \mid X_0 = i\} &= \sum_{k \in S} \mathbb{P}\{X_n = k \mid X_0 = i\} \mathbb{P}\{X_{n+1} = j \mid X_n = k\} \\ &= \sum_{k \in S} p_n(i, k) p(k, j). \end{aligned}$$

But if $p_n(i, k)$ is the (i, k) entry of \mathbf{P}^n , the last sum is exactly the (i, j) entry of $\mathbf{P}^n \mathbf{P} = \mathbf{P}^{n+1}$.

An initial probability distribution can be given by a vector

$$\bar{\phi}_0 = (\phi_0(1), \dots, \phi_0(N)).$$

[We will denote the vector $(v(1), \dots, v(N))$ by \bar{v} . We will use the same notation whether \bar{v} is to be considered a row vector or a column vector. For example, we can write either $\bar{v}\mathbf{P}$, or $\mathbf{P}\bar{v}$ although \bar{v} is a row vector in the first case and a column vector in the second.] If ϕ_0 is given, the distribution at time n , $\phi_n(i) = \mathbb{P}\{X_n = i\}$ is given by

$$\bar{\phi}_n = \bar{\phi}_0 \mathbf{P}^n.$$

Example 6. Consider Example 1 and assume the phone is free at time 0. Assume $p = 1/4$ and $q = 1/6$. Let $n = 6$. Then

$$\mathbf{P}^6 = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}^6 = \begin{bmatrix} .424 & .576 \\ .384 & .616 \end{bmatrix}.$$

If the phone is free at time 0, $\bar{\phi}_0 = (1, 0)$. If we want to know the probability that the phone is busy at time 6 given that it was free at time 0, we compute

$$(\bar{\phi}_0 \mathbf{P}^6)(1) = .576.$$

1.2 Large-Time Behavior and Invariant Probability

Understanding the large-time behavior of a Markov chain boils down to understanding the behavior of \mathbf{P}^n for large n values. Let us start by considering a particular example,

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}.$$

Taking powers of this matrix is easy (with a computer) and one can quickly see that

$$\mathbf{P}^n \approx \begin{bmatrix} .4 & .6 \\ .4 & .6 \end{bmatrix},$$

for large n , i.e., a limit matrix

$$\Pi = \lim_{n \rightarrow \infty} \mathbf{P}^n$$

exists and the rows of Π are identical. If \bar{v} is any probability vector [we say a vector $\bar{v} = (v(1), \dots, v(N))$ is a probability vector if the components are nonnegative and sum to 1], then

$$\lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^n = \bar{\pi},$$

where $\bar{\pi} = (2/5, 3/5)$ is one of the rows of Π . For another example, consider Example 2 of Section 1.1 with $p = 1/4, q = 1/6$,

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/8 & 2/3 & 5/24 \\ 0 & 1/6 & 5/6 \end{bmatrix}. \quad (1.7)$$

We see the same phenomenon. In this case for large n ,

$$\mathbf{P}^n \approx \begin{bmatrix} .182 & .364 & .455 \\ .182 & .364 & .455 \\ .182 & .364 & .455 \end{bmatrix} = \begin{bmatrix} \bar{\pi} \\ \bar{\pi} \\ \bar{\pi} \end{bmatrix},$$

where $\bar{\pi} = (2/11, 4/11, 5/11)$ and hence for every probability vector \bar{v} ,

$$\lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^n = \bar{\pi}.$$

At any large time, the probability that the phone has no callers is about $\pi(0) = 2/11$, regardless of what the state of the system was at time 0.

Suppose $\bar{\pi}$ is a limiting probability vector, i.e., for some initial probability vector \bar{v} ,

$$\bar{\pi} = \lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^n.$$

Then

$$\bar{\pi} = \lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^{n+1} = (\lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^n)\mathbf{P} = \bar{\pi}\mathbf{P}.$$

We call a probability vector $\bar{\pi}$ an *invariant probability distribution* for \mathbf{P} if

$$\bar{\pi} = \bar{\pi}\mathbf{P}. \quad (1.8)$$

Such a $\bar{\pi}$ is also called a *stationary*, *equilibrium*, or *steady-state* probability distribution. Note that an invariant probability vector is a left eigenvector of \mathbf{P} with eigenvalue 1.

There are three natural questions to ask about invariant probability distributions for stochastic matrices:

- 1) Does every stochastic matrix \mathbf{P} have an invariant probability distribution $\bar{\pi}$?
- 2) Is the invariant probability distribution unique?
- 3) When can we conclude that

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \begin{bmatrix} \bar{\pi} \\ \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix},$$

and hence that for all initial probability distributions \bar{v} ,

$$\lim_{n \rightarrow \infty} \bar{v}\mathbf{P}^n = \bar{\pi}?$$

Let us start by considering the two-state Markov chain with

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix},$$

where $0 < p, q < 1$. This matrix has eigenvalues 1 and $1 - p - q$. We can diagonalize \mathbf{P} ,

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{P}\mathbf{Q},$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & -p \\ 1 & q \end{bmatrix}, \quad \mathbf{Q}^{-1} = \begin{bmatrix} q/(p+q) & p/(p+q) \\ -1/(p+q) & 1/(p+q) \end{bmatrix}.$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1-p-q \end{bmatrix}.$$

The columns of \mathbf{Q} are right eigenvectors of \mathbf{P} and the rows of \mathbf{Q}^{-1} are left eigenvectors. The eigenvectors are unique up to a multiplicative constant.

We have chosen the constant in the left eigenvector for eigenvalue 1 so that it is a probability vector. $\bar{\pi} = (q/(p+q), p/(p+q))$ is the unique invariant probability distribution for \mathbf{P} . Once \mathbf{P} is diagonalized it is easy to raise \mathbf{P} to powers,

$$\begin{aligned}\mathbf{P}^n &= (\mathbf{Q}\mathbf{D}\mathbf{Q}^{-1})^n \\ &= \mathbf{Q}\mathbf{D}^n\mathbf{Q}^{-1} \\ &= \mathbf{Q} \begin{bmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{bmatrix} \mathbf{Q}^{-1} \\ &= \begin{bmatrix} [q+p(1-p-q)^n]/(p+q) & [p-p(1-p-q)^n]/(p+q) \\ [q-q(1-p-q)^n]/(p+q) & [p+q(1-p-q)^n]/(p+q) \end{bmatrix}.\end{aligned}$$

Since $|1-p-q| < 1$, we see that

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \begin{bmatrix} q/(p+q) & p/(p+q) \\ q/(p+q) & p/(p+q) \end{bmatrix} = \begin{bmatrix} \bar{\pi} \\ \bar{\pi} \end{bmatrix}.$$

The key to the computation of the limit is the fact that the second eigenvalue $1-p-q$ has absolute value less than 1 and so the dominant contribution to \mathbf{P}^n comes from the eigenvector with eigenvalue 1, i.e., the invariant probability distribution.

Suppose \mathbf{P} is any stochastic matrix. It is easy to check that the vector $\bar{1} = (1, 1, \dots, 1)$ is a *right* eigenvector with eigenvalue 1. Hence at least one left eigenvector for eigenvalue 1 exists. Suppose we can show that:

The left eigenvector can be chosen to have all nonnegative entries, (1.9)

The eigenvalue 1 is simple and all other eigenvalues

have absolute value less than 1. (1.10)

Then we can show that essentially the same thing happens as in the two-state case. It is not always true that we can diagonalize \mathbf{P} ; however, we can do well enough using a Jordon decomposition (consult a text in linear algebra for details): there exists a matrix \mathbf{Q} such that

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{P}\mathbf{Q},$$

where the first row of \mathbf{Q}^{-1} is the unique invariant probability vector $\bar{\pi}$; the first column of \mathbf{Q} contains all 1s. The matrix \mathbf{D} is not necessarily diagonal but it does have the form

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{M} & \\ 0 & & & \end{bmatrix},$$

where $\mathbf{M}^n \rightarrow 0$. Then in the same way as the two-state example,

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \lim_{n \rightarrow \infty} \mathbf{Q} \mathbf{D}^n \mathbf{Q}^{-1} = \mathbf{Q} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{bmatrix} \mathbf{Q}^{-1} = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}.$$

This leads us to ask which matrices satisfy (1.9) and (1.10). The Perron–Frobenius Theorem from linear algebra gives one large class of matrices for which this is true. Suppose that \mathbf{P} is a stochastic matrix such that all of the entries are strictly positive. Then the Perron–Frobenius Theorem implies that: 1 is a simple eigenvalue for \mathbf{P} ; the left eigenvector of 1 can be chosen to have all positive entries (and hence can be made into a probability vector by multiplying by an appropriate constant); and all the other eigenvalues have absolute value strictly less than 1. We sketch a proof of this theorem in Exercise 1.20.

While this includes a large number of matrices, it does not cover all stochastic matrices with the appropriate limit behavior. For example, consider the matrix \mathbf{P} in (1.7). Although \mathbf{P} does not have all positive entries, note that

$$\mathbf{P}^2 = \begin{bmatrix} .594 & .354 & .052 \\ .177 & .510 & .312 \\ .021 & .250 & .729 \end{bmatrix},$$

and hence \mathbf{P}^2 satisfies the conditions of the theorem. Therefore, 1 is a simple eigenvalue for \mathbf{P}^2 with invariant probability $\bar{\pi}$ and the other eigenvalues of \mathbf{P}^2 have absolute value strictly less than 1. Since the eigenvalues for \mathbf{P}^2 are the squares of the eigenvalues of \mathbf{P} , and eigenvectors of \mathbf{P} are eigenvectors of \mathbf{P}^2 , we see that \mathbf{P} also satisfies (1.9) and (1.10). We then get a general rule.

Fact. *If \mathbf{P} is a stochastic matrix such that for some n , \mathbf{P}^n has all entries strictly positive, then \mathbf{P} satisfies (1.9) and (1.10).*

In the next section we classify all stochastic matrices \mathbf{P} that have the property that \mathbf{P}^n has all positive entries for some n .

1.3 Classification of States

In this section we investigate under what conditions on a stochastic matrix \mathbf{P} we can conclude that \mathbf{P}^n has all positive entries for some sufficiently large n . We start by considering some examples where this is not true.

Example 1. Simple random walk with reflecting boundary on $\{0, \dots, 4\}$. In this case,

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

If one takes powers of this matrix, one quickly sees that \mathbf{P}^n looks different depending on whether n is even or odd. For large n , if n is even,

$$\mathbf{P}^n \approx \begin{bmatrix} .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \end{bmatrix},$$

whereas if n is odd,

$$\mathbf{P}^n \approx \begin{bmatrix} 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \end{bmatrix}.$$

It is easy to see why there should be many zeroes in \mathbf{P}^n . At each step, the random walker moves from an “even” step to an “odd” step or vice versa. If the walker starts on an even site, then after an even number of steps the walker will be on an even site, i.e., $p_n(i, j) = 0$ if i is even, j is odd, n is even. Similarly, after an odd number of steps, a walker who started on an even point will be at an odd point. In this example we say that \mathbf{P} has period 2.

Example 2. Simple random walk with absorbing boundary on $\{0, \dots, 4\}$. Here,

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

If n is large, we see that

$$\mathbf{P}^n \approx \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ .75 & 0 & 0 & 0 & .25 \\ .50 & 0 & 0 & 0 & .50 \\ .25 & 0 & 0 & 0 & .75 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In this case the random walker eventually gets to 0 or 4 and then stays at that state forever. Look at the second row and observe that $p_n(1, 0) \rightarrow 3/4$ and $p_n(1, 4) \rightarrow 1/4$. This implies that the probability that a random walker starting at 1 will eventually stick at 0 is $3/4$, whereas with probability $1/4$ she eventually sticks at 4. We will call states such as 1, 2, 3 transient states of the Markov chain.

Example 3. Suppose $S = \{1, 2, 3, 4, 5\}$ and

$$\mathbf{P} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1/2 & 1/2 & 0 & 0 & 0 \\ 2 & 1/6 & 5/6 & 0 & 0 & 0 \\ 3 & 0 & 0 & 3/4 & 1/4 & 0 \\ 4 & 0 & 0 & 1/8 & 2/3 & 5/24 \\ 5 & 0 & 0 & 0 & 1/6 & 5/6 \end{bmatrix}.$$

For large n ,

$$\mathbf{P}^n \approx \begin{bmatrix} .25 & .75 & 0 & 0 & 0 \\ .25 & .75 & 0 & 0 & 0 \\ 0 & 0 & .182 & .364 & .455 \\ 0 & 0 & .182 & .364 & .455 \\ 0 & 0 & .182 & .364 & .455 \end{bmatrix}.$$

In this case the chain splits into two smaller, noninteracting chains: a chain with state space $\{1, 2\}$ and a chain with state space $\{3, 4, 5\}$. Each “subchain” converges to an equilibrium distribution, but one cannot change from a state in $\{1, 2\}$ to a state in $\{3, 4, 5\}$. We call such a system a reducible Markov chain.

The main goal of this section is to show that the above examples illustrate all the ways that a Markov chain can fail to satisfy (1.9) and (1.10).

1.3.1 Reducibility

We say two states i and j of a Markov chain *communicate* with each other, written $i \leftrightarrow j$, if there exist $m, n \geq 0$ such that $p_m(i, j) > 0$ and $p_n(j, i) > 0$. In other words, two states communicate if and only if each state has a positive probability of eventually being reached by a chain starting in the other state.

The relation \leftrightarrow is an equivalence relation on the state space, i.e., it is: *reflexive*, $i \leftrightarrow i$ [since $p_0(i, i) = 1 > 0$]; *symmetric*, $i \leftrightarrow j$ implies that $j \leftrightarrow i$ (this is immediate from the definition); and *transitive*, $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$. To see that transitivity holds, note that if $p_{m_1}(i, j) > 0$ and $p_{m_2}(j, k) > 0$ then

$$\begin{aligned} p_{m_1+m_2}(i, k) &= \mathbb{P}\{X_{m_1+m_2} = k \mid X_0 = i\} \\ &\geq \mathbb{P}\{X_{m_1+m_2} = k, X_{m_1} = j \mid X_0 = i\} \\ &= \mathbb{P}\{X_{m_1} = j \mid X_0 = i\} \mathbb{P}\{X_{m_1+m_2} = k \mid X_{m_1} = j\} \\ &= p_{m_1}(i, j) p_{m_2}(j, k) > 0, \end{aligned}$$

and similarly $p_{n_1}(j, i) > 0, p_{n_2}(k, j) > 0$ imply $p_{n_1+n_2}(k, i) > 0$. This equivalence relation partitions the state space into disjoint sets called *communication classes*. For example, in Example 3 of this section there are two communication classes $\{1, 2\}$ and $\{3, 4, 5\}$.

If there is only one communication class, i.e., if for all i, j there exists an $n = n(i, j)$ with $p_n(i, j) > 0$, then the chain is called *irreducible*. Any matrix satisfying (1.9) and (1.10) is irreducible. However, one can also check that Example 1 of this section is also irreducible. Example 2 has three communication classes, $\{0\}$, $\{1, 2, 3\}$, and $\{4\}$. In this example, if the chain starts in the class $\{1, 2, 3\}$, then with probability 1 it eventually leaves this class and never returns. Classes with this property are called *transient* classes and the states are called transient states. Other classes are called *recurrent* classes with recurrent states. A Markov chain starting in a recurrent class never leaves that class.

Suppose \mathbf{P} is the matrix for a reducible Markov chain with recurrent communication classes R_1, \dots, R_r and transient classes T_1, \dots, T_s . It is easy to see that there must be at least one recurrent class. For each recurrent class R , the submatrix of \mathbf{P} obtained from considering only the rows and columns for states in R is a stochastic matrix. Hence we can write \mathbf{P} in the following form (after, perhaps, reordering the states):

$$\mathbf{P} = \left[\begin{array}{cc|c} \mathbf{P}_1 & & \mathbf{0} & \mathbf{0} \\ & \mathbf{P}_2 & & \\ & & \mathbf{P}_3 & \\ & \mathbf{0} & & \ddots \\ & & & \mathbf{P}_r \\ \hline \mathbf{S} & & & \mathbf{Q} \end{array} \right],$$

where \mathbf{P}_k is the matrix associated with R_k . Then,

$$\mathbf{P}^n = \left[\begin{array}{cc|c} \mathbf{P}_1^n & & \\ \mathbf{P}_2^n & \mathbf{0} & \\ \mathbf{P}_3^n & & \\ & \ddots & \\ \mathbf{0} & & \mathbf{P}_r^n \\ \hline \mathbf{S}_n & & \mathbf{Q}^n \end{array} \right],$$

for some matrix \mathbf{S}_n . To analyze the large time behavior of the Markov chain on the class R_k we need only consider the matrix \mathbf{P}_k . We discuss the behavior of \mathbf{Q}^n in Section 1.5.

1.3.2 Periodicity

Suppose that \mathbf{P} is the matrix for an irreducible Markov chain (if \mathbf{P} is reducible we can consider separately each of the recurrent communication classes). We define the *period* of a state i , $d = d(i)$, to be the greatest common divisor of

$$J_i := \{n \geq 0 : p_n(i, i) > 0\}.$$

In Example 1 of this section, the period of each state is 2; in fact, in this case $p_{2n}(i, i) > 0$ and $p_{2n+1}(i, i) = 0$ for all n, i .

Suppose J is any nonempty subset of the nonnegative integers that is closed under addition, i.e., $m, n \in J \Rightarrow m + n \in J$. An example of such a J is the set J_i since $p_{m+n}(i, i) \geq p_m(i, i)p_n(i, i)$. Let d be the greatest common divisor of the elements of J . Then $J \subset \{0, d, 2d, \dots\}$. Moreover, it can be shown (Exercise 1.21) that J must contain all but a finite number of the elements of $\{0, d, 2d, \dots\}$, i.e., there is some M such that $md \in J$ for all $m > M$. Hence J_i contains md for all m greater than some $M = M_i$. If j is another state and m, n are such that $p_m(i, j) > 0, p_n(j, i) > 0$, then $m + n \in J_i, m + n \in J_j$. Hence $m + n = kd$ for some integer k . Also, if $l \in J_j$, then

$$p_{m+n+l}(i, j) \geq p_m(i, j)p_l(j, j)p_n(j, i) > 0,$$

and so d divides l . We have just shown that if d divides every element of J_i then it divides every element of J_j . From this we see that all states have the same period and hence we can talk about the period of \mathbf{P} . (We have used the fact that \mathbf{P} is irreducible. If \mathbf{P} is reducible, it is possible for states in different communication classes to have different periods.)

Example 4. Consider simple random walk on a graph (see Example 5, Section 1.1). The chain is irreducible if and only if the graph is connected, i.e., if any two vertices can be connected by a path of edges in the graph. Every vertex in a connected graph (with at least two vertices) is adjacent to at least

one other point. If $v \sim w$ then $p_2(v, v) \geq p_1(v, w)p_1(w, v) > 0$. Therefore, the period is either 1 or 2. It is easy to see that the period is 2 if and only if the graph is bipartite, i.e., if and only if the vertices can be partitioned into two disjoint sets V_1, V_2 such that all edges of the graph connect one vertex of V_1 and one vertex V_2 . Note that symmetric random walk with reflecting boundaries gives an example of simple random walk on a bipartite graph.

1.3.3 Irreducible, aperiodic chains

We call an irreducible matrix \mathbf{P} *aperiodic* if $d = 1$. What we will show now is the following: if \mathbf{P} is irreducible and aperiodic, then there exists an $M > 0$ such that for all $n \geq M$, \mathbf{P}^n has all entries strictly positive. To see this, take any i, j . Since \mathbf{P} is irreducible there exists some $m(i, j)$ such that $p_{m(i,j)}(i, j) > 0$. Moreover, since \mathbf{P} is aperiodic, there exists some $M(i)$ such that for all $n \geq M(i)$, $p_n(i, i) > 0$. Hence for all $n \geq M(i)$,

$$p_{n+m(i,j)}(i, j) \geq p_n(i, i)p_{m(i,j)}(i, j) > 0.$$

Let M be the maximum value of $M(i) + m(i, j)$ over all pairs (i, j) (the maximum exists since the state space is finite). Then $p_n(i, j) > 0$ for all $n \geq M$ and all i, j . Using the rule at the end of Section 1.2 we can now summarize with the following theorem.

Theorem. *If \mathbf{P} is the transition matrix for an irreducible, aperiodic Markov chain, then there exists a unique invariant probability vector $\bar{\pi}$ satisfying*

$$\bar{\pi}\mathbf{P} = \bar{\pi}.$$

If $\bar{\phi}$ is any initial probability vector,

$$\lim_{n \rightarrow \infty} \bar{\phi}\mathbf{P}^n = \bar{\pi}.$$

Moreover, $\pi(i) > 0$ for each i .

1.3.4 Reducible or periodic chains

We finish this section by discussing how \mathbf{P}^n behaves when \mathbf{P} is not irreducible and aperiodic. First, assume \mathbf{P} is reducible with recurrent classes R_1, \dots, R_r and transient classes T_1, \dots, T_s . Each recurrent class acts as a small Markov chain; hence, there exists r different invariant probability vectors $\bar{\pi}^1, \dots, \bar{\pi}^r$ with $\bar{\pi}^k$ concentrated on R_k ($\pi^k(i) = 0$ if $i \notin R_k$). In other words, the eigenvalue 1 has multiplicity r with one eigenvector for each recurrent class. Assume, for ease, that the submatrix \mathbf{P}_k for each recurrent class is aperiodic. Then if $i \in R_k$,

$$\lim_{n \rightarrow \infty} p_n(i, j) = \pi^k(j), \quad j \in R_k,$$

$$p_n(i, j) = 0, \quad j \notin R_k.$$

If i is any transient state, then the chain starting at i eventually ends up in a recurrent state. This means that for each transient state j ,

$$\lim_{n \rightarrow \infty} p_n(i, j) = 0.$$

Let $\alpha_k(i), k = 1, \dots, r$ be the probability that the chain starting in state i eventually ends up in recurrent class R_k [in Section 1.5 we will discuss how to calculate $\alpha_k(i)$]. Once the chain reaches a state in R_k it will settle down to the equilibrium distribution on R_k . From this we see that if $j \in R_k$,

$$\lim_{n \rightarrow \infty} p_n(i, j) = \alpha_k(i) \pi^k(j).$$

If $\bar{\phi}$ is an initial probability vector,

$$\lim_{n \rightarrow \infty} \bar{\phi} \mathbf{P}^n$$

exists but depends on $\bar{\phi}$.

Suppose now that \mathbf{P} is irreducible but has period $d > 1$. In this case the state space splits nicely into d sets, A_1, \dots, A_d , such that the chain always moves from A_i to A_{i+1} (or A_d to A_1). To illustrate the large-time behavior of \mathbf{P}^n , we will consider Example 1 of this section which has period 2. Let

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The eigenvalues for \mathbf{P} are $1, -1, 0, 1/\sqrt{2}, -1/\sqrt{2}$. The eigenvalue 1 is simple and there is a unique invariant probability $\bar{\pi} = (1/8, 1/4, 1/4, 1/4, 1/8)$. However, when powers of \mathbf{P} are taken the eigenvector for -1 becomes important as well as $\bar{\pi}$. We can diagonalize \mathbf{P} ,

$$\mathbf{D} = \mathbf{Q}^{-1} \mathbf{P} \mathbf{Q},$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & -1/2 & 1/4 & -1 & \sqrt{2}/4 \\ 1 & 1/2 & 0 & -\sqrt{2}/2 & -1/4 \\ 1 & -1/2 & -1/4 & 0 & 0 \\ 1 & 1/2 & 0 & \sqrt{2}/2 & 1/4 \\ 1 & -1/2 & 1/4 & 1 & -\sqrt{2}/4 \end{bmatrix},$$

$$\mathbf{Q}^{-1} = \begin{bmatrix} 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ -1/4 & 1/2 & -1/2 & 1/2 & -1/4 \\ 1 & 0 & -2 & 0 & 1 \\ -1/4 & -\sqrt{2}/4 & 0 & \sqrt{2}/4 & 1/4 \\ \sqrt{2}/2 & -1 & 0 & 1 & -\sqrt{2}/2 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & -1/\sqrt{2} \end{bmatrix}.$$

We then see that for \mathbf{P}^n , the eigenvectors for the three eigenvalues with absolute value less than 1 become irrelevant and for large n

$$\mathbf{P}^n \approx \begin{bmatrix} 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \end{bmatrix} + (-1)^n \begin{bmatrix} 1/8 & -1/4 & 1/4 & -1/4 & 1/8 \\ -1/8 & 1/4 & -1/4 & 1/4 & -1/8 \\ 1/8 & -1/4 & 1/4 & -1/4 & 1/8 \\ -1/8 & 1/4 & -1/4 & 1/4 & -1/8 \\ 1/8 & -1/4 & 1/4 & -1/4 & 1/8 \end{bmatrix}.$$

The asymptotic value for \mathbf{P}^n varies depending on whether n is even or odd. In this case the invariant probability at a state i , $\pi(i)$, does not represent the limit of $p_n(j, i)$. However, it does represent the average amount of time that is spent in site i . In fact, one can check that for large n , the average of $p_n(j, i)$ and $p_{n+1}(j, i)$ approaches $\pi(i)$ for each initial state j ,

$$\pi(i) = \lim_{n \rightarrow \infty} \frac{1}{2} [p_n(j, i) + p_{n+1}(j, i)].$$

In general, if \mathbf{P} is irreducible with period d , \mathbf{P} will have d eigenvalues with absolute value 1, the d complex numbers z with $z^d = 1$. Each is simple; in particular, the eigenvalue 1 is simple and there exists a unique invariant probability $\bar{\pi}$. Given any initial probability distribution $\bar{\phi}$, for large n , $\bar{\phi}\mathbf{P}^n$ will cycle through d different distributions, but they will average to $\bar{\pi}$,

$$\lim_{n \rightarrow \infty} \frac{1}{d} [\bar{\phi}\mathbf{P}^{n+1} + \cdots + \bar{\phi}\mathbf{P}^{n+d}] = \bar{\pi}.$$

1.4 Return Times

Let X_n be an irreducible (but perhaps periodic) Markov chain with transition matrix \mathbf{P} . Consider the amount of time spent in state j up to and

including time n ,

$$Y(j, n) = \sum_{m=0}^n I\{X_m = j\}.$$

Here we write I to denote the “indicator function” of an event, i.e., the random variable which equals 1 if the event occurs and 0 otherwise. If $\bar{\pi}$ denotes the invariant probability distribution for \mathbf{P} , then it follows from the results of the previous sections that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E}(Y(j, n) | X_0 = i) &= \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{m=0}^n \mathbb{P}\{X_m = j | X_0 = i\} \\ &= \pi(j), \end{aligned}$$

i.e., $\pi(j)$ represents the fraction of time that the chain spends in state j . In this section we relate $\pi(j)$ to the first return time to the state j .

Fix a state i and assume that $X_0 = i$. Let T be the first time after 0 that the Markov chain is in state i ,

$$T = \min\{n \geq 1 : X_n = i\}.$$

Since the chain is irreducible, we know that $T < \infty$ with probability 1. In fact (see Exercise 1.7) it is not too difficult to show that $\mathbb{E}(T) < \infty$.

Consider the time until the k th return to the state i . This time is given by a sum of independent random variables, $T_1 + \dots + T_k$, each with the distribution of T . Here, T_m denotes the time between the $(m-1)$ st and m th return. For k large, the law of large numbers tells us that

$$\frac{1}{k}(T_1 + \dots + T_k) \approx \mathbb{E}(T),$$

i.e., there are about k visits to the state i in $k\mathbb{E}(T)$ steps of the chain. But we have already seen that in n steps we expect about $n\pi(i)$ visits to the state i . Hence setting $n = k\mathbb{E}(T)$ we get the relation

$$\mathbb{E}(T) = \frac{1}{\pi(i)}. \quad (1.11)$$

This says that the expected number of steps to return to i , assuming that the chain starts at i , is given by the reciprocal of the invariant probability. The above argument is, of course, not completely rigorous, but it does not take too much work to supply the details to prove that (1.11) always holds. See Exercise 1.15 for another derivation of (1.11).

Example. Consider the two-state Markov chain with $S = \{0, 1\}$ and

$$\mathbf{P} = \begin{smallmatrix} 0 & 1 \\ 1-p & p \\ q & 1-q \end{smallmatrix}, \quad 0 < p, q < 1.$$

Assume the chain starts in state 0 and let T be the return time to 0. In Section 1.2, we showed that $\bar{\pi} = (q/(p+q), p/(p+q))$ and hence

$$\mathbb{E}(T) = \frac{1}{\pi(0)} = \frac{p+q}{q} \quad (1.12)$$

In this example we can write down the distribution for T explicitly and verify (1.12). For $n > 1$,

$$\mathbb{P}\{T \geq n\} = \mathbb{P}\{X_1 = 1, \dots, X_{n-1} = 1 \mid X_0 = 0\} = p(1-q)^{n-2}.$$

If Y is any random variable taking values in the nonnegative integers,

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{n=1}^{\infty} n \mathbb{P}\{Y = n\} = \sum_{n=1}^{\infty} \sum_{k=1}^n \mathbb{P}\{Y = n\} \\ &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{P}\{Y = n\} = \sum_{k=1}^{\infty} \mathbb{P}\{Y \geq k\}. \end{aligned} \quad (1.13)$$

Therefore,

$$\begin{aligned} \mathbb{E}(T) &= \sum_{n=1}^{\infty} n \mathbb{P}\{T = n\} = \sum_{n=1}^{\infty} \mathbb{P}\{T \geq n\} \\ &= 1 + \sum_{n=2}^{\infty} p(1-q)^{n-2} = \frac{p+q}{q}. \end{aligned}$$

It should be pointed out that (1.11) only gives the expected value of the random variable T and says nothing else about its distribution. In general, one can say very little else about the distribution of T given only the invariant probability $\bar{\pi}$. To illustrate this, consider the two-state example above with $p = q$ so that $\mathbb{E}(T) = 2$. If p is close to 1, then $T = 2$ most of the time and $\text{Var}(T)$ is small. If p is close to 0, then $T = 1$ most of the time, but occasionally T takes on a very high value. In this case, $\text{Var}(T)$ is large.

In the next section, we discuss how to compute the expected number of steps from i to j when $i \neq j$.

1.5 Transient States

Let \mathbf{P} be the transition matrix for a Markov chain X_n . Recall that a state i is called transient if with probability 1 the chain visits i only a finite number

of times. Suppose \mathbf{P} has some transient states and let \mathbf{Q} be the submatrix of \mathbf{P} which includes only the rows and columns for the transient states. Hence (after rearranging the order of the states) we can write

$$\mathbf{P} = \left[\begin{array}{c|c} \tilde{\mathbf{P}} & \mathbf{0} \\ \hline \mathbf{S} & \mathbf{Q} \end{array} \right], \quad \mathbf{P}^n = \left[\begin{array}{c|c} \tilde{\mathbf{P}}^n & \mathbf{0} \\ \hline \mathbf{S}_n & \mathbf{Q}^n \end{array} \right].$$

As an example, we consider the random walk with absorbing boundaries (Example 2, Section 1.3). We order the state space $S = \{0, 4, 1, 2, 3\}$ so that we can write

$$\mathbf{P} = \left[\begin{array}{ccccc} 0 & 4 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 & 0 \\ 1 & 1/2 & 0 & 1/2 & 0 \\ 2 & 0 & 0 & 1/2 & 0 \\ 3 & 0 & 1/2 & 0 & 1/2 \end{array} \right], \quad \mathbf{Q} = \left[\begin{array}{ccc} 1 & 2 & 3 \\ \hline 1 & 0 & 1/2 & 0 \\ 2 & 1/2 & 0 & 1/2 \\ 3 & 0 & 1/2 & 0 \end{array} \right]. \quad (1.14)$$

The matrix \mathbf{Q} is a *substochastic matrix*, i.e., a matrix with nonnegative entries whose row sums are less than or equal to 1. Since the states represented by \mathbf{Q} are transient, $\mathbf{Q}^n \rightarrow \mathbf{0}$. This implies that all of the eigenvalues of \mathbf{Q} have absolute values strictly less than 1. Hence, $\mathbf{I} - \mathbf{Q}$ is an invertible matrix and there is no problem in defining the matrix

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}.$$

Let i be a transient state and consider Y_i , the total number of visits to i ,

$$Y_i = \sum_{n=0}^{\infty} I\{X_n = i\}.$$

Since i is transient, $Y_i < \infty$ with probability 1. Suppose $X_0 = j$, where j is another transient state. Then,

$$\begin{aligned} \mathbb{E}(Y_i | X_0 = j) &= \mathbb{E}\left[\sum_{n=0}^{\infty} I\{X_n = i\} | X_0 = j\right] \\ &= \sum_{n=0}^{\infty} \mathbb{P}\{X_n = i | X_0 = j\} \\ &= \sum_{n=0}^{\infty} p_n(j, i). \end{aligned}$$

In other words, $\mathbb{E}(Y_i | X_0 = j)$ is the (j, i) entry of the matrix

$$\mathbf{I} + \mathbf{P} + \mathbf{P}^2 + \dots$$

which is the same as the (j, i) entry of the matrix $\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots$. However, a simple calculation shows that

$$(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots)(\mathbf{I} - \mathbf{Q}) = \mathbf{I},$$

or

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots = (\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{M}.$$

We have just shown that the expected number of visits to i starting at j is given by \mathbf{M}_{ji} , the (j, i) entry of \mathbf{M} . If we want to compute the expected number of steps until the chain enters a recurrent class, assuming $X_0 = j$, we need only sum \mathbf{M}_{ji} over all transient states i .

In the particular example (1.14),

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 3/2 & 1 & 1/2 \\ 2 & 1 & 2 & 1 \\ 3 & 1/2 & 1 & 3/2 \end{array}.$$

Starting in state 1, the expected number of visits to state 3 before absorption is $1/2$, and the expected total number of steps until absorption is $3/2 + 1 + 1/2 = 3$.

We can also use this technique to determine the expected number of steps that an irreducible Markov chain takes to go from one state j to another state i . We first write the transition matrix \mathbf{P} for the chain with i being the first site:

$$\mathbf{P} = \left[\frac{p(i, i)}{\mathbf{S}} \middle| \mathbf{R} \right].$$

We then change i to an absorbing site, and hence have the new matrix

$$\tilde{\mathbf{P}} = \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{S} & \mathbf{Q} \end{array} \right].$$

Let T_i be the number of steps needed to reach state i . In other words, T_i is the smallest time n such that $X_n = i$. For any other state k let $T_{i,k}$ be the number of visits to k before reaching i (if we start at state k , we include this as one visit to k). Then,

$$\mathbb{E}(T_i | X_0 = j) = \mathbb{E} \left[\sum_{k \neq i} T_{i,k} | X_0 = j \right] = \sum_{k \neq i} \mathbf{M}_{jk}.$$

In other words, $\bar{\mathbf{M}}_1$ gives a vector whose j th component is the number of steps starting at j until reaching i .

Example 1. Suppose \mathbf{P} is the matrix for random walk with reflecting boundary,

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

If we let $i = 0$, then

$$\mathbf{Q} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 1/2 & 0 & 0 \\ 2 & 1/2 & 0 & 1/2 & 0 \\ 3 & 0 & 1/2 & 0 & 1/2 \\ 4 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 2 & 2 & 1 \\ 2 & 2 & 4 & 4 & 2 \\ 3 & 2 & 4 & 6 & 3 \\ 4 & 2 & 4 & 6 & 4 \end{bmatrix},$$

$$\mathbf{M}\bar{1} = (7, 12, 15, 16).$$

Hence, the expected number of steps to get from 4 to 0 is 16.

We now suppose that there are at least two different recurrent classes and ask the question: starting at a given transient state j , what is the probability that the Markov chain eventually ends up in a particular recurrent class? In order to answer this question, we can assume that the recurrent classes consist of single points r_1, \dots, r_k with $p(r_i, r_i) = 1$. If we order the states so that the recurrent states r_1, \dots, r_k precede the transient states t_1, \dots, t_s , then

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{S} & \mathbf{Q} \end{array} \right].$$

For $i = 1, \dots, s$, $j = 1, \dots, k$, let $\alpha(t_i, r_j)$ be the probability that the chain starting at t_i eventually ends up in recurrent state r_j . We set $\alpha(r_i, r_i) = 1$ and $\alpha(r_i, r_j) = 0$ if $i \neq j$. For any transient state t_i ,

$$\begin{aligned} \alpha(t_i, r_j) &= \mathbb{P}\{X_n = r_j \text{ eventually} \mid X_0 = t_i\} \\ &= \sum_{x \in S} \mathbb{P}\{X_1 = x \mid X_0 = t_i\} \mathbb{P}\{X_n = r_j \text{ eventually} \mid X_1 = x\} \\ &= \sum_{x \in S} p(t_i, x) \alpha(x, r_j). \end{aligned}$$

If \mathbf{A} is the $s \times k$ matrix with entries $\alpha(t_i, r_j)$, then the above can be written in matrix form

$$\mathbf{A} = \mathbf{S} + \mathbf{Q}\mathbf{A},$$

or

$$\mathbf{A} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{S} = \mathbf{MS}.$$

Example 2. Consider a random walk with absorbing boundary on $\{0, \dots, 4\}$. If we order the states $\{0, 4, 1, 2, 3\}$ so that the recurrent states precede the transient states then

$$\mathbf{P} = \begin{array}{c|cc|ccc} & 0 & 4 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 & 0 & 0 \\ \hline 1 & 1/2 & 0 & 0 & 1/2 & 0 \\ 2 & 0 & 0 & 1/2 & 0 & 1/2 \\ 3 & 0 & 1/2 & 0 & 1/2 & 0 \end{array}$$

$$\mathbf{S} = \begin{array}{c} 0 \quad 4 \\ 1 \left[\begin{matrix} 1/2 & 0 \end{matrix} \right] \\ 2 \left[\begin{matrix} 0 & 0 \end{matrix} \right] \\ 3 \left[\begin{matrix} 0 & 1/2 \end{matrix} \right] \end{array}, \quad \mathbf{M} = \begin{array}{c} 1 \quad 2 \quad 3 \\ 2 \left[\begin{matrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \end{matrix} \right] \\ 3 \left[\begin{matrix} 1/2 & 1 & 3/2 \end{matrix} \right] \end{array}, \quad \mathbf{MS} = \begin{array}{c} 0 \quad 4 \\ 2 \left[\begin{matrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{matrix} \right] \\ 3 \left[\begin{matrix} 1/4 & 3/4 \end{matrix} \right] \end{array}.$$

Hence, starting at state 1 the probability that the the walk is eventually absorbed at state 0 is $3/4$.

Example 3. Gambler's Ruin. Consider the random walk with absorbing boundary on $\{0, \dots, N\}$. Let $\alpha(j) = \alpha(j, N)$ be the probability that the walker starting at state j eventually ends up absorbed in state N . Clearly, $\alpha(0) = 0, \alpha(N) = 1$. For $0 < j < N$, we can consider one step as above and note that

$$\alpha(j) = (1-p)\alpha(j-1) + p\alpha(j+1). \quad (1.15)$$

This gives us $N-1$ linear equations in $N-1$ unknowns, $\alpha(1), \dots, \alpha(N-1)$. To find the solution, we need to know how to solve linear difference equations. By (0.5) and (0.6), the general solution of (1.15) is

$$\alpha(j) = c_1 + c_2 \left(\frac{1-p}{p} \right)^j, \quad p \neq 1/2,$$

$$\alpha(j) = c_1 + c_2 j, \quad p = 1/2.$$

The boundary conditions $\alpha(0) = 0, \alpha(N) = 1$ allow us to determine the constants c_1, c_2 , so we get

$$\alpha(j) = \frac{1 - (\frac{1-p}{p})^j}{1 - (\frac{1-p}{p})^N}, \quad p \neq 1/2,$$

$$\alpha(j) = \frac{j}{N}, \quad p = 1/2. \quad (1.16)$$

Note that if $p \leq 1/2$, then for any fixed j ,

$$\lim_{N \rightarrow \infty} \alpha(j) = 0.$$

This says that if a gambler with fixed resources j plays a fair (or unfair) game in which the gambler wins or loses one unit with each play, then the chance that a gambler will beat a house with very large resources N is very small. However, if $p > 1/2$,

$$\lim_{N \rightarrow \infty} \alpha(j) = 1 - \left(\frac{1-p}{p} \right)^j > 0.$$

This says that there is a positive chance that the gambler playing a game in the gambler's favor will never lose all the resources and will be able to play forever.

Suppose $p = 1/2$, and let T be the time it takes for the random walk to reach 0 or N , and let

$$G(j) = G(j, N) = \mathbb{E}[T \mid X_0 = j].$$

Clearly, $G(0) = 0, G(N) = 0$ and by considering one step we can see that

$$G(j) = 1 + \frac{1}{2} G(j-1) + \frac{1}{2} G(j+1), \quad j = 1, \dots, n-1. \quad (1.17)$$

This is an example of an inhomogeneous linear difference equation. One solution of the equation is given by $G_0(j) = j^2$. Also, if G_1, G_2 are two solutions to (1.17), we can see that $g = G_1 - G_2$ satisfies the homogeneous equation

$$g(j) = \frac{1}{2} g(j-1) + \frac{1}{2} g(j+1), \quad j = 1, \dots, n-1.$$

Using this, we can see that all solutions of (1.17) are of the form

$$G(j) = j^2 + c_1 + c_2 j.$$

Plugging in the boundary conditions $G(0) = G(N) = 0$, allows us to determine the constants c_1, c_2 , and we get

$$\mathbb{E}[T \mid X_0 = j] = j(N-j). \quad (1.18)$$

1.6 Examples

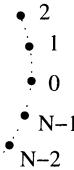
Simple Random Walk on a Graph (Example 5, Section 1.1). Assume the graph is connected so that the walk is irreducible. Let e denote the total

number of edges in the graph and $d(v)$ the number of edges that have v as one of their endpoints. Since each edge has two endpoints, the sum of $d(v)$ over the vertices in the graph is $2e$. It is easy to check that

$$\pi(v) = d(v)/2e,$$

is the invariant probability measure for this chain.

Simple Random Walk on a Circle. Let $N \geq 2$ be an integer. We can consider $\{0, 1, \dots, N-1\}$ to be a “circle” by assuming that $N-1$ is adjacent to 0 as well as $N-2$.



Let X_n be simple random walk on the circle. The transition probabilities are

$$p(k, k-1) = p(k-1, k) = \frac{1}{2}, \quad k = 1, \dots, N-1,$$

$$p(0, N-1) = p(N-1, 0) = \frac{1}{2}.$$

The invariant probability is the uniform distribution. Assume that $X_0 = 0$ and let T_k denote the first time at which the number of distinct points visited equals k . Then T_N is the first time that every point has been visited. By definition $T_1 = 0$, and clearly $T_2 = 1$. We will compute $r(k) = \mathbb{E}[T_k - T_{k-1}]$ for $k = 3, \dots, N$; a little thought will show that the value depends only on k and not on N . Note that at time T_{k-1} the chain is at a boundary point so that one of the neighbors of $X_{T_{k-1}}$ has been visited and the other has not. In the next step we will either visit the new point or we will go to an interior point. If we go to the interior point, the random walk has to continue until it reaches a boundary point and then we start afresh. By (1.18), the expected time that it takes the random walk from the interior point (next to the boundary point) to reach a boundary point is $k-3$. We therefore get the equation

$$r(k) = 1 + \frac{1}{2} [(k-3) + r(k)],$$

or $r(k) = k - 1$. Therefore,

$$\mathbb{E}[T_N] = 1 + \sum_{k=3}^N \mathbb{E}[T_k - T_{k-1}] = 1 + \sum_{k=3}^N (k - 1) = \frac{N(N - 1)}{2}.$$

We can also ask for the distribution of X_{T_N} , the last point to be visited by the chain. It turns out that the distribution of this random variable is uniform on $\{1, \dots, N - 1\}$. We leave the derivation of this fact to the exercises (Exercise 1.16).

Urn Model. Suppose there is an urn with N balls. Each ball is colored either red or green. In each time period, one ball is chosen at random from the urn and with probability $1/2$ is replaced with a ball of the other color; otherwise, the ball is returned to the urn. Let X_n denote the number of red balls after n picks. Then X_n is an irreducible Markov chain with state space $\{0, \dots, N\}$. The transition matrix is given by

$$p(j, j+1) = \frac{N-j}{2N}, \quad p(j, j-1) = \frac{j}{2N}, \quad p(j, j) = \frac{1}{2}, \quad j = 0, 1, \dots, N.$$

One might guess that this chain would tend to keep the number of red balls and green balls about the same. In fact, the invariant probability is given by the binomial distribution

$$\pi(j) = \binom{N}{j} 2^{-N}.$$

It is straightforward to show that this is an invariant probability,

$$\begin{aligned} (\bar{\pi}\mathbf{P})(j) &= \sum_{k=0}^N \pi(k)p(k, j) \\ &= \pi(j-1)p(j-1, j) + \pi(j)p(j, j) + \pi(j+1)p(j+1, j) \\ &= 2^{-N} \binom{N}{j-1} \frac{N-(j-1)}{2N} + 2^{-N} \binom{N}{j} \frac{1}{2} + 2^{-N} \binom{N}{j+1} \frac{j+1}{2N} \\ &= 2^{-N} \binom{N}{j} = \pi(j). \end{aligned}$$

Hence the probability distribution in equilibrium for the number of red balls is the same as the distribution for the number of heads in N flips of a coin. Recall by the central limit theorem, the number of heads is $N/2$ with a random fluctuation which is of order \sqrt{N} . We could have guessed the invariant distribution by considering the problem slightly differently: suppose we always keep the same N balls, but when a ball is chosen we paint it the other color with probability $1/2$. Then in the long run, we would expect the colors of the N balls to become independent with each ball having probability $1/2$ of being red.

Cell Genetics. Consider the following Markov chain which models reproduction of cells. Suppose each cell contains N particles each of either one of two types, I or II. Let j be the number of particles of type I. In reproduction, we assume that the cell duplicates itself and then splits, randomly distributing the particles. After duplication, the cell has $2j$ particles of type I and $2(N - j)$ particles of type II. It then selects N of these $2N$ particles for the next cell. By using the hypergeometric distribution we see that this gives rise to transition probabilities

$$p(j, k) = \frac{\binom{2j}{k} \binom{2(N-j)}{N-k}}{\binom{2N}{N}}.$$

This Markov chain has two absorbing states, 0 and N . Eventually all cells will have only particles of type I or of type II.

Suppose we start with a large number of cells each with j particles of type I. After a long time the population will be full of cells all with one type of particle. What fraction of these will be all type I? Since the fraction of type I particles does not change in this procedure we would expect that the fraction would be j/N . In other words, if we let $\alpha(j)$ be the probability that the Markov chain starting in state j is eventually absorbed in state N , then we expect that

$$\alpha(j) = \frac{j}{N}.$$

For $1 \leq j \leq N - 1$ we can, in fact, verify that this choice of $\alpha(j)$ satisfies

$$\alpha(j) = \sum_{k=0}^N p(j, k) \alpha(k),$$

and hence gives the absorption probabilities.

Card Shuffling. Consider a deck of cards numbered $1, \dots, n$. At each time we will shuffle the cards by drawing a card at random and placing it at the top of the deck. This can be thought of as a Markov chain on S_n , the set of permutations of n elements. If λ denotes any permutation (one-to-one correspondence of $\{1, \dots, n\}$ with itself), and ν_j denotes the permutation corresponding to moving the j th card to the top of the deck, then the transition probabilities for this chain are given by

$$p(\lambda, \nu_j \lambda) = \frac{1}{n}, \quad j = 1, \dots, n.$$

This chain is irreducible and aperiodic. It is easy to verify that the unique invariant probability is the uniform measure on S_n , the measure that assigns

probability $1/n!$ to each permutation. Therefore, if we start with any ordering of the cards, after enough moves of this kind the deck will be well shuffled.

A much harder question which we will not discuss in this book is how many such moves are “enough” so the deck of cards is shuffled. Other questions, such as the expected number of moves from a given permutation to another given permutation, theoretically can be answered by the methods described in this chapter yet cannot be answered from a practical perspective. The reason is that the transition matrix is $n! \times n!$ which (except for small n) is too large to do the necessary matrix operations.

1.7 Exercises

1.1 The Smiths receive the paper every morning and place it on a pile after reading it. Each afternoon, with probability $1/3$, someone takes all the papers in the pile and puts them in the recycling bin. Also, if ever there are at least five papers in the pile, Mr. Smith (with probability 1) takes the papers to the bin. Consider the number of papers in the pile in the evening. Is it reasonable to model this by a Markov chain? If so, what are the state space and transition matrix?

1.2 Consider a Markov chain with state space $\{0, 1\}$ and transition matrix

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

Assuming that the chain starts in state 0 at time $n = 0$, what is the probability that it is in state 1 at time $n = 3$?

1.3 Consider a Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$\mathbf{P} = \begin{smallmatrix} & 1 & 2 & 3 \\ 1 & .4 & .2 & .4 \\ 2 & .6 & 0 & .4 \\ 3 & .2 & .5 & .3 \end{smallmatrix}.$$

What is the probability in the long run that the chain is in state 1? Solve this problem two different ways: 1) by raising the matrix to a high power; and 2) by directly computing the invariant probability vector as a left eigenvector.

1.4 Do the same for the transition matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} .2 & .4 & .4 \\ .1 & .5 & .4 \\ .6 & .3 & .1 \end{bmatrix} \end{matrix}.$$

1.5 Consider the Markov chain with state space $S = \{0, \dots, 5\}$ and transition matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} .5 & .5 & 0 & 0 & 0 & 0 \\ .3 & .7 & 0 & 0 & 0 & 0 \\ 0 & 0 & .1 & 0 & .9 & 0 \\ .25 & .25 & 0 & 0 & .25 & .25 \\ 0 & 0 & .7 & 0 & .3 & 0 \\ 0 & .2 & 0 & .2 & .2 & .4 \end{bmatrix} \end{matrix}.$$

What are the communication classes? Which ones are recurrent and which are transient? Suppose the system starts in state 0. What is the probability that it will be in state 0 at some large time? Answer the same question assuming the system starts in state 5.

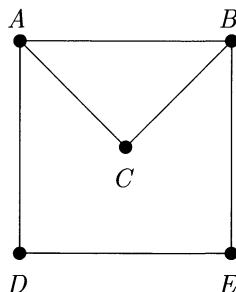
1.6 Assume that the chain in Exercise 1.3 starts in state 2. What is the expected number of time intervals until the chain is in state 2 again?

1.7 Let X_n be an irreducible Markov chain on the state space $\{1, \dots, N\}$. Show that there exist $C < \infty$ and $\rho < 1$ such that for any states i, j ,

$$\mathbb{P}\{X_m \neq j, m = 0, \dots, n \mid X_0 = i\} \leq C\rho^n.$$

Show that this implies that $\mathbb{E}(T) < \infty$, where T is the first time that the Markov chain reaches the state j . (Hint: there exists a $\delta > 0$ such that for all i , the probability of reaching j some time in the first N steps, starting at i , is greater than δ . Why?)

1.8 Consider simple random walk on the graph below. (Recall that simple random walk on a graph is the Markov chain which at each time moves to an adjacent vertex, each adjacent vertex having the same probability.)



- (a) In the long run, about what fraction of time is spent in vertex A ?
- (b) Suppose a walker starts in vertex A . What is the expected number of steps until the walker returns to A ?
- (c) Suppose a walker starts in vertex C . What is the expected number of visits to B before the walker reaches A ?
- (d) Suppose the walker starts in vertex B . What is the probability that the walker reaches A before the walker reaches C ?
- (e) Again assume the walker starts in C . What is the expected number of steps until the walker reaches A ?

1.9 Consider the Markov chain with state space $\{1, 2, 3, 4, 5\}$ and matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \left[\begin{matrix} 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{matrix} \right] \end{matrix}.$$

- (a) Is the chain irreducible?
- (b) What is the period of the chain?
- (c) What are $p_{1,000}(2, 1), p_{1,000}(2, 2), p_{1,000}(2, 4)$ (approximately)?
- (d) Let T be the first return time to the state 1, starting at state 1. What is the distribution of T and what is $\mathbb{E}(T)$? What does this say, without any further calculation, about $\pi(1)$?
- (e) Find the invariant probability $\bar{\pi}$. Use this to find the expected return time to state 2, starting in state 2.

1.10 Suppose X_n is a Markov chain with state space $\{0, 1, \dots, 6\}$ and transition probabilities

$$p(0, 0) = \frac{3}{4}, \quad p(0, 1) = \frac{1}{4},$$

$$p(1, 0) = \frac{1}{2}, \quad p(1, 1) = \frac{1}{4}, \quad p(1, 2) = \frac{1}{4},$$

$$p(6, 0) = \frac{1}{4}, \quad p(6, 5) = \frac{1}{4}, \quad p(6, 6) = \frac{1}{2},$$

and for $j = 2, 3, 4, 5$,

$$p(j, 0) = p(j, j - 1) = p(j, j) = p(j, j + 1) = \frac{1}{4}.$$

- (a) Is this chain irreducible? Is it aperiodic?

- (b) Suppose the chain has been running for a long time and we start watching the chain. What is the probability that the next three states will be 4, 5, 0 in that order?
- (c) Suppose the chain starts in state 1. What is the probability that it reaches state 6 before reaching state 0?
- (d) Suppose the chain starts in state 3. What is the expected number of steps until the chain is in state 3 again?
- (e) Suppose the chain starts in state 0. What is the expected number of steps until the chain is in state 6?

1.11 Let X_1, X_2, \dots be the successive values from independent rolls of a standard six-sided die. Let $S_n = X_1 + \dots + X_n$. Let

$$T_1 = \min\{n \geq 1 : S_n \text{ is divisible by } 8\},$$

$$T_2 = \min\{n \geq 1 : S_n - 1 \text{ is divisible by } 8\}.$$

Find $\mathbb{E}(T_1)$ and $\mathbb{E}(T_2)$. (Hint: consider the remainder of S_n after division by 8 as a Markov chain.)

1.12 Let X_n, Y_n be independent Markov chains with state space $\{0, 1, 2\}$ and transition matrix

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

Suppose $X_0 = 0, Y_0 = 2$ and let

$$T = \inf\{n : X_n = Y_n\}.$$

- (a) Find $\mathbb{E}(T)$.
- (b) What is $\mathbb{P}\{X_T = 2\}$?
- (c) In the long run, what percentage of the time are both chains in the same state?

[Hint: consider the nine-state Markov chain $Z_n = (X_n, Y_n)$.]

1.13 Consider the Markov chain described in Exercise 1.1.

- (a) After a long time, what would be the expected number of papers in the pile?

- (b) Assume the pile starts with 0 papers. What is the expected time until the pile will again have 0 papers?

1.14 Let X_n be a Markov chain on state space $\{1, 2, 3, 4, 5\}$ with transition matrix

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

- (a) Is this chain irreducible? Is it aperiodic?
- (b) Find the stationary probability vector.
- (c) Suppose the chain starts in state 1. What is the expected number of steps until it is in state 1 again?
- (d) Again, suppose $X_0 = 1$. What is the expected number of steps until the chain is in state 4?
- (e) Again, suppose $X_0 = 1$. What is the probability that the chain will enter state 5 before it enters state 3?

1.15 Let X_n be an irreducible Markov chain with state space S starting at state i with transition matrix \mathbf{P} . Let

$$T = \min\{n > 0 : X_n = i\}$$

be the first time that the chain returns to state i . For each state j let $r(j)$ be the expected number of visits to j before returning to i ,

$$r(j) = \mathbb{E} \left[\sum_{n=0}^{T-1} I\{X_n = j\} \right].$$

Note that $r(i) = 1$.

- (a) Let \bar{r} be the vector whose j th component is $r(j)$. Show that $\bar{r}\mathbf{P} = \bar{r}$.
- (b) Show that

$$\mathbb{E}(T) = \sum_{j \in S} r(j).$$

- (c) Conclude that $\mathbb{E}(T) = \pi(i)^{-1}$, where $\bar{\pi}$ denotes the invariant probability.

1.16 Consider simple random walk on the circle $\{0, 1, \dots, N-1\}$ started at 0 as described in Section 1.6. Show that the distribution of X_{T_N} is uniform on $\{1, 2, \dots, N-1\}$.

1.17 The complete graph on $\{1, \dots, N\}$ is the simple graph with these vertices such that any pair of distinct points is adjacent. Let X_n denote simple

random walk on this graph and let T be the first time that the walk reaches the state 1.

- (a) Give the distribution of T assuming $X_0 = 1$. Verify (1.11) for this chain.
- (b) What is $\mathbb{E}[T \mid X_0 = 2]$?
- (c) Find the expected number of steps needed until every point has been visited at least once.

1.18 Suppose we take a standard deck of cards with 52 cards and do the card shuffling procedure as in Section 1.6. Suppose we do one move every second. What is the expected amount of time in years until the deck returns to the original order?

1.19 Suppose we flip a fair coin repeatedly until we have flipped four consecutive heads. What is the expected number of flips that are needed? (Hint: consider a Markov chain with state space $\{0, 1, \dots, 4\}$.)

1.20 In this exercise we outline a proof of the Perron–Frobenius Theorem about matrices with positive entries. Let $\mathbf{A} = (a_{ij})$ be an $N \times N$ matrix with $a_{ij} > 0$ for all i, j . For vectors $\bar{u} = (u^1, \dots, u^N)$ and $\bar{v} = (v^1, \dots, v^N)$ we write $\bar{u} \geq \bar{v}$ if $u^i \geq v^i$ for each i and $\bar{u} > \bar{v}$ if $u^i > v^i$ for each i . We write $\bar{0} = (0, \dots, 0)$.

(a) Show that if $\bar{v} \geq \bar{0}$ and $\bar{v} \neq \bar{0}$, then $\mathbf{A}\bar{v} > \bar{0}$.

For any vector $\bar{v} \geq \bar{0}$, let $g(\bar{v})$ be the largest λ such that

$$\mathbf{A}\bar{v} \geq \lambda\bar{v}.$$

(b) Show that $g(\bar{v}) > 0$ for any nonzero $\bar{v} \geq \bar{0}$ and if $c > 0$ then $g(c\bar{v}) = g(\bar{v})$. Let

$$\alpha = \sup g(\bar{v}),$$

where the supremum is over all nonzero $\bar{v} \geq \bar{0}$. By (b) we can consider the supremum over all v with

$$\|v\| = \sqrt{(v^1)^2 + \dots + (v^N)^2} = 1.$$

By continuity of the function g on $\{\|v\| = 1\}$ it can be shown that there exists at least one vector $\bar{v} \geq \bar{0}$ with $g(\bar{v}) = \alpha$.

(c) Show that for any \bar{v} with $g(\bar{v}) = \alpha$,

$$\mathbf{A}\bar{v} = \alpha\bar{v},$$

i.e., \bar{v} is an eigenvector with eigenvalue α . [Hint: we know by definition that $\mathbf{A}\bar{v} \geq \alpha\bar{v}$. Assume that they are not equal and consider

$$\mathbf{A}[\mathbf{A}\bar{v} - \alpha\bar{v}],$$

using (a).]

- (d) Show that there is a unique $\bar{v} \geq \bar{0}$ with $g(\bar{v}) = \alpha$ and $\sum_{i=1}^N v^i = 1$.
 [Hint: assume there were two such vectors, \bar{v}_1, \bar{v}_2 , and consider $g(\bar{v}_1 - \bar{v}_2)$ and $g(|\bar{v}_1 - \bar{v}_2|)$ where

$$|\bar{v}| = (|v^1|, \dots, |v^n|).]$$

- (e) Show that all the components of the \bar{v} in (c) are strictly positive. [Hint: if $\mathbf{A}\bar{v} \geq \lambda\bar{v}$ then $\mathbf{A}(\mathbf{A}\bar{v}) \geq \lambda\mathbf{A}\bar{v}$.]

- (f) Show that if λ is any other eigenvalue of \mathbf{A} , then $|\lambda| < \alpha$. (Hint: assume $\mathbf{A}\bar{u} = \lambda\bar{u}$ and consider $\mathbf{A}|\bar{u}|$.)

- (g) Show that if \mathbf{B} is any $(N-1) \times (N-1)$ submatrix of A , then all the eigenvalues of \mathbf{B} have absolute value strictly less than α . [Hint: since \mathbf{B} is a matrix with positive entries, (a)–(f) apply to \mathbf{B} .]

- (h) Consider

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}).$$

Show that

$$f'(\lambda) = - \sum_{i=1}^N \det(\mathbf{B}_i - \lambda\mathbf{I}),$$

where \mathbf{B}_i denotes the submatrix of \mathbf{A} obtained by deleting the i th row and i th column.

- (i) Use (g) and (h) to conclude that

$$f'(\alpha) > 0,$$

and hence that α is a simple eigenvalue for \mathbf{A} .

- (j) Explain why every stochastic matrix with strictly positive entries has a unique invariant probability with all positive components. (Apply the above results to the transpose of the stochastic matrix.)

- 1.21** An elementary theorem in number theory states that if two integers m and n are relatively prime (i.e., greatest common divisor equal to 1), then there exist integers x and y (positive or negative) such that

$$mx + ny = 1.$$

Using this theorem show the following:

- (a) If m and n are relatively prime then the set

$$\{xm + ny : x, y \text{ positive integers}\}$$

contains all but a finite number of the positive integers.

- (b) Let J be a set of nonnegative integers whose greatest common divisor is d . Suppose also that J is closed under addition, $m, n \in J \Rightarrow m + n \in J$. Then J contains all but a finite number of integers in the set $\{0, d, 2d, \dots\}$.

Chapter 2

Countable Markov Chains

2.1 Introduction

In this chapter, we consider (time-homogeneous) Markov chains with a countably infinite state space. A set is countably infinite if it can be put into one-to-one correspondence with the set of nonnegative integers $\{0, 1, 2, \dots\}$. Examples of such sets are: \mathbb{Z} , the set of all integers; $2\mathbb{Z}$, the set of even integers; and \mathbb{Z}^2 , the set of lattice points in the plane,

$$\mathbb{Z}^2 = \{(i, j) : i, j \text{ integers}\}.$$

(The reader may wish to consider how \mathbb{Z}^2 and $\{0, 1, 2, \dots\}$ can be put into one-to-one correspondence.) Not all infinite sets are countably infinite; for example, the set of real numbers cannot be put into one-to-one correspondence with the positive integers.

We will again let X_n denote a Markov chain. Some of that which was described for finite-state Markov chains holds equally well in the infinite case; however, some things become a bit trickier. We again can speak of the transition matrix, but in this case it becomes an infinite matrix. We will choose not to use the matrix notation here, but simply write the transition probabilities as

$$p(x, y) = \mathbb{P}\{X_1 = y \mid X_0 = x\}, \quad x, y \in S.$$

The transition probabilities are nonnegative and the “rows” add up to 1, i.e., for each $x \in S$,

$$\sum_{y \in S} p(x, y) = 1.$$

We have chosen to use x, y, z for elements of the state space S . We also define the n -step transition probabilities

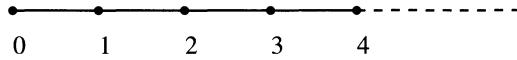
$$p_n(x, y) = \mathbb{P}\{X_n = y \mid X_0 = x\}.$$

If $0 < m, n < \infty$,

$$\begin{aligned} p_{m+n}(x, y) &= \mathbb{P}\{X_{m+n} = y \mid X_0 = x\} \\ &= \sum_{z \in S} \mathbb{P}\{X_{m+n} = y, X_m = z \mid X_0 = x\} \\ &= \sum_{z \in S} p_m(x, z) p_n(z, y). \end{aligned}$$

This equation is sometimes called the Chapman–Kolmogorov equation. It can be considered the definition of matrix multiplication for infinite matrices.

Example 1. Random Walk with Partially Reflecting Boundary at 0. Let $0 < p < 1$ and $S = \{0, 1, 2, \dots\}$.



The transition probabilities are given by

$$p(x, x-1) = 1-p, \quad p(x, x+1) = p, \quad x > 0,$$

and

$$p(0, 0) = 1-p, \quad p(0, 1) = p.$$

Example 2. Simple Random Walk on the Integer Lattice. Let \mathbb{Z}^d be the d -dimensional integer lattice, i.e.,

$$\mathbb{Z}^d = \{(z_1, \dots, z_d) : z_i \in \mathbb{Z}\}.$$

Note that each element x of \mathbb{Z}^d has $2d$ “nearest neighbors” in \mathbb{Z}^d which are distance 1 from x . Simple random walk on \mathbb{Z}^d is the process X_n taking values in \mathbb{Z}^d which at each time moves to one of the $2d$ nearest neighbors of its current position, choosing equally among all the nearest neighbors. More precisely, it is the Markov chain with state space $S = \mathbb{Z}^d$ and

$$p(x, y) = \begin{cases} 1/2d, & \text{if } |x - y| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example 3. Queueing Model. Let X_n be the number of customers waiting in line for some service. We think of the first person in line as being serviced while all others are waiting their turn. During each time interval there is a probability p that a new customer arrives. With probability q , the service

for the first customer is completed and that customer leaves the queue. We put no limit on the number of customers waiting in line. This is a Markov chain with state space $\{0, 1, 2, \dots\}$ and transition probabilities (see Example 2, Section 1.1):

$$p(x, x - 1) = q(1 - p), \quad p(x, x) = qp + (1 - q)(1 - p),$$

$$p(x, x + 1) = p(1 - q), \quad x > 0;$$

$$p(0, 0) = 1 - p, \quad p(0, 1) = p.$$

As in the case of finite Markov chains, our goal will be to understand the behavior for large time. Some of the ideas for finite chains apply equally well to the infinite case. For example, the notion of communication classes applies equally well here. Again, we call a Markov chain irreducible if all the states communicate. All the examples discussed in this chapter are irreducible except for a couple of cases where all the states but one communicate and that one state x is absorbing, $p(x, x) = 1$. We can also talk of the period of an irreducible chain; Examples 1 and 3 above are aperiodic, whereas Example 2 has period 2. It will not always be the case that an irreducible, aperiodic Markov chain with infinite state space converges to an equilibrium probability distribution.

2.2 Recurrence and Transience

Suppose X_n is an irreducible Markov chain with countably infinite state space S and transition probabilities $p(x, y)$. We say that X_n is a *recurrent* chain if for each state x ,

$$\mathbb{P}\{X_n = x \text{ for infinitely many } n\} = 1,$$

i.e., if the chain returns infinitely often to x . If an irreducible chain visits a certain state x infinitely often then it must visit every state infinitely often. (The basic reason is that if y is another state there is a positive probability of reaching y from x . If x is visited infinitely often then we get this chance of reaching y infinitely often. If a certain event has a positive probability of occurring, and we get an infinite number of trials, then the event will occur an infinite number of times.) If the chain is not recurrent, then every state is visited only a finite number of times. In this case, the chain is called *transient*. It is not always easy to determine whether a given Markov chain is recurrent or transient. In this section we give two criteria for determining this.

Fix a site x and assume that $X_0 = x$. Consider the random variable R which gives the total number of visits to the site x , including the initial visit. We can write R as

$$R = \sum_{n=0}^{\infty} I\{X_n = x\},$$

where again we use I to denote the indicator function, which equals 1 if the event occurs and 0 otherwise. If the chain is recurrent then R is identically equal to infinity; if the chain is transient, then $R < \infty$ with probability 1. We can compute the expectation of R (assuming $X_0 = x$),

$$\mathbb{E}(R) = \mathbb{E} \sum_{n=0}^{\infty} I\{X_n = x\} = \sum_{n=0}^{\infty} \mathbb{P}\{X_n = x\} = \sum_{n=0}^{\infty} p_n(x, x).$$

We will now compute $\mathbb{E}(R)$ in a different way. Let T be the time of first return to x ,

$$T = \min\{n > 0 : X_n = x\}.$$

We say that $T = \infty$ if the chain never returns to x . Suppose $\mathbb{P}\{T < \infty\} = 1$. Then with probability one, the chain always returns and by continuing we see that the probability that the chain returns infinitely often is 1 and the chain is recurrent. Now suppose $\mathbb{P}\{T < \infty\} = q < 1$, and let us compute the distribution of R in terms of q . First, $R = 1$ if and only if the chain never returns; hence, $\mathbb{P}\{R = 1\} = 1 - q$. If $m > 1$, then $R = m$ if and only if the chain returns $m - 1$ times and then does not return for the m th time. Hence, $\mathbb{P}\{R = m\} = q^{m-1}(1 - q)$. Therefore, in the transient case, $q < 1$,

$$\mathbb{E}(R) = \sum_{m=1}^{\infty} m \mathbb{P}\{R = m\} = \sum_{m=1}^{\infty} m q^{m-1} (1 - q) = \frac{1}{1 - q} < \infty.$$

We have concluded the following:

Fact. *An irreducible Markov chain is transient if and only if the expected number of returns to a state is finite, i.e., if and only if*

$$\sum_{n=0}^{\infty} p_n(x, x) < \infty.$$

Example. Simple Random Walk in \mathbb{Z}^d . We first take $d = 1$, and consider the Markov chain on the integers with transition probabilities

$$p(x, x+1) = p(x, x-1) = \frac{1}{2}.$$

We will concentrate on the state $x = 0$ and assume $X_0 = 0$. Since this chain has period 2, $p_n(0, 0) = 0$ for n odd. We will write down an exact expression for $p_{2n}(0, 0)$. Suppose the walker is to be at 0 after $2n$ steps. Then the walker must take exactly n steps to the right and n steps to the left. Any “path” of length $2n$ that takes exactly n steps to the right and n steps to the left is equally likely.

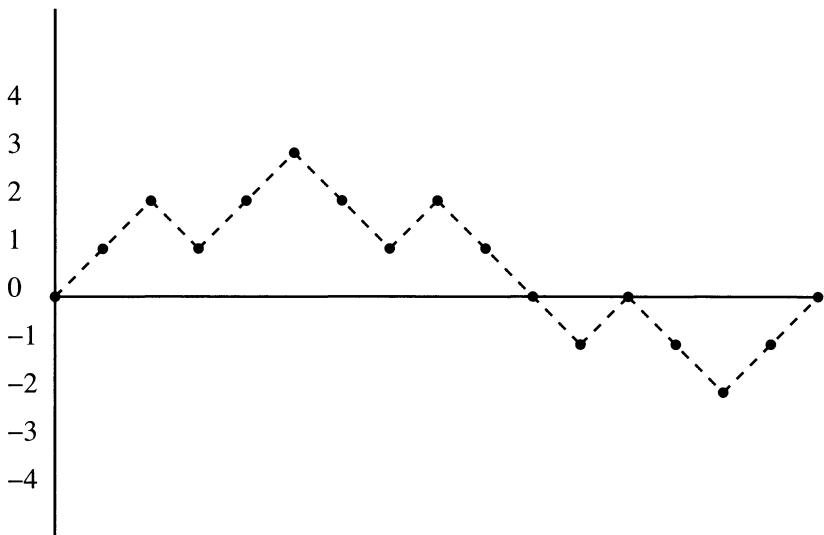


FIGURE 2.1: The graph of a random walk path that is at the origin after 16 steps.

Each such path has probability $(1/2)^{2n}$ of occurring since it combines $2n$ events each with probability $1/2$. There are $\binom{2n}{n}$ ways of choosing which n of the $2n$ steps should be to the right, and then the other n are to the left. Therefore,

$$p_{2n}(0, 0) = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!n!} \left(\frac{1}{2}\right)^{2n}.$$

It is not so easy to see what this looks like for large values of n . However, we can use Stirling’s formula to estimate the factorials. Stirling’s formula (see Exercise 2.18) states that

$$n! \sim \sqrt{2\pi n} n^n e^{-n},$$

where \sim means that the ratio of the two sides approaches 1 as n goes to

infinity. If we plug this into the above expressions we get that

$$p_{2n}(0, 0) \sim \frac{1}{\sqrt{\pi n}}. \quad (2.1)$$

In particular, since $\sum n^{-1/2} = \infty$,

$$\sum_{n=0}^{\infty} p_{2n}(0, 0) = \infty,$$

and simple random walk in one dimension is recurrent.

We now take $d > 1$ so that the chain is on the d -dimensional integer lattice \mathbb{Z}^d and has transition probabilities

$$p(x, y) = 1/2d, \quad |x - y| = 1.$$

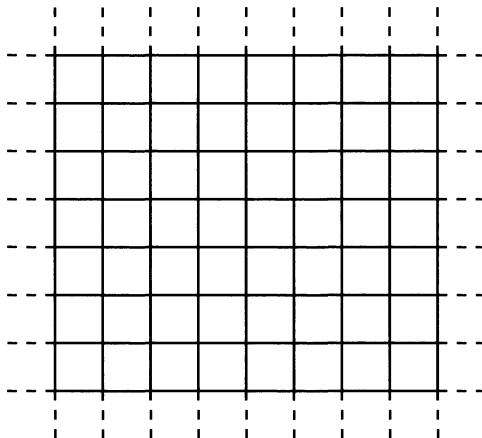


FIGURE 2.2: The lattice \mathbb{Z}^2 .

Again we start the walk at $0 = (0, \dots, 0)$. We will try to get an asymptotic expression for $p_{2n}(0, 0)$ [again $p_n(0, 0) = 0$ for n odd]. The combinatorics are somewhat more complicated in this case, so we will give only a sketch of the derivation. Suppose a walker takes $2n$ steps. Then by the law of large numbers, for large values of n , we expect that $2n/d$ of these steps will be taken in each of the d components. We will need the number of steps in each component to be even if we have any chance of being at 0 in n steps. For large n the probability of this occurring is about $(1/2)^{d-1}$ (whether or not an even number of steps have been taken in each of the first $d-1$ components are almost independent events; however, we know that if an even number of steps

have been taken in the first $d - 1$ components then an even number of steps have been taken in the last component as well since the total number of steps taken is even). In each component, if about $2n/d$ steps have been taken, then by (2.1) we would expect that the probability that that component equals 0 is about $(\pi(n/d))^{-1/2}$. Combining this, we get an asymptotic expression

$$p_{2n}(0,0) \sim \left(\frac{1}{2}\right)^{d-1} \left(\frac{d}{n\pi}\right)^{d/2}.$$

Recall that $\sum n^{-a} < \infty$ if and only if $a > 1$. Hence,

$$\sum_{n=0}^{\infty} p_{2n}(0,0) \begin{cases} = \infty, & d = 1, 2, \\ < \infty, & d \geq 3. \end{cases}$$

We have derived the following.

Fact. *Simple random walk in \mathbb{Z}^d is recurrent if $d = 1$ or 2 and is transient if $d \geq 3$.*

We now consider another method for determining recurrence or transience. Suppose X_n is an irreducible Markov chain and consider a fixed state which we will denote z . For each state x , we set

$$\alpha(x) = \mathbb{P}\{X_n = z \text{ for some } n \geq 0 \mid X_0 = x\}.$$

Clearly, $\alpha(z) = 1$. If the chain is recurrent, then $\alpha(x) = 1$ for all x . However, if the chain is transient there must be states x with $\alpha(x) < 1$. In fact, although not quite as obviously, if the chain is transient there must be points “farther and farther” away from z with $\alpha(x)$ as small as we like.

If $x \neq z$, then

$$\begin{aligned} \alpha(x) &= \mathbb{P}\{X_n = z \text{ for some } n \geq 0 \mid X_0 = x\} \\ &= \mathbb{P}\{X_n = z \text{ for some } n \geq 1 \mid X_0 = x\} \\ &= \sum_{y \in S} \mathbb{P}\{X_1 = y \mid X_0 = x\} \mathbb{P}\{X_n = z \text{ for some } n \geq 1 \mid X_1 = y\} \\ &= \sum_{y \in S} p(x, y) \alpha(y). \end{aligned}$$

Summarizing, $\alpha(x)$ satisfies the following:

$$0 \leq \alpha(x) \leq 1, \tag{2.2}$$

$$\alpha(z) = 1, \quad \inf\{\alpha(x) : x \in S\} = 0, \tag{2.3}$$

and

$$\alpha(x) = \sum_{y \in S} p(x, y) \alpha(y), \quad x \neq z. \tag{2.4}$$

It turns out that if X_n is transient, then there is a unique solution to (2.2) – (2.4) that must correspond to the appropriate probability. Moreover, it can be shown (we prove this in Chapter 5, Section 5.5, Example 5) that if X_n is recurrent there is no solution to (2.2) – (2.4). This then gives another method to determine recurrence or transience:

Fact. *An irreducible Markov chain is transient if and only if for any z we can find a function $\alpha(x)$ satisfying (2.2) – (2.4).*

Example. Consider Example 1 in the previous section, random walk with partially reflecting boundary. Let $z = 0$ and let us try to find a solution to (2.2) – (2.4). The third equation states that

$$\alpha(x) = (1 - p)\alpha(x - 1) + p\alpha(x + 1), \quad x > 0.$$

From (0.5) and (0.6) we see that the only solutions to the above equation are of the form

$$\alpha(x) = c_1 + c_2 \left(\frac{1-p}{p} \right)^x, \quad p \neq 1/2,$$

$$\alpha(x) = c_1 + c_2 x, \quad p = 1/2.$$

The first condition in (2.3) gives $\alpha(0) = 1$; plugging this in gives

$$\alpha(x) = (1 - c_2) + c_2 \left(\frac{1-p}{p} \right)^x, \quad p \neq 1/2 \tag{2.5}$$

$$\alpha(x) = 1 + c_2 x, \quad p = 1/2. \tag{2.6}$$

If we choose $c_2 = 0$, we get $\alpha(x) = 1$ for all x which clearly does not satisfy (2.3). If $p = 1/2$ and $c_2 \neq 0$, then the solution is not bounded and hence cannot satisfy (2.2). Similarly, if $p < 1/2$, the solution to (2.5) will be unbounded for $c_2 \neq 0$. In this case, we can conclude that the chain is recurrent for $p \leq 1/2$. For $p > 1/2$, we can find a solution. The second condition in (2.3) essentially boils down to $\alpha(x) \rightarrow 0$ as $x \rightarrow \infty$, and we get

$$\alpha(x) = \left(\frac{1-p}{p} \right)^x.$$

Therefore, for $p > 1/2$, the chain is transient.

2.3 Positive Recurrence and Null Recurrence

Suppose X_n is an irreducible, aperiodic Markov chain on the infinite state space S . In this section we investigate when a limiting probability distribution

exists. A limiting probability $\pi(x), x \in S$ is a probability distribution on S such that for each $x, y \in S$,

$$\lim_{n \rightarrow \infty} p_n(y, x) = \pi(x).$$

If X_n is transient, then

$$\lim_{n \rightarrow \infty} p_n(y, x) = 0, \quad (2.7)$$

for all x, y , so no limiting probability distribution exists. It is possible, however, for (2.7) to hold for a recurrent chain. Consider, for example, simple random walk on \mathbb{Z} described in the last section (this is actually a periodic chain, but a small modification can be made to give an aperiodic example). It is recurrent but $p_{2n}(0, 0) \rightarrow 0$ as $n \rightarrow \infty$. We call a chain *null recurrent* if it is recurrent but

$$\lim_{n \rightarrow \infty} p_n(x, y) = 0.$$

Otherwise, a recurrent chain is called *positive recurrent*.

Positive recurrent chains behave very similarly to finite Markov chains. If X_n is an irreducible, aperiodic, positive recurrent Markov chain, then for every x, y , the limit

$$\lim_{n \rightarrow \infty} p_n(y, x) = \pi(x) > 0,$$

exists and is independent of the initial state y . The $\pi(x)$ give an invariant probability distribution on S , i.e.,

$$\sum_{y \in S} \pi(y)p(y, x) = \pi(x). \quad (2.8)$$

Moreover, if we consider the return time to a state x ,

$$T = \min\{n > 0 \mid X_n = x\},$$

then for a positive recurrent chain,

$$\mathbb{E}(T \mid X_n = x) = 1/\pi(x).$$

If X_n is null recurrent, then $T < \infty$ with probability 1, but $\mathbb{E}(T) = \infty$. If X_n is transient, then $T = \infty$ with positive probability.

One way to determine whether or not a chain is positive recurrent is to try to find an invariant probability distribution. It can be proved that if an irreducible chain is positive recurrent, then there exists a unique probability distribution satisfying (2.8); moreover, if a chain is not positive recurrent, there is no probability distribution satisfying (2.8). This gives a good criterion:

try to find an invariant probability distribution. If it exists, then the chain is positive recurrent; if none exists, then it is either null recurrent or transient.

Example. Consider again the example of random walk with partially reflecting boundary. We will try to find a probability distribution that satisfies (2.8), i.e., a nonnegative function $\pi(x)$ satisfying (2.8) and

$$\sum_{x \in S} \pi(x) = 1. \quad (2.9)$$

In this example, (2.8) gives

$$\pi(x+1)(1-p) + \pi(x-1)p = \pi(x), \quad x > 0, \quad (2.10)$$

$$\pi(1)(1-p) + \pi(0)(1-p) = \pi(0). \quad (2.11)$$

By (0.5) and (0.6), the general solution to (2.10) is

$$\pi(x) = c_1 + c_2 \left(\frac{p}{1-p} \right)^x, \quad p \neq 1/2,$$

$$\pi(x) = c_1 + c_2 x, \quad p = 1/2.$$

Equation (2.11) gives $\pi(0) = [(1-p)/p]\pi(1)$. Plugging this into the above gives

$$\pi(x) = c_2 \left(\frac{p}{1-p} \right)^x, \quad p \neq 1/2,$$

$$\pi(x) = c_1, \quad p = 1/2.$$

Now we impose the condition (2.9): can we choose the constant c_1 or c_2 so that $\sum \pi(x) = 1$? For $p = 1/2$, it clearly cannot be done. Suppose $p \neq 1/2$. Clearly, we would need $c_2 \neq 0$. If $p > 1/2$, $\sum [p/(1-p)]^x = \infty$ and we cannot find such a c_2 (we already knew the chain was transient in this case, so it could not possibly be positive recurrent). However if $p < 1/2$, the sum is finite and we can choose

$$\pi(x) = \left(\frac{p}{1-p} \right)^x \left[\sum_{y=0}^{\infty} \left(\frac{p}{1-p} \right)^y \right]^{-1} = \left(\frac{1-2p}{1-p} \right) \left(\frac{p}{1-p} \right)^x.$$

In this case the chain is positive recurrent and this gives the invariant probability. Summarizing the discussion in the last two sections we see that random walk with partially reflecting boundary is

$$\begin{aligned} &\text{positive recurrent if } p < 1/2, \\ &\text{null recurrent if } p = 1/2, \\ &\text{transient if } p > 1/2. \end{aligned}$$

2.4 Branching Process

In this section we study a stochastic model for population growth. Consider a population of individuals. We let X_n denote the number of individuals at time n . At each time interval, the population will change according to the following rule: each individual will produce a random number of offspring; after producing the offspring, the individual dies and leaves the system. We make two assumptions about the reproduction process:

1. Each individual produces offspring with the same probability distribution: there are given nonnegative numbers p_0, p_1, p_2, \dots summing to 1 such that the probability that an individual produces exactly k offspring is p_k .
2. The individuals reproduce independently.

The number of individuals at stage n , X_n , is then a Markov chain with state space $\{0, 1, 2, \dots\}$. Note that 0 is an absorbing state; once the population dies out, no individuals can be produced. It is not so easy to write down explicitly the transition probabilities for this chain. Suppose that $X_n = k$. Then k individuals produce offspring for the $(n+1)$ st generation. If Y_1, \dots, Y_k are independent random variables each with distribution $\mathbb{P}\{Y_i = j\} = p_j$, then

$$p(k, j) = \mathbb{P}\{X_{n+1} = j \mid X_n = k\} = \mathbb{P}\{Y_1 + \dots + Y_k = j\}.$$

The actual distribution of $Y_1 + \dots + Y_k$ can be expressed in terms of convolutions, but we will not need the exact form here. Let μ denote the mean number of offspring produced by an individual,

$$\mu = \sum_{i=0}^{\infty} i p_i.$$

Then,

$$\mathbb{E}(X_{n+1} \mid X_n = k) = \mathbb{E}(Y_1 + \dots + Y_k) = k\mu.$$

It is relatively straightforward to calculate the mean number of individuals, $\mathbb{E}(X_n)$,

$$\begin{aligned} \mathbb{E}(X_n) &= \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\} \mathbb{E}(X_n \mid X_{n-1} = k) \\ &= \sum_{k=0}^{\infty} k \mu \mathbb{P}\{X_{n-1} = k\} = \mu \mathbb{E}(X_{n-1}). \end{aligned}$$

Or, if we do this n times,

$$\mathbb{E}(X_n) = \mu^n \mathbb{E}(X_0).$$

Some interesting conclusions can be reached from this expression. If $\mu < 1$, then the mean number of offspring goes to 0 as n gets large. The easy estimate

$$\mathbb{E}(X_n) = \sum_{k=0}^{\infty} k \mathbb{P}\{X_n = k\} \geq \sum_{k=1}^{\infty} \mathbb{P}\{X_n = k\} = \mathbb{P}\{X_n \geq 1\}$$

can then be used to deduce that the population eventually dies out,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{X_n = 0\} = 1.$$

If $\mu = 1$, the expected population size remains constant while for $\mu > 1$, the expected population size grows. It is not so clear in these cases whether or not the population dies out with probability 1. [It is possible for X_n to be 0 with probability very near 1, yet $\mathbb{E}(X_n)$ not be small.] Below we investigate how to determine the probability that the population dies out. In order to avoid trivial cases we will assume that

$$p_0 > 0; \quad p_0 + p_1 < 1. \quad (2.12)$$

Let

$$a_n(k) = \mathbb{P}\{X_n = 0 \mid X_0 = k\}$$

and let $a(k)$ be the probability that the population eventually dies out assuming that there are k individuals initially,

$$a(k) = \lim_{n \rightarrow \infty} a_n(k).$$

If the population has k individuals at a certain time, then the only way for the population to die out is for all k branches to die out. Since the branches act independently,

$$a(k) = [a(1)]^k.$$

It suffices therefore to determine $a(1)$ which we will denote by just a and call the *extinction probability*. Assume now that $X_0 = 1$. If we look at one generation, we get

$$\begin{aligned} a &= \mathbb{P}\{\text{population dies out} \mid X_0 = 1\} \\ &= \sum_{k=0}^{\infty} \mathbb{P}\{X_1 = k \mid X_0 = 1\} \mathbb{P}\{\text{population dies out} \mid X_1 = k\} \\ &= \sum_{k=0}^{\infty} p_k a(k) = \sum_{k=0}^{\infty} p_k a^k. \end{aligned}$$

The quantity on the right is of sufficient interest to give it a name. If X is a random variable taking values in $\{0, 1, 2, \dots\}$, the *generating function* of X is the function

$$\phi(s) = \phi_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \mathbb{P}\{X = k\}.$$

Note that $\phi(s)$ is an increasing function of $s \geq 0$ with $\phi(0) = \mathbb{P}\{X = 0\}$ and $\phi(1) = 1$. Differentiating, we get

$$\phi'(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}\{X = k\},$$

$$\phi''(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} \mathbb{P}\{X = k\}.$$

Hence,

$$\phi'(1) = \sum_{k=1}^{\infty} k \mathbb{P}\{X = k\} = \mathbb{E}(X), \quad (2.13)$$

and for $s > 0$, if $\mathbb{P}\{X \geq 2\} > 0$,

$$\phi''(s) > 0. \quad (2.14)$$

If X_1, \dots, X_m are independent random variables taking values in the nonnegative integers, then

$$\phi_{X_1 + \dots + X_m}(s) = \phi_{X_1}(s) \cdots \phi_{X_m}(s).$$

The easiest way to see this is to use the expression $\phi_X(s) = \mathbb{E}(s^X)$ and the product rule for expectation of independent random variables.

Returning to the branching process we see that the extinction probability a satisfies the equation

$$a = \phi(a).$$

Clearly, $a = 1$ satisfies this equation, but there could well be other solutions. Again, we assume $X_0 = 1$. Then the generating function of the random variable X_0 is a and the generating function of X_1 is $\phi(a)$. Let $\phi^n(a)$ be the generating function of X_n . We will now show that

$$\phi^n(a) = \phi(\phi^{n-1}(a)).$$

To see this, we first note

$$\begin{aligned} \phi^n(a) &= \sum_{k=0}^{\infty} \mathbb{P}\{X_n = k\} a^k \\ &= \sum_{k=0}^{\infty} \left[\sum_{j=0}^{\infty} \mathbb{P}\{X_1 = j\} \mathbb{P}\{X_n = k \mid X_1 = j\} \right] a^k \\ &= \sum_{j=0}^{\infty} p_j \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k \mid X_0 = j\} a^k. \end{aligned}$$

Now, if $X_0 = j$, then X_{n-1} is the sum of j independent random variables each with the distribution of X_{n-1} given $X_0 = 1$. Hence the sum over k is the generating function of the sum of j independent random variables each with generating function $\phi^{n-1}(a)$ and hence

$$\sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k \mid X_0 = j\} a^k = [\phi^{n-1}(a)]^j,$$

and

$$\phi^n(a) = \sum_{j=0}^{\infty} p_j [\phi^{n-1}(a)]^j = \phi(\phi^{n-1}(a)).$$

We now have a recursive way to find $\phi^n(a)$ and hence to find

$$a_n(1) = \mathbb{P}\{X_n = 0 \mid X_0 = 1\} = \phi^n(0).$$

We are now ready to demonstrate the following: the extinction probability a is the *smallest* positive root of the equation $a = \phi(a)$. We have already seen that a must satisfy this equation. Let \hat{a} be the smallest positive root. We will show by induction that for every n , $a_n = \mathbb{P}\{X_n = 0\} \leq \hat{a}$ (which implies that $a = \lim a_n \leq \hat{a}$). This is obviously true for $n = 0$ since $a_0 = 0$. Assume that $a_{n-1} \leq \hat{a}$. Then

$$\mathbb{P}\{X_n = 0\} = \phi^n(0) = \phi(\phi^{n-1}(0)) = \phi(a_{n-1}) \leq \phi(\hat{a}) = \hat{a}.$$

The inequality follows from the fact that ϕ is an increasing function.

Example 1. Suppose $p_0 = 1/4, p_1 = 1/4, p_2 = 1/2$. Then $\mu = 5/4$ and

$$\phi(a) = \frac{1}{4} + \frac{1}{4}a + \frac{1}{2}a^2.$$

Solving $a = \phi(a)$ gives the solutions $a = 1, 1/2$. The extinction probability is $1/2$.

Example 2. Suppose $p_0 = 1/2, p_1 = 1/4, p_2 = 1/4$. Then $\mu = 3/4$ and

$$\phi(a) = \frac{1}{2} + \frac{1}{4}a + \frac{1}{4}a^2.$$

Solving $a = \phi(a)$ gives the solutions $a = 1, 2$. The extinction probability is 1. (We had already demonstrated this fact since $\mu < 1$.)

Example 3. Suppose $p_0 = 1/4, p_1 = 1/2, p_2 = 1/4$. Then $\mu = 1$ and

$$\phi(a) = \frac{1}{4} + \frac{1}{4}a + \frac{1}{4}a^2.$$

Solving $a = \phi(a)$ gives the solutions $a = 1, 1$. The extinction probability is 1.

We finish by establishing a criterion to determine whether or not $a < 1$. We have already seen that if $\mu < 1$, then $a = 1$. Suppose $\mu = 1$. By (2.13), $\phi'(1) = 1$ and therefore by (2.14), $\phi'(s) < 1$ for $s < 1$. Hence for any $s < 1$,

$$1 - \phi(s) = \int_s^1 \phi'(s)ds < 1 - s,$$

i.e., $\phi(s) > s$. Therefore, if $\mu = 1$, the extinction probability is 1. This is an interesting result—even though the expected population size stays at 1, the probability that the population has died out increases to 1. One corollary of this is that the conditional size of the population conditioned that it has not died out must increase with time. That is to say, if one is told at some large time that the population has *not* died out, then one would expect the population to be large.

Now assume $\mu > 1$. Then $\phi'(1) > 1$ and hence there must be some $s < 1$ with $\phi(s) < s$. But $\phi(0) > 0$. By standard continuity arguments, we see that there must be some $a \in (0, s)$ with $\phi(a) = a$. Since $\phi''(s) > 0$ for $s \in (0, 1)$, the curve is convex and there can be at most one $s \in (0, 1)$ with $\phi(s) = s$. In this case, with positive probability the population lives forever. We summarize these results as a theorem.

Theorem. *If $\mu \leq 1$ and $p_0 > 0$, the extinction probability $a = 1$, i.e., the population eventually dies out. If $\mu > 1$, then the extinction probability $a < 1$ and equals the unique root of the equation*

$$t = \phi(t),$$

with $0 < t < 1$.

2.5 Exercises

2.1 Consider the queueing model (Example 3 of Section 2.1). For which values of p, q is the chain null recurrent, positive recurrent, transient?

For the positive recurrent case give the limiting probability distribution π . What is the average length of the queue in equilibrium?

For the transient case, give $\alpha(x) =$ the probability starting at x of ever reaching state 0.

2.2 Consider the following Markov chain with state space $S = \{0, 1, \dots\}$. A sequence of positive numbers p_1, p_2, \dots is given with $\sum_{i=1}^{\infty} p_i = 1$. Whenever the chain reaches state 0 it chooses a new state according to the p_i . Whenever

the chain is at a state other than 0 it proceeds deterministically, one step at a time, toward 0. In other words, the chain has transition probability

$$p(x, x - 1) = 1, \quad x > 0,$$

$$p(0, x) = p_x, \quad x > 0.$$

This is a recurrent chain since the chain keeps returning to 0. Under what conditions on the p_x is the chain positive recurrent? In this case, what is the limiting probability distribution π ? [Hint: it may be easier to compute $\mathbb{E}(T)$ directly where T is the time of first return to 0 starting at 0.]

2.3 Consider the Markov chain with state space $S = \{0, 1, 2, \dots\}$ and transition probabilities:

$$p(x, x + 1) = 2/3; \quad p(x, 0) = 1/3.$$

Show that the chain is positive recurrent and give the limiting probability π .

2.4 Consider the Markov chain with state space $S = \{0, 1, 2, \dots\}$ and transition probabilities:

$$p(x, x + 2) = p, \quad p(x, x - 1) = 1 - p, \quad x > 0.$$

$$p(0, 2) = p, \quad p(0, 0) = 1 - p.$$

For which values of p is this a transient chain?

2.5 Let X_n be the Markov chain with state space \mathbb{Z} and transition probability

$$p(n, n + 1) = p, \quad p(n, n - 1) = 1 - p,$$

where $p > 1/2$. Assume $X_0 = 0$.

- (a) Let $Y = \min\{X_0, X_1, \dots\}$. What is the distribution of Y ?
- (b) For positive integer k , let $T_k = \min\{n : X_n = k\}$ and let $e(k) = \mathbb{E}[T_k]$. Explain why $e(k) = k e(1)$.
- (c) Find $e(1)$. (Hint: (b) might be helpful.)
- (d) Use (c) to give another proof that $e(1) = \infty$ if $p = 1/2$.

2.6 Suppose J_1, J_2, \dots are independent random variables with $\mathbb{P}\{J_j = 1\} = 1 - \mathbb{P}\{J_j = 0\} = p$. Let k be a positive integer and let T_k be the first time that k consecutive 1s have appeared. In other words, $T_k = n$ if $J_n = J_{n-1} = \dots = J_{n-(k-1)} = 1$ and there is no $m < n$ such that $J_m = J_{m-1} = \dots = J_{m-(k-1)} = 1$. Let $X_0 = 0$ and for $n > 0$, let X_n be the number of consecutive 1s in the last run, i.e., $X_n = k$ if $J_{n-k} = 0$ and $J_i = 1$ for $n - k < i \leq n$.

- (a) Explain why X_n is a Markov chain with state space $\{0, 1, 2, \dots\}$ and give the transition probabilities.
- (b) Show that the chain is irreducible and positive recurrent and give the invariant probability π .
- (c) Find $\mathbb{E}[T_k]$ by writing an equation for $\mathbb{E}[T_k]$ in terms of $\mathbb{E}[T_{k-1}]$ and then solving the recursive equation.
- (d) Find $\mathbb{E}[T_k]$ in a different way. Suppose the chain starts in state k , and let \hat{T}_k be the time until returning to state k and \hat{T}_0 the time until the chain reaches state 0. Explain why

$$\mathbb{E}[\hat{T}_k] = \mathbb{E}[\hat{T}_0] + \mathbb{E}[T_k],$$

find $\mathbb{E}[\hat{T}_0]$, and use part (b) to determine $\mathbb{E}[\hat{T}_k]$.

2.7 Let X_n be a Markov chain with state space $S = \{0, 1, 2, \dots\}$. For each of the following transition probabilities, state if the chain is positive recurrent, null recurrent, or transient. If it is positive recurrent, give the stationary probability distribution:

- (a) $p(x, 0) = 1/(x+2)$, $p(x, x+1) = (x+1)/(x+2)$;
- (b) $p(x, 0) = (x+1)/(x+2)$, $p(x, x+1) = 1/(x+2)$;
- (c) $p(x, 0) = 1/(x^2+2)$, $p(x, x+1) = (x^2+1)/(x^2+2)$.

2.8 Given a branching process with the following offspring distributions, determine the extinction probability a .

- (a) $p_0 = .25, p_1 = .4, p_2 = .35$.
- (b) $p_0 = .5, p_1 = .1, p_3 = .4$.
- (c) $p_0 = .91, p_1 = .05, p_2 = .01, p_3 = .01, p_6 = .01, p_{13} = .01$.
- (d) $p_i = (1-q)q^i$, for some $0 < q < 1$.

2.9 Consider the branching process with offspring distribution as in Exercise 2.8(b) and suppose $X_0 = 1$.

(a) What is the probability that the population is extinct in the second generation ($X_2 = 0$), given that it did not die out in the first generation ($X_1 > 0$)?

(b) What is the probability that the population is extinct in the third generation, given that it was not extinct in the second generation?

2.10 Consider a branching process with offspring distribution given by $\{p_n\}$. We will make the process into an irreducible Markov chain by asserting that if the population ever dies out, then the next generation will have one new individual [in other words, $p(0, 1) = 1$]. For which $\{p_n\}$ will this chain be positive recurrent, null recurrent, transient?

2.11 Consider the following variation of the branching process. At each time n , each individual produces offspring independently using offspring distribution $\{p_n\}$, and then the individual dies with probability $q \in (0, 1)$. Hence,

each individual reproduces j times where j is the lifetime of the individual. For which values of q , $\{p_n\}$ do we have eventual extinction with probability one?

2.12 Consider the branching process with $p_0 = 1/3, p_1 = 1/3, p_2 = 1/3$. Find, with the aid of a computer, the probability that the population dies out after n steps where $n = 20, 100, 200, 1000, 1500, 2000, 5000$. Do the same with values $p_0 = .35, p_1 = .33, p_2 = .32$, and then do it with values $p_0 = .32, p_1 = .33, p_2 = .35$.

2.13 Consider a population of animals with the following rule for (asexual) reproduction: an individual that is born has probability q of surviving long enough to produce offspring. If the individual does produce offspring, she produces one or two offspring, each with equal probability. After this the individual no longer reproduces and eventually dies. Suppose the population starts with four individuals.

(a) For which values of q is it guaranteed that the population will eventually die out?

(b) If $q = .9$, what is the probability that the population survives forever?

2.14 Let X_n be the number of individuals at time n of a branching process with $\mu > 1$. Assume $X_0 = 1$. Let ϕ be the generating function for the offspring distribution, and let $a < 1$ be the extinction probability.

(a) Explain why $\phi'(a) < 1$.

(b) Let $a_n = \mathbb{P}\{X_n = 0\}$. Using part (a) show that there is a $\rho < 1$ such that for all n sufficiently large

$$a - a_{n+1} \leq \rho(a - a_n).$$

(c) Show that there exist $b > 0, c < \infty$ such that for all n ,

$$\mathbb{P}\{\text{extinction} \mid X_n \neq 0\} \leq c e^{-bn}.$$

In other words, if the population is going to go extinct it is very likely to do it in the first few generations.

2.15 Let X_1, X_2, \dots be independent identically distributed random variables taking values in the integers with mean 0. Let $S_0 = 0$ and

$$S_n = X_1 + \cdots + X_n.$$

(a) Let

$$G_n(x) = \mathbb{E} \left[\sum_{j=0}^n I\{S_j = x\} \right]$$

be the expected number of visits to x in the first n steps. Show that for all n and x , $G_n(0) \geq G_n(x)$. (Hint: consider the first j with $S_j = x$.)

(b) Recall that the law of large numbers implies that for each $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|S_n| \leq n\epsilon\} = 1.$$

Show that this implies that for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|x| \leq \epsilon n} G_n(x) = 1.$$

(c) Using (a) and (b), show that for each $M < \infty$ there is an n such that $G_n(0) \geq M$.

(d) Conclude that S_n is a recurrent Markov chain.

2.16 Let p_1, p_0, p_{-1}, \dots be a probability distribution on $\{\dots, -2, -1, 0, 1\}$ with negative mean

$$\sum_n np_n = \mu < 0.$$

Define a Markov chain X_n on the nonnegative integers with transition probabilities

$$p(n, m) = p_{m-n}, \quad m > 0,$$

$$p(n, 0) = \sum_{m \leq 0} p_{m-n}.$$

In other words, X_n acts like a random walk with increments given by the p_i except that the walk is forbidden to jump below 0. The purpose of this exercise is to show that the chain is positive recurrent.

(a) Let $\pi(n)$ be an invariant probability for the chain. Show that for each $n > 0$,

$$\pi(n) = \sum_{m=n-1}^{\infty} \pi(m)p_{n-m}.$$

(b) Let $q_n = p_{1-n}$. Show there exists an $\alpha \in (0, 1)$ such that

$$\alpha = q_0 + q_1\alpha + q_2\alpha^2 + \dots.$$

(Hint: q_n is the probability distribution of a random variable with mean greater than 1. The right-hand side is the generating function of the q_n .)

(c) Use the α from (b) to find the invariant probability distribution for the chain.

2.17 Let $p(x, y)$ be the transition probability for a Markov chain on a state space S . Call a function f superharmonic at x for p if

$$\sum_{y \in S} p(x, y)f(y) \leq f(x).$$

Fix a state $z \in S$.

(a) Let \mathcal{A} be the set of all functions f with $f(z) = 1$; $0 \leq f(y) \leq 1$ for all $y \in S$; and that are superharmonic at all $y \neq z$. Let g be defined by

$$g(x) = \inf_{f \in \mathcal{A}} f(x).$$

Show that $g \in \mathcal{A}$.

(b) Show that for all $x \neq z$,

$$\sum_{y \in S} p(x, y)g(y) = g(x).$$

[Hint: suppose $\sum_y p(x, y)g(y) < g(x)$ for some x . Show how you can decrease g a little at x so that the function stays superharmonic.]

(c) Let g be as in (a). Show that if $g(x) < 1$ for some x , then

$$\inf_{x \in S} g(x) = 0.$$

[Hint: let $\epsilon = \inf_x g(x)$ and consider $h(x) = (g(x) - \epsilon)/(1 - \epsilon)$.]

(d) Conclude the following: suppose that an irreducible Markov chain with transition probabilities $p(x, y)$ is given and there is a function f that is superharmonic for p at all $y \neq z$; $f(z) = 1$; $0 \leq f(y) \leq 1$, $y \in S$; and such that $f(x) < 1$ for some $x \in S$. Then the chain is transient.

2.18 In this exercise, we will establish Stirling's formula

$$n! \sim \sqrt{2\pi} n^{n+(1/2)} e^{-n}. \quad (2.15)$$

Let X_1, X_2, \dots be independent Poisson random variables with mean 1 and let $Y_n = X_1 + \dots + X_n$ which is a Poisson random variable with mean n . Let

$$p(n, k) = \mathbb{P}\{Y_n = k\} = e^{-n} \frac{n^k}{k!}.$$

(a) Use the central limit theorem to show that if $a > 0$,

$$\lim_{n \rightarrow \infty} \sum_{n \leq k \leq n+a\sqrt{n}} p(n, k) = \int_0^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

(b) Show that if $a > 0$, n is a positive integer, and $n \leq k < n + a\sqrt{n}$, then

$$e^{-a^2} p(n, n) \leq p(n, k) \leq p(n, n).$$

(c) Use (a) and (b) to conclude that

$$p(n, n) \sim \frac{1}{\sqrt{2\pi n}}.$$

Stirling's formula (2.15) follows immediately.

Chapter 3

Continuous-Time Markov Chains

3.1 Poisson Process

Consider X_t the number of customers arriving at a store by time t . Time is now continuous so t takes values in the nonnegative real numbers. Suppose we make three assumptions about the rate at which customers arrive. Intuitively, they are as follows:

1. The number of customers arriving during one time interval does not affect the number arriving during a different time interval.
2. The “average” rate at which customers arrive remains constant.
3. Customers arrive one at a time.

We now make these assumptions mathematically precise. The first assumption is easy: for $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, the random variables $X_{t_1} - X_{s_1}, \dots, X_{t_n} - X_{s_n}$ are independent. For the second two assumptions, let λ be the rate at which customers arrive, i.e., on the average we expect λt customers in time t . In a small time interval $[t, t + \Delta t]$, we expect that a new customer arrives with probability about $\lambda \Delta t$. The third assumption states that the probability that more than one customer comes in during a small time interval is significantly smaller than this. Rigorously, this becomes

$$\mathbb{P}\{X_{t+\Delta t} = X_t\} = 1 - \lambda \Delta t + o(\Delta t), \quad (3.1)$$

$$\mathbb{P}\{X_{t+\Delta t} = X_t + 1\} = \lambda \Delta t + o(\Delta t), \quad (3.2)$$

$$\mathbb{P}\{X_{t+\Delta t} \geq X_t + 2\} = o(\Delta t). \quad (3.3)$$

Here $o(\Delta t)$ represents some function that is much smaller than Δt for Δt small, i.e.,

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

A stochastic process X_t with $X_0 = 0$ satisfying these assumptions is called a *Poisson process with rate parameter λ* .

We will now determine the distribution of X_t . We will actually derive the distribution in two different ways. First, consider a large number n and write

$$X_t = \sum_{j=1}^n [X_{jt/n} - X_{(j-1)t/n}]. \quad (3.4)$$

We have written X_t as the sum of n independent, identically distributed random variables. If n is large, the probability that any of these random variables is 2 or more is small; in fact,

$$\begin{aligned} & \mathbb{P}\{X_{jt/n} - X_{(j-1)t/n} \geq 2 \text{ for some } j \leq n\} \\ & \leq \sum_{j=1}^n \mathbb{P}\{X_{jt/n} - X_{(j-1)t/n} \geq 2\} \\ & = n \mathbb{P}\{X_{t/n} \geq 2\}. \end{aligned}$$

The last term goes to 0 as $n \rightarrow \infty$ by (3.3). Hence we can approximate the sum in (3.4) by a sum of independent random variables which equal 1 with probability $\lambda(t/n)$ and 0 with probability $1 - \lambda(t/n)$. By the formula for the binomial distribution,

$$\mathbb{P}\{X_t = k\} \approx \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}.$$

Rigorously, we can then show:

$$\mathbb{P}\{X_t = k\} = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}.$$

To take this limit, note that

$$\lim_{n \rightarrow \infty} \binom{n}{k} n^{-k} = \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-k+1)}{k! n^k} = \frac{1}{k!},$$

and

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^{n-k} = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^{-k} = e^{-\lambda t}.$$

Hence,

$$\mathbb{P}\{X_t = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

i.e., X_t has a Poisson distribution with parameter λt .

We now derive this formula in a different way. Let

$$P_k(t) = \mathbb{P}\{X_t = k\}.$$

Note that $P_0(0) = 1$ and $P_k(0) = 0$, $k > 0$. Equations (3.1) through (3.3) can be used to give a system of differential equations for $P_k(t)$. The definition of the derivative gives

$$P'_k(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\mathbb{P}\{X_{t+\Delta t} = k\} - \mathbb{P}\{X_t = k\}).$$

Note that

$$\begin{aligned} \mathbb{P}\{X_{t+\Delta t} = k\} &= \mathbb{P}\{X_t = k\} \mathbb{P}\{X_{t+\Delta t} = k \mid X_t = k\} \\ &\quad + \mathbb{P}\{X_t = k-1\} \mathbb{P}\{X_{t+\Delta t} = k \mid X_t = k-1\} \\ &\quad + \mathbb{P}\{X_t \leq k-2\} \mathbb{P}\{X_{t+\Delta t} = k \mid X_t \leq k-2\} \\ &= P_k(t)(1 - \lambda\Delta t) + P_{k-1}(t)\lambda\Delta t + o(\Delta t). \end{aligned}$$

Therefore,

$$P'_k(t) = \lambda P_{k-1}(t) - \lambda P_k(t).$$

We can solve these equations recursively. For $k = 0$, the differential equation

$$P'_0(t) = -\lambda P_0(t), \quad P_0(0) = 1$$

has the solution

$$P_0(t) = e^{-\lambda t}.$$

To solve for $k > 0$ it is convenient to consider

$$f_k(t) = e^{\lambda t} P_k(t).$$

Then $f_0(t) = 1$ and the differential equation becomes

$$f'_k(t) = \lambda f_{k-1}(t), \quad f_k(0) = 0.$$

It is then easy to check inductively that the solution is

$$f_k(t) = \lambda^k t^k / k!,$$

and hence

$$P_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

which is what we derived previously.

Another way to view the Poisson process is to consider the waiting times between customers. Let $T_n, n = 1, 2, \dots$ be the time between the arrivals of the $(n-1)$ st and n th customers. Let $Y_n = T_1 + \dots + T_n$ be the total amount of time until n customers arrive. We can write

$$Y_n = \inf\{t : X_t = n\},$$

$$T_n = Y_n - Y_{n-1}.$$

Here inf stands for “infimum” or least upper bound which is the generalization of minimum for infinite sets; e.g., the infimum of the set of positive numbers is 0. The T_i should be independent, identically distributed random variables. One property that the T_i should satisfy is the loss of memory property: if we have waited s time units for a customer and no one has arrived, the chance that a customer will come in the next t time units is exactly the same as if there had been some customers before. Mathematically, this property is written

$$\mathbb{P}\{T_i \geq s + t \mid T_i \geq s\} = \mathbb{P}\{T_i \geq t\}.$$

The only real-valued functions satisfying $f(s+t) = f(s)f(t)$ are of the form $f(t) = e^{-bt}$. Hence the distribution of T_i must be an exponential distribution with parameter b . [Recall that a random variable Z has an exponential distribution with rate parameter b if it has density

$$f(z) = be^{-bz}, \quad 0 < z < \infty,$$

or equivalently, if it has distribution function

$$F(z) = \mathbb{P}\{Z \leq z\} = 1 - e^{-bz}, \quad z \geq 0.$$

An easy calculation gives $\mathbb{E}(Z) = 1/b$.] It is easy to see what b should be. For large t values we expect for there to be about λt customers. Hence, $Y_{\lambda t} \approx t$. But $Y_n \approx n\mathbb{E}(T_i) = n/b$. Hence $\lambda = b$. This gives a means of constructing a Poisson process: take independent random variables T_1, T_2, \dots , each exponential with rate λ , and define

$$Y_n = T_1 + \cdots + T_n,$$

$$X_t = n, \quad \text{if } Y_n \leq t < Y_{n+1}.$$

From this we could then conclude in a third way that the random variables X_t have a Poisson distribution. Conversely, given that we already have the Poisson process, it is easy to compute the distribution of T_i since

$$\mathbb{P}\{T_1 > t\} = \mathbb{P}\{X_t = 0\} = e^{-\lambda t}.$$

3.2 Finite State Space

In this section we discuss continuous-time Markov chains on a finite state space. We start by discussing some facts about exponential random variables.

Suppose T_1, \dots, T_n are independent random variables, each exponential with rates b_1, \dots, b_n , respectively. Intuitively, we can think of n alarm clocks which will go off at times T_1, \dots, T_n . Consider the first time when any of the alarm clocks goes off; more precisely, consider the random variable

$$T = \min\{T_1, \dots, T_n\}.$$

Note that

$$\begin{aligned}\mathbb{P}\{T \geq t\} &= \mathbb{P}\{T_1 \geq t, \dots, T_n \geq t\} \\ &= \mathbb{P}\{T_1 \geq t\} \mathbb{P}\{T_2 \geq t\} \cdots \mathbb{P}\{T_n \geq t\} \\ &= e^{-b_1 t} e^{-b_2 t} \cdots e^{-b_n t} = e^{-(b_1 + \cdots + b_n)t}.\end{aligned}$$

In other words, T has an exponential distribution with parameter $b_1 + \cdots + b_n$. Moreover, it is easy to give the probabilities for which of the clocks goes off first,

$$\begin{aligned}\mathbb{P}\{T_1 = T\} &= \int_0^\infty \mathbb{P}\{T_2 > t, \dots, T_n > t\} d\mathbb{P}\{T_1 = t\} \\ &= \int_0^\infty e^{-(b_2 + \cdots + b_n)t} b_1 e^{-b_1 t} dt \\ &= \frac{b_1}{b_1 + \cdots + b_n}.\end{aligned}$$

In other words, the probability that the i th clock goes off first is the ratio of b_i to $b_1 + \cdots + b_n$. If we are given an infinite sequence of exponential random variables T_1, T_2, \dots , with parameters b_1, b_2, \dots , the same result holds provided that $b_1 + b_2 + \cdots < \infty$.

Suppose now that we have a finite state space S . We will define a continuous-time process X_t on S that has the Markov property,

$$\mathbb{P}\{X_t = y \mid X_r, 0 \leq r \leq s\} = \mathbb{P}\{X_t = y \mid X_s\},$$

and that is time-homogeneous,

$$\mathbb{P}\{X_t = y \mid X_s = x\} = \mathbb{P}\{X_{t-s} = y \mid X_0 = x\}.$$

For each $x, y \in S, x \neq y$ we assign a nonnegative number $\alpha(x, y)$ that we think of as the rate at which the chain changes from state x to state y . We let $\alpha(x)$ denote the total rate at which the chain is changing from state x , i.e.,

$$\alpha(x) = \sum_{y \neq x} \alpha(x, y).$$

A (*time-homogeneous*) *continuous-time Markov chain* with rates α is a stochastic process X_t taking values in S satisfying

$$\mathbb{P}\{X_{t+\Delta t} = x \mid X_t = x\} = 1 - \alpha(x)\Delta t + o(\Delta t), \quad (3.5)$$

$$\mathbb{P}\{X_{t+\Delta t} = x \mid X_t = y\} = \alpha(y, x)\Delta t + o(\Delta t), \quad y \neq x. \quad (3.6)$$

In other words, the probability that the chain in state y jumps to a different state x in a small time interval of length Δt is about $\alpha(y, x)\Delta t$. For the Poisson process, we used the description for small Δt to write differential equations for the probabilities. We do the same in this case. If we let $p_x(t) = \mathbb{P}\{X_t = x\}$, then the equations above can be shown to give a system of linear differential equations,

$$p'_x(t) = -\alpha(x)p_x(t) + \sum_{y \neq x} \alpha(y, x)p_y(t).$$

If we impose an initial condition, $p_x(0), x \in S$, then we can solve the system. This system is often written in matrix form. Let \mathbf{A} be the matrix whose (x, y) entry equals $\alpha(x, y)$ if $x \neq y$ and equals $-\alpha(x)$ if $x = y$. Then if $\bar{p}(t)$ denotes the vector of probabilities, the system can be written

$$\bar{p}'(t) = \bar{p}(t)\mathbf{A}. \quad (3.7)$$

The matrix \mathbf{A} is called the *infinitesimal generator* of the chain. Note that the row sums of \mathbf{A} equal 0, the nondiagonal entries of \mathbf{A} are nonnegative, and the diagonal entries are nonpositive. From differential equations (see Section 0.2), we can give the solution

$$\bar{p}(t) = \bar{p}(0)e^{t\mathbf{A}}.$$

We can also write this in terms of transition matrices. Let $p_t(x, y) = \mathbb{P}\{X_t = y \mid X_0 = x\}$ and let \mathbf{P}_t be the matrix whose (x, y) entry is $p_t(x, y)$. The system of differential equations can be written as a single matrix equation:

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\mathbf{A}, \quad \mathbf{P}_0 = \mathbf{I}. \quad (3.8)$$

The matrix \mathbf{P}_t is then given by

$$\mathbf{P}_t = e^{t\mathbf{A}}.$$

Example 1. Consider a chain with two states—0, 1. Assume $\alpha(0, 1) = 1$ and $\alpha(1, 0) = 2$. Then the infinitesimal generator is

$$\mathbf{A} = \begin{smallmatrix} 0 & 1 \\ -1 & 1 \\ 1 & -2 \end{smallmatrix}.$$

In order to compute $e^{t\mathbf{A}}$, we diagonalize the matrix. The eigenvalues are 0, -3. We can write

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q},$$

where

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & -3 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}, \quad \mathbf{Q}^{-1} = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & -1/3 \end{bmatrix}.$$

We use the diagonalization to compute the exponential $e^{t\mathbf{A}}$.

$$\begin{aligned} \mathbf{P}_t = e^{t\mathbf{A}} &= \sum_{n=0}^{\infty} \frac{(t\mathbf{A})^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{\mathbf{Q}(t\mathbf{D})^n \mathbf{Q}^{-1}}{n!} \\ &= \mathbf{Q} \begin{bmatrix} 1 & 0 \\ 0 & e^{-3t} \end{bmatrix} \mathbf{Q}^{-1} \\ &= \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} + e^{-3t} \begin{bmatrix} 1/3 & -1/3 \\ -2/3 & 2/3 \end{bmatrix}. \end{aligned}$$

Note that

$$\lim_{t \rightarrow \infty} \mathbf{P}_t = \begin{bmatrix} \bar{\pi} \\ \bar{\pi} \end{bmatrix},$$

where $\bar{\pi} = (2/3, 1/3)$.

Example 2. Consider a chain with four states—0, 1, 2, 3—and infinitesimal generator

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ -1 & 1 & 0 & 0 \\ 1 & -3 & 1 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & 1 & -2 \end{bmatrix}.$$

The eigenvalues of \mathbf{A} are 0, -1, -3, -4 with right eigenvectors (which are left eigenvectors as well since \mathbf{A} is symmetric) (1, 1, 1, 1), (1, 0, -1/2, -1/2), (0, 0, -1/2, 1/2), and (-1/3, 1, -1/3, 1/3). Then,

$$\mathbf{D} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q},$$

where

$$\begin{aligned} \mathbf{D} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 1 & 1 & 0 & -1/3 \\ 1 & 0 & 0 & 1 \\ 1 & -1/2 & -1/2 & -1/3 \\ 1 & -1/2 & 1/2 & -1/3 \end{bmatrix}, \\ \mathbf{Q}^{-1} &= \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 2/3 & 0 & -1/3 & -1/3 \\ 0 & 0 & -1 & 1 \\ -1/4 & 3/4 & -1/4 & -1/4 \end{bmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{P}_t &= e^{t\mathbf{A}} = \mathbf{Q}e^{t\mathbf{D}}\mathbf{Q}^{-1} = \\ &\left[\begin{array}{cccc} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right] + e^{-t} \left[\begin{array}{cccc} 2/3 & 0 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 \\ -1/3 & 0 & 1/6 & 1/6 \\ -1/3 & 0 & 1/6 & 1/6 \end{array} \right] \\ &+ e^{-3t} \left[\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/2 -1/2 \end{array} \right] + e^{-4t} \left[\begin{array}{cccc} 1/12 & -1/4 & 1/12 & 1/12 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ 1/12 & -1/4 & 1/12 & 1/12 \\ 1/12 & -1/4 & 1/12 & 1/12 \end{array} \right]. \end{aligned}$$

Note that

$$\lim_{t \rightarrow \infty} \mathbf{P}_t = \left[\begin{array}{cccc} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right].$$

We can use exponential waiting times to give an alternative description of the Markov chain. Suppose rates $\alpha(x, y)$ have been given. Suppose $X_0 = x$. Let

$$T = \inf\{t : X_t \neq x\},$$

i.e., T is the time at which the process first changes state. The Markov property can be used to see that T must have the loss of memory property, and hence T must have an exponential distribution. By (3.5),

$$\mathbb{P}\{T \leq \Delta t\} = \alpha(x)\Delta t + o(\Delta t).$$

In order for this to be true, T must be exponential with parameter $\alpha(x)$. What state does the chain move to? The infinitesimal characterization (3.6) can be used to check that the probability that the state changes to y is exactly $\alpha(x, y)/\alpha(x)$. By the discussion of exponential distributions above we can think of this in another way. Independent “alarm clocks” are placed at each state y , with each alarm going off at an exponential time with rate $\alpha(x, y)$. The chain stays in state x until the first such clock goes off and then it moves to the state corresponding to that clock.

As in the case for discrete time, we are interested in the large-time behavior. As Examples 1 and 2 in this section demonstrate we expect

$$\lim_{t \rightarrow \infty} \mathbf{P}_t = \Pi_t = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix},$$

where $\bar{\pi}$ represents a limiting probability. The limiting probability should not change with time; hence, by (3.7),

$$\bar{\pi}\mathbf{A} = \bar{0}.$$

In this case, $\bar{\pi}$ is an eigenvector of \mathbf{A} with eigenvalue 0. The limit theory now parallels that for discrete time. Suppose for ease that the chain is irreducible. [A continuous-time Markov chain is irreducible if all states communicate, i.e., for each $x, y \in S$, there exist $z_1, \dots, z_j \in S$ with $\alpha(x, z_1), \alpha(z_1, z_2), \dots, \alpha(z_{j-1}, z_j), \alpha(z_j, y)$ all strictly positive.] In this case, one can show (see Exercise 3.4) using the results for stochastic matrices that:

1. There is a unique probability vector $\bar{\pi}$ satisfying

$$\bar{\pi}\mathbf{A} = \bar{0}.$$

2. All other eigenvalues of \mathbf{A} have negative real part.

By analyzing the matrix differential equation it is not too difficult to show that

$$\lim_{t \rightarrow \infty} \mathbf{P}_t = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}.$$

If the chain is reducible, we must analyze the chain on each communication class. We have not discussed periodicity. This phenomenon does not occur for continuous-time chains; in fact, one can prove (see Exercise 3.7) that for any irreducible continuous-time chain, \mathbf{P}_t has strictly positive entries for all $t > 0$.

A number of the methods for analyzing discrete-time chains have analogues for continuous-time chains. Suppose X_t is an irreducible continuous-time chain on finite state space S and suppose z is some fixed state in S . We will compute the *mean passage time* to z starting at state x , i.e., $b(x) = \mathbb{E}(Y | X_0 = x)$, where

$$Y = \inf\{t : X_t = z\}.$$

Clearly, $b(z) = 0$. For $x \neq z$, assume $X_0 = x$ and let T be the first time that the chain changes state as above. Then

$$\mathbb{E}(Y | X_0 = x) = \mathbb{E}(T | X_0 = x) + \sum_{y \in S} \mathbb{P}\{X_T = y | X_0 = x\} \mathbb{E}(Y | X_0 = y).$$

Since T is exponential with parameter $\alpha(x)$ the first term on the right hand side equals $1/\alpha(x)$. Also from the above discussion, $\mathbb{P}\{X_t = y | X_0 = x\} =$

$\alpha(x, y)/\alpha(x)$. Finally, since $b(z) = 0$, we do not need to include the $y = z$ term in the sum. Therefore, the equation becomes

$$\alpha(x) b(x) = 1 + \sum_{y \neq x, z} \alpha(x, y) b(y).$$

If we let $\tilde{\mathbf{A}}$ be the matrix obtained from \mathbf{A} by deleting the row and column associated to the state z , we get the matrix equation

$$\bar{0} = \bar{1} + \tilde{\mathbf{A}}\bar{b},$$

or

$$\bar{b} = [-\tilde{\mathbf{A}}]^{-1}\bar{1}.$$

(The matrix $\tilde{\mathbf{A}}$ is a square matrix whose row sums are all nonpositive and at least one of whose row sums is strictly negative. From this one can conclude that all the eigenvalues of $\tilde{\mathbf{A}}$ have strictly negative real part, and hence $\tilde{\mathbf{A}}$ is invertible.)

Example 3. Consider Example 2 in this section and let us compute the expected time to get from state 0 to state 3. Then $z = 3$,

$$\tilde{\mathbf{A}} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \left[\begin{matrix} -1 & 1 & 0 \\ 1 & -3 & 1 \\ 0 & 1 & -2 \end{matrix} \right]. \end{matrix}$$

and

$$\bar{b} = [-\tilde{\mathbf{A}}]^{-1}\bar{1} = (8/3, 5/3, 4/3).$$

Therefore the expected time to get from state 0 to state 3 is $8/3$.

3.3 Birth-and-Death Processes

In this section we consider a large class of infinite state space, continuous-time Markov chains that are known by the name of *birth-and-death processes*. The state space will be $\{0, 1, 2, \dots\}$, and changes of state will always be from n to $n+1$ or n to $n-1$. Intuitively we can view the state of the system as the size of a population that can increase or decrease by 1 by a “birth” or a “death,” respectively. To describe the chain, we give birth rates $\lambda_n, n = 0, 1, 2, \dots$ and death rates $\mu_n, n = 1, 2, 3, \dots$. If the population is currently n , then new individuals arrive at rate λ_n and individuals leave at rate μ_n (note if the population is 0 there can be no deaths, so $\mu_0 = 0$).

If we let X_t denote the state of the chain at time t , then

$$\mathbb{P}\{X_{t+\Delta t} = n \mid X_t = n\} = 1 - (\mu_n + \lambda_n)\Delta t + o(\Delta t),$$

$$\mathbb{P}\{X_{t+\Delta t} = n+1 \mid X_t = n\} = \lambda_n \Delta t + o(\Delta t),$$

$$\mathbb{P}\{X_{t+\Delta t} = n-1 \mid X_t = n\} = \mu_n \Delta t + o(\Delta t).$$

As before, we can convert these equations into differential equations for $P_n(t) = \mathbb{P}\{X_t = n\}$ and get the system

$$P'_n(t) = \mu_{n+1}P_{n+1}(t) + \lambda_{n-1}P_{n-1}(t) - (\mu_n + \lambda_n)P_n(t). \quad (3.9)$$

To compute the transition probabilities

$$p_t(m, n) = \mathbb{P}\{X_t = n \mid X_0 = m\}$$

we need only solve the system with initial conditions,

$$P_m(0) = 1, \quad P_n(0) = 0, \quad n \neq m.$$

Example 1. The Poisson process with rate parameter λ is a birth-and-death process with $\lambda_n = \lambda$ and $\mu_n = 0$.

Example 2. Markovian Queueing Models. Suppose X_t denotes the number of people on line for some service. We assume that people arrive at a rate λ ; more precisely, the arrival rate of customers follows a Poisson process with rate λ . Customers are also serviced at an exponential rate μ . We note three different service rules:

(a) $M/M/1$ queue. In this case there is one server and only the first person in line is being serviced. This gives a birth-and-death process with $\lambda_n = \lambda$ and $\mu_n = \mu$ ($n \geq 1$). The two M s in the notation refer to the fact that both the arrival and the service times are exponential and hence the process is Markovian. The 1 denotes the fact that there is one server.

(b) $M/M/k$ queue. In this case there are k servers and anyone in the first k positions in the line can be served. If there are k people being served, and each one is served at rate μ , then the rate at which people are leaving the system is $k\mu$. This gives a birth-and-death process with $\lambda_n = \lambda$ and

$$\mu_n = \begin{cases} n\mu, & \text{if } n \leq k, \\ k\mu, & \text{if } n \geq k. \end{cases}$$

(c) $M/M/\infty$ queue. In this case there are an infinite number of servers, so everyone in line has a chance of being served. In this case $\lambda_n = \lambda$ and $\mu_n = n\mu$.

Example 3. Population Model. Imagine that the state of the chain represents the number of individuals in a population. Each individual at a certain rate λ produces another individual. Similarly each individual dies at rate μ . If all the individuals act independently this can be modelled by a birth-and-death process with $\lambda_n = n\lambda$ and $\mu_n = n\mu$. Note that 0 is an absorbing state in this model. When $\mu = 0$, this is sometimes called the Yule process.

Example 4. Population Model with Immigration. Assume that individuals die and reproduce with rates μ and λ , respectively, as in the previous model. We also assume that new individuals arrive at a constant rate ν . This gives a birth-and-death process with $\lambda_n = n\lambda + \nu$ and $\mu_n = n\mu$.

Example 5. Fast-Growing Population Model. Imagine that a population grows at a rate proportional to the square of the number of individuals. Then if we assume no deaths, we have a process with $\lambda_n = n^2\lambda$ and $\mu_n = 0$. The population in this case grows very fast, and we will see later that it actually reaches an “infinite population” in finite time.

We will look more closely at all of these examples, but first we develop some general theory. We call the birth-and-death chain irreducible if all the states communicate. It is not very difficult to see that this happens if and only if all the λ_n ($n \geq 0$) and all the μ_n ($n \geq 1$) are positive. An irreducible chain is recurrent if one always returns to a state; otherwise, it is called transient. For any birth-and-death process, there is a discrete-time Markov chain on $\{0, 1, 2, \dots\}$ that follows the continuous-time chain “when it moves.” It has transition probabilities

$$p(n, n-1) = \frac{\mu_n}{\mu_n + \lambda_n}, \quad p(n, n+1) = \frac{\lambda_n}{\mu_n + \lambda_n}.$$

One can check that the continuous-time chain is recurrent if and only if the corresponding discrete-time chain is recurrent. Let $a(n)$ be the probability that the chain starting at state n ever reaches state 0. Note that $a(0) = 1$ and the value of $a(n)$ is the same whether one considers the continuous-time or the discrete-time chain. From our discussion of discrete-time chains, we see that $a(n)$ satisfies

$$a(n)(\mu_n + \lambda_n) = a(n-1)\mu_n + a(n+1)\lambda_n, \quad n > 0. \quad (3.10)$$

If the chain is transient, $a(n) \rightarrow 0$ as $n \rightarrow \infty$. If the chain is recurrent, no solution of this equation will exist with $a(0) = 1, 0 \leq a(n) \leq 1, a(n) \rightarrow 0$ ($n \rightarrow \infty$).

We now give a necessary and sufficient condition for a birth-and-death chain to be transient. We will try to find the function $a(n)$. Equation (3.10) can be rewritten

$$a(n) - a(n+1) = \frac{\mu_n}{\lambda_n} [a(n-1) - a(n)], \quad n \geq 1.$$

If we continue, we get

$$a(n) - a(n+1) = \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} [a(0) - a(1)].$$

Hence,

$$\begin{aligned} a(n+1) &= [a(n+1) - a(0)] + a(0) \\ &= \sum_{j=0}^n [a(j+1) - a(j)] + 1 \\ &= [a(1) - 1] \sum_{j=0}^n \frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j} + 1, \end{aligned}$$

where the $j = 0$ term of the sum equals 1 by convention. We can find a nontrivial solution if the sum converges. We have established the following.

Fact. *The birth-and-death chain is transient if and only if*

$$\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} < \infty. \quad (3.11)$$

As an example, consider the queueing models (Example 2). For the $M/M/1$ queue,

$$\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \sum_{n=1}^{\infty} \left(\frac{\mu}{\lambda}\right)^n,$$

which converges if and only if $\mu < \lambda$. Consider now the $M/M/k$ queue. For any $n \geq k$,

$$\frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \frac{k!}{k^k} \left(\frac{k\mu}{\lambda}\right)^n.$$

Therefore, in this case the sum is finite and the chain is transient if and only if $k\mu < \lambda$. Finally for the $M/M/\infty$ queue,

$$\sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} = \sum_{n=1}^{\infty} n! \left(\frac{\mu}{\lambda}\right)^n = \infty.$$

Hence, for all values of μ and λ the chain is recurrent. These three results can be summarized by saying that the queueing models are transient (and hence the lines grow longer and longer) if and only if the (maximal) service rate is strictly less than the arrival rate.

For recurrent chains, there may or may not be a limiting probability. Again, we call an irreducible chain positive recurrent if there exists a probability distribution $\pi(n)$ such that

$$\lim_{t \rightarrow \infty} \mathbb{P}\{X_t = n \mid X_0 = m\} = \pi(n).$$

for all states m . Otherwise a recurrent chain is called null recurrent. If the system is in the limiting probability, i.e., if $P_n(t) = \pi(n)$, where $P_n(t)$ is as in (3.9), then $P'_n(t)$ should equal 0. In other words π should satisfy

$$0 = \lambda_{n-1} \pi(n-1) + \mu_{n+1} \pi(n+1) - (\lambda_n + \mu_n) \pi(n). \quad (3.12)$$

Again, as for the case of discrete-time chains, we can find π by solving these equations. If we can find a probability distribution that satisfies (3.12), then the chain is positive recurrent and that distribution is the unique equilibrium distribution.

We can solve (3.12) directly. First, the equation for $n = 0$ gives

$$\pi(1) = \frac{\lambda_0}{\mu_1} \pi(0).$$

For $n \geq 1$, the equation can be written

$$\mu_{n+1} \pi(n+1) - \lambda_n \pi(n) = \mu_n \pi(n) - \lambda_{n-1} \pi(n-1).$$

If we iterate this equation, we get

$$\mu_{n+1} \pi(n+1) - \lambda_n \pi(n) = \mu_1 \pi(1) - \lambda_0 \pi(0) = 0.$$

Hence, $\pi(n+1) = (\lambda_n / \mu_{n+1}) \pi(n)$, and by iterating we get the solution

$$\pi(n) = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi(0).$$

We now impose the condition that π be a probability measure. We can arrange this if and only if $\sum \pi(x) < \infty$. We have established the following.

Fact. *A birth-and-death chain is positive recurrent if and only if*

$$q = \sum_{n=0}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$$

(by convention, the $n = 0$ term in this sum is equal to 1). In this case the invariant probability is given by

$$\pi(n) = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} q^{-1}. \quad (3.13)$$

As an example, consider the queueing models again. For the $M/M/1$ queue,

$$\sum_{n=0}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n = \left(1 - \frac{\lambda}{\mu} \right)^{-1},$$

provided $\lambda < \mu$ and is infinite otherwise. Hence this chain is positive recurrent for $\lambda < \mu$ in which case the equilibrium distribution is

$$\pi(n) = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^n.$$

Note that the expected length of the queue in equilibrium is

$$\sum_{n=0}^{\infty} n\pi(n) = \sum_{n=0}^{\infty} n \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^n = \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu} \right)^{-1} = \frac{\lambda}{\mu - \lambda}.$$

In particular, the expected length gets large as λ approaches μ . In the case of the $M/M/k$ queue, the exact form of π is a little messy, but it is easy to verify that the chain is positive recurrent if and only if $\lambda < k\mu$. Finally for the $M/M/\infty$ queue,

$$\sum_{n=0}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n = e^{\lambda/\mu}.$$

Hence, the chain is positive recurrent for all λ, μ and has equilibrium distribution

$$\pi(n) = e^{-\lambda/\mu} \frac{(\lambda/\mu)^n}{n!},$$

i.e., the equilibrium distribution is a Poisson distribution with parameter λ/μ . The mean queue length in equilibrium is λ/μ .

Conditions under which the population models are positive recurrent, null recurrent, or transient are discussed in Exercises 3.12 and 3.13.

We finish by considering two pure birth processes. A birth-and-death process is a pure birth process if $\mu_n = 0$ for all n . We first consider the Yule process with $\lambda_n = n\lambda$. Let us assume that the population starts with one individual; hence, $P_1(0) = 1, P_n(t) = 0$ ($n > 1$), where again $P_n(t) = \mathbb{P}\{X_t = n\}$. The $P_n(t)$ s satisfy the differential equations

$$P'_n(t) = (n-1)\lambda P_{n-1}(t) - n\lambda P_n(t), \quad n \geq 1.$$

One can solve these equations recursively, but since the computations are a little messy, we will skip them and simply state that the solution is

$$P_n(t) = e^{-\lambda t} [1 - e^{-\lambda t}]^{n-1}, \quad n \geq 1.$$

(It is not too difficult to verify that $P_n(t)$ defined as above does satisfy these equations.) The form for $P_n(t)$ is nice; in fact, for a fixed t , X_t has a geometric distribution with parameter $p = e^{-\lambda t}$. This allows us immediately to compute the expected population size at time t ,

$$\mathbb{E}(X_t) = \sum_{n=1}^{\infty} n P_n(t) = e^{\lambda t}.$$

We could derive this last result in a different way. Let $f(t) = \mathbb{E}(X_t)$. Then

$$\begin{aligned} f'(t) &= \sum_{n=1}^{\infty} n P'_n(t) = \sum_{n=1}^{\infty} n [(n-1)\lambda P_{n-1}(t) - n\lambda P_n(t)] \\ &= \sum_{n=1}^{\infty} n\lambda P_n(t) = \lambda f(t). \end{aligned}$$

Therefore, $f(t)$ satisfies the standard equation for exponential growth and the initial condition $f(0) = 1$ immediately gives the solution $f(t) = e^{\lambda t}$. There is one other way we can look at the Yule process. Consider the time Y_n when the population first reaches n , i.e.,

$$Y_n = \inf\{t : X_t = n\}.$$

Then $Y_n = T_1 + \dots + T_{n-1}$, where T_i measures the time between the arrival of the i th and $(i+1)$ st individual. The random variables T_i are independent and T_i has an exponential distribution with parameter $i\lambda$. In particular $\mathbb{E}(T_i) = 1/(i\lambda)$ and $\text{Var}(T_i) = 1/(i\lambda)^2$. Therefore,

$$\mathbb{E}(Y_n) = \sum_{i=1}^{n-1} \frac{1}{i\lambda} \sim \frac{\ln n}{\lambda}.$$

Also $\text{Var}(Y_n) \leq \sum_{i=1}^{\infty} (i\lambda)^{-2} < \infty$. Hence, Y_n equals $\ln n/\lambda$ up to a small random error which is bounded as n gets large. If it takes time $\ln n/\lambda$ to reach a population of n individuals, then in time t we would expect $e^{\lambda t}$ individuals.

Now consider the fast-growing population model, Example 5, with $\lambda_n = n^2\lambda$. Again let us consider Y_n the time until the n th individual enters the population. In this case, an interesting phenomenon occurs. Consider

$$Y_{\infty} = T_1 + T_2 + T_3 + \dots$$

Then

$$\mathbb{E}(Y_{\infty}) = \sum_{i=1}^{\infty} \mathbb{E}(T_i) = \sum_{i=1}^{\infty} \frac{1}{i^2\lambda} < \infty.$$

In particular, with probability 1, $Y_{\infty} < \infty$! This says that in finite time the population grows to an infinite size. This phenomenon is often called

explosion. For a pure birth process, explosion occurs if and only if $\mathbb{E}(Y_\infty) < \infty$, i.e., if and only if

$$\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty.$$

3.4 General Case

Suppose we have a countable (perhaps infinite) state space S and rates $\alpha(x, y)$ denoting the rate at which the state is changing from x to y . Suppose for each x ,

$$\alpha(x) = \sum_{y \neq x} \alpha(x, y) < \infty.$$

Then we can use the “exponential alarm clocks” at each state in order to construct a time-homogeneous, continuous-time Markov chain X_t such that for each $x \neq y$,

$$\mathbb{P}\{X_{t+\Delta t} = y \mid X_t = x\} = \alpha(x, y)\Delta t + o_x(\Delta t).$$

Here we write $o_x(\cdot)$ to show that the size of the error term can depend on the state x . If the rates α are not bounded, it is possible for the chain to have explosion in finite time as was seen in the case of the fast-growing population model in Section 3.3. Let us assume for the time being that we have a chain for which explosion does not occur (it is sometimes difficult to determine whether or not explosion occurs).

We will consider the transition probabilities

$$p_t(x, y) = \mathbb{P}\{X_t = y \mid X_0 = x\} = \mathbb{P}\{X_{t+s} = y \mid X_s = x\}.$$

To derive a differential equation for the transition probabilities in the same manner as in the previous sections, we write

$$\begin{aligned} p_{t+\Delta t}(x, y) &= p_t(x, y)p_{\Delta t}(y, y) + \sum_{z \neq y} p_t(x, z)p_{\Delta t}(z, y) \\ &= p_t(x, y)[1 - \alpha(y)\Delta t + o_y(\Delta t)] \\ &\quad + \sum_{z \neq y} p_t(x, z)[\alpha(z, y)\Delta t + o_z(\Delta t)] \\ &= p_t(x, y)[1 - \alpha(y)\Delta t] + \sum_{z \neq y} p_t(x, z)\alpha(z, y)\Delta t \\ &\quad + \sum_z p_t(x, z)o_z(\Delta t). \end{aligned}$$

If we can combine the last error term so that

$$\sum_z p_t(x, z) o_z(\Delta t) = o(\Delta t), \quad (3.14)$$

then we can conclude that the transition probabilities satisfy the system of equations

$$p'_t(x, y) = -\alpha(y)p_t(x, y) + \sum_{z \neq y} \alpha(z, y)p_t(x, z),$$

where the derivative is with respect to time. These are sometimes called the *forward equations* for the chain. In most cases of interest, including all the examples in the first three sections, (3.14) can be justified. There are examples, however, where the forward equations cannot be justified.

There is another set of equations called the *backward equations* which always hold. For the backward equations we write

$$\begin{aligned} p_{t+\Delta t}(x, y) &= \sum_z p_{\Delta t}(x, z)p_t(z, y) \\ &= \sum_{z \neq x} [\alpha(x, z)\Delta t + o_x(\Delta t)]p_t(z, y) \\ &\quad + [1 - \alpha(x)\Delta t + o_x(\Delta t)]p_t(x, y). \end{aligned}$$

The error term depends only on x . With a little work one can show that one can always take the limit as Δt goes to 0 and get

$$p'_t(x, y) = -\alpha(x)p_t(x, y) + \sum_{z \neq x} \alpha(x, z)p_t(z, y).$$

In the case of a finite state space with infinitesimal generator \mathbf{A} , the backward equations for the transition matrix \mathbf{P}_t becomes in matrix form

$$\frac{d}{dt} \mathbf{P}_t = \mathbf{A}\mathbf{P}_t,$$

which can be compared to the forward equation (3.8). Both equations (with initial condition $\mathbf{P}_0 = \mathbf{I}$) have the solution

$$\mathbf{P}_t = e^{t\mathbf{A}}.$$

3.5 Exercises

3.1 Suppose that the number of calls per hour arriving at an answering service follows a Poisson process with $\lambda = 4$.

- (a) What is the probability that fewer than two calls come in the first hour?
- (b) Suppose that six calls arrive in the first hour. What is the probability that at least two calls will arrive in the second hour?
- (c) The person answering the phones waits until fifteen phone calls have arrived before going to lunch. What is the expected amount of time that the person will wait?
- (d) Suppose it is known that exactly eight calls arrived in the first two hours. What is the probability that exactly five of them arrived in the first hour?
- (e) Suppose it is known that exactly k calls arrived in the first four hours. What is the probability that exactly j of them arrived in the first hour?

3.2 Let X_t and Y_t be two independent Poisson processes with rate parameters λ_1 and λ_2 , respectively, measuring the number of customers arriving in stores 1 and 2, respectively.

- (a) What is the probability that a customer arrives in store 1 before any customers arrive in store 2?
- (b) What is the probability that in the first hour, a total of exactly four customers have arrived at the two stores?
- (c) Given that exactly four customers have arrived at the two stores, what is the probability that all four went to store 1?
- (d) Let T denote the time of arrival of the first customer at store 2. Then X_T is the number of customers in store 1 at the time of the first customer arrival at store 2. Find the probability distribution of X_T (i.e., for each k , find $\mathbb{P}\{X_T = k\}$).

3.3 Suppose X_t and Y_t are independent Poisson processes with parameters λ_1 and λ_2 , respectively, measuring the number of calls arriving at two different phones. Let $Z_t = X_t + Y_t$.

- (a) Show that Z_t is a Poisson process. What is the rate parameter for Z ?
- (b) What is the probability that the first call comes on the first phone?
- (c) Let T denote the first time that at least one call has come from each of the two phones. Find the density and distribution function of the random variable T .

3.4 Let \mathbf{A} be the infinitesimal generator for an irreducible, continuous-time Markov chain with finite state space. Then the rows of \mathbf{A} add up to 0 and the nondiagonal elements of \mathbf{A} are nonnegative.

- (a) Let a be some positive number greater than all the entries of \mathbf{A} . Let $\mathbf{P} = (1/a)\mathbf{A} + \mathbf{I}$. Show that \mathbf{P} is the transition matrix for a discrete-time, irreducible, aperiodic Markov chain.
- (b) Use this to conclude: \mathbf{A} has a unique left eigenvector with eigenvalue 0 that is a probability vector and all the other eigenvalues of \mathbf{A} have real part strictly less than 0.

3.5 Let X_t be a Markov chain with state space $\{1, 2\}$ and rates $\alpha(1, 2) = 1, \alpha(2, 1) = 4$. Find \mathbf{P}_t .

3.6 Repeat Exercise 3.5 with state space $\{1, 2, 3\}$ and rates $\alpha(1, 2) = 1, \alpha(2, 1) = 4, \alpha(2, 3) = 1, \alpha(3, 2) = 4, \alpha(1, 3) = 0, \alpha(3, 1) = 0$.

3.7 Let X_t be an irreducible, continuous-time Markov chain. Show that for each i, j and every $t > 0$,

$$\mathbb{P}\{X_t = j \mid X_0 = i\} > 0.$$

3.8 Consider the continuous-time Markov chain with state space $\{1, 2, 3, 4\}$ and infinitesimal generator

$$\mathbf{A} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \left[\begin{matrix} -3 & 1 & 1 & 1 \\ 0 & -3 & 2 & 1 \\ 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & -1 \end{matrix} \right] \end{matrix}.$$

- (a) Find the equilibrium distribution $\bar{\pi}$.
- (b) Suppose the chain starts in state 1. What is the expected amount of time until it changes state for the first time?
- (c) Again assume the chain starts in state 1. What is the expected amount of time until the chain is in state 4?

3.9 Repeat Exercise 3.8 with

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \left[\begin{matrix} -2 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 1 & -3 & 1 \\ 0 & 0 & 1 & -1 \end{matrix} \right] \end{matrix}.$$

3.10 Suppose α gives the rates for an irreducible continuous-time Markov chain on a finite state space. Suppose the invariant probability measure is π . Let

$$p(x, y) = \alpha(x, y)/\alpha(x), \quad x \neq y,$$

be the transition probability for the discrete-time Markov chain corresponding to the continuous-time chain “when it moves.” Find the invariant probability for the discrete-time chain in terms of π and α .

3.11 Let X_t be a continuous-time birth-and-death process with birth rate $\lambda_n = 1 + (1/(n+1))$ and death rate $\mu_n = 1$. Is this process positive recurrent, null recurrent, or transient? What if $\lambda_n = 1 - (1/(n+2))$?

3.12 Consider the population model (Example 3, Section 3.3). For which values of μ and λ is extinction certain, i.e., when is the probability of reaching state 0 equal to 1?

3.13 Consider the population model with immigration (Example 4, Section 3.3). For which values of μ, λ, ν is the chain positive recurrent, null recurrent, transient?

3.14 Consider a birth-and-death process with $\lambda_n = 1/(n + 1)$ and $\mu_n = 1$. Show that the process is positive recurrent and give the stationary distribution.

3.15 Suppose one has a *deterministic* model for population where the population grows proportionately to the square of the current population. In other words, the population $p(t)$ satisfies the differential equation

$$\frac{dp}{dt} = c[p(t)]^2,$$

for some constant $c > 0$. Assume $p(0) = 1$. Solve this differential equation (by separation of variables) and describe what happens as time increases.

3.16 Consider a birth-and-death process with birth rates λ_n and death rates μ_n . What are the backward equations for the transition probabilities $p_t(m, n)$?

Chapter 4

Optimal Stopping

4.1 Optimal Stopping of Markov Chains

Imagine the following simple game. A player rolls a die. If the player rolls a 6 the player wins no money. Otherwise, the player may either quit the game and win k dollars, where k is the roll of the die, or may roll again. If the player rolls again, the game continues until either a 6 is rolled or the player quits. The total payoff for the game is always k dollars, where k is the value of the last roll (unless the roll is a 6 in which case the payoff is 0). What is the optimal strategy for the player?

In order to determine the optimal strategy, it is necessary to decide what should be optimized. For example, if the player only wants to guarantee that the payoff is positive, then the game should be stopped after the first roll—either the player has already lost (if a 6 is rolled) or the player can guarantee a positive payoff by stopping. However, it is reasonable to consider what happens if the player decides to maximize the *expected* payoff. Let us analyze this problem and then show how this applies to more general Markov chain problems.

We first let $f(k)$ denote the payoff associated with each roll. In this example $f(k) = k$ if $k \leq 5$ and $f(6) = 0$. We let $v(k)$ be the expected winnings of the player given that the first roll is k *assuming that the player takes the optimal strategy*. At this moment we may not know what the optimal strategy is, but it still makes sense to discuss v . We will, in fact, write down an equation that v satisfies and use this to determine v and the optimal strategy. We first note that $v(6) = 0$ and $v(5) = 5$. The latter is true since it clearly does not pay to roll again if the first roll is 5, so the optimal strategy is to stop and pick up \$5. It is not so clear what $v(k)$ is for $k \leq 4$.

Now let $u(k)$, $k \leq 5$ be the amount of payoff that is expected if the player does not stop after rolling a k , but from then on plays according to the optimal strategy. [In this particular example, $u(k)$ is actually the same for all k .] Then it is easy to see that

$$u(k) = \frac{1}{6}v(1) + \frac{1}{6}v(2) + \frac{1}{6}v(3) + \frac{1}{6}v(4) + \frac{1}{6}v(5) + \frac{1}{6}v(6).$$

We now can write the optimal strategy in terms of $u(k)$ —if $f(k) > u(k)$, the

player should stop and take the money; if $f(k) < u(k)$, the player should roll again. In other words,

$$v(k) = \max\{f(k), u(k)\}.$$

In particular, $v(k) \geq f(k)$. This fact implies that $u(k) \geq (f(1) + \dots + f(6))/6 = 5/2$. We now know more about the optimal strategy—if the first roll is a 1 or a 2 the player should roll again. Hence,

$$v(1) = \frac{v(1) + \dots + v(6)}{6} = \frac{v(1) + \dots + v(4)}{6} + \frac{5}{6},$$

$$v(2) = \frac{v(1) + \dots + v(4)}{6} + \frac{5}{6}.$$

Suppose the first roll is a 4. Suppose that the optimal strategy were to continue playing. Then clearly that would also be the optimal strategy if the first roll is a 3. Under this strategy, the game would continue until a 5 or a 6 is rolled and each of these ending rolls would be equally likely. This would give an expected payoff of $(5+0)/2 = 5/2$, which is less than 4. Hence this cannot be the optimal strategy starting with a 4. The player, therefore, should stop with a 4 and $v(4) = f(4) = 4$. We finally consider what happens if the first roll is a 3. Suppose the player rolls again whenever a 3 comes up and uses the optimal strategy otherwise. Let u be the expected winnings in this case. Then

$$u = \mathbb{P}\{\text{roll } \leq 3\}u + \frac{1}{6}4 + \frac{1}{6}5 = \frac{1}{2}u + \frac{1}{6}4 + \frac{1}{6}5.$$

Solving for u we get $u = 9/3$. Since this equals $f(3)$, the expected payoff for playing is the same as for stopping and $v(3) = 3$. With these values, we can solve for $v(1)$ and $v(2)$, getting $v(1) = v(2) = 3$. The optimal strategy is to play if the first roll is 1 or 2; stop if the first roll is 4, 5, 6; and either play or stop if the first roll is a 3.

We now generalize these ideas. Suppose \mathbf{P} is the transition matrix for a discrete-time Markov chain X_n with state space S . For ease we will assume that S is finite, but much of what follows can be applied to the infinite state space case. Assume there is a payoff function f that assigns to each state the payoff if the chain is stopped when it reaches that state. In cases of interest, \mathbf{P} will not be irreducible since otherwise one could always continue until one reached the state that has the maximum payoff. A *stopping rule* or *stopping time* will be a random variable T that gives the time at which the chain is stopped. It is important that one must decide whether or not to stop based only on what has happened up through step n ; in other words, one cannot look into the future to decide whether or not to stop. Because we are dealing with a time-homogeneous Markov chain it does not take too much work to convince oneself that the only reasonable stopping rules that do not look into

the future are of the following form: the state space is divided into two sets S_1 and S_2 ; if the state of the chain is in S_1 one continues, if it is in S_2 it stops. The goal is to maximize the expected payoff over all stopping rules. We let $v(x)$ be the *value* of a state x , i.e., the expected payoff assuming that the optimal stopping strategy is used. We can write

$$v(x) = \max_T \mathbb{E} [f(X_T) \mid X_0 = x],$$

where the maximum is over all legal stopping rules.

There are two main inequalities that v satisfies. First, v is greater than or equal to the payoff available by stopping,

$$v(x) \geq f(x). \quad (4.1)$$

Second, v is greater than or equal to the maximum expected payoff if one continues,

$$v(x) \geq \mathbf{P}v(x) = \sum_{y \in S} p(x, y) v(y). \quad (4.2)$$

In fact, v is equal to the maximum of these values:

$$v(x) = \max\{f(x), \mathbf{P}v(x)\}. \quad (4.3)$$

If we let S_1 be the set of states where one continues and S_2 the set of states where one stops (assuming the optimal strategy), and we let

$$T = \min\{j \geq 0 : X_j \in S_2\},$$

then

$$v(x) = \mathbb{E} [f(X_T) \mid X_0 = x].$$

We will characterize the function v . We call a function u *superharmonic* with respect to \mathbf{P} if it satisfies (4.2), i.e.,

$$u(x) \geq \mathbf{P}u(x).$$

Suppose u is superharmonic and T is the time associated to a stopping rule as above. Consider the time $T_n = \min\{T, n\}$ We claim that

$$u(x) \geq \mathbb{E} [u(X_{T_n}) \mid X_0 = x].$$

To see this, note that it is trivially true for $n = 0$. Assume it is true for $n - 1$.

Then

$$\begin{aligned}
& \mathbb{E}[u(X_{T_n}) \mid X_0 = x] \\
&= \sum_{y \in S} \mathbb{P}\{X_{T_n} = y \mid X_0 = x\} u(y) \\
&= \sum_{y \in S} \sum_{z \in S} \mathbb{P}\{X_{T_n} = y \mid X_{T_{n-1}} = z\} \mathbb{P}\{X_{T_{n-1}} = z \mid X_0 = x\} u(y) \\
&= \sum_{z \in S_2} \sum_{y \in S} \mathbb{P}\{X_{T_n} = y \mid X_{T_{n-1}} = z\} \mathbb{P}\{X_{T_{n-1}} = z \mid X_0 = x\} u(y) + \\
&\quad \sum_{z \in S_1} \sum_{y \in S} \mathbb{P}\{X_{T_n} = y \mid X_{T_{n-1}} = z\} \mathbb{P}\{X_{T_{n-1}} = z \mid X_0 = x\} u(y).
\end{aligned}$$

If $z \in S_2$, then $\mathbb{P}\{X_{T_n} = z \mid X_{T_{n-1}} = z\} = 1$ and hence the first double sum in the last expression equals

$$\sum_{z \in S_2} \mathbb{P}\{X_{T_{n-1}} = z \mid X_0 = x\} u(z).$$

If $z \in S_1$, $\mathbb{P}\{X_{T_n} = y \mid X_{T_{n-1}} = z\} = p(z, y)$ and hence

$$\sum_{y \in S} \mathbb{P}\{X_{T_n} = y \mid X_{T_{n-1}} = z\} u(y) = \mathbf{P}u(z) \leq u(z).$$

Hence,

$$\begin{aligned}
\mathbb{E}[u(X_{T_n}) \mid X_0 = x] &\leq \sum_{z \in S} \mathbb{P}\{X_{T_{n-1}} = z \mid X_0 = x\} u(z) \\
&= \mathbb{E}[u(X_{T_{n-1}}) \mid X_0 = x] \leq u(x).
\end{aligned}$$

Since u is a bounded function, we can let $n \rightarrow \infty$ and get

$$u(x) \geq \lim_{n \rightarrow \infty} \mathbb{E}[u(X_{T_n}) \mid X_0 = x] = \mathbb{E}[u(X_T) \mid X_0 = x].$$

Now suppose that $u(x) \geq f(x)$ for all x . Then

$$u(x) = \mathbb{E}[u(X_T) \mid X_0 = x] \geq \mathbb{E}[f(X_T) \mid X_0 = x] = v(x).$$

Hence every superharmonic function that is larger than f is greater than or equal to the value function v . Also we note (see Exercise 4.7) that if $\{u_i(x)\}$ is any collection of superharmonic functions, then

$$u(x) = \inf_i u_i(x)$$

is also superharmonic. We have derived the following.

Fact. v is the smallest superharmonic function with respect to \mathbf{P} that is greater than equal to f ; equivalently,

$$v(x) = \inf u(x),$$

where the infimum is over all superharmonic functions u with $u(x) \geq f(x)$.

The characterization leads to an algorithm for determining v . Start with the function $u_1(x)$ that equals $f(x)$ if x is an absorbing state and otherwise equals the maximum value of f . This gives a superharmonic function that is greater than f . Let

$$u_2(x) = \max\{\mathbf{P}u_1(x), f(x)\}.$$

Since u_1 is superharmonic and $u_1 \geq f$, $u_2(x) \leq u_1(x)$. Also,

$$\mathbf{P}u_2(x) \leq \mathbf{P}u_1(x) \leq u_2(x).$$

Hence, u_2 is a superharmonic function greater than f . Continuing, we define

$$u_n(x) = \max\{\mathbf{P}u_{n-1}(x), f(x)\},$$

and we see that u_n is a superharmonic function greater than f but less than u_{n-1} . We will show at the end of this section that

$$v(x) = \lim_{n \rightarrow \infty} u_n(x).$$

Example 1. If we consider the game that we already analyzed and started with the function $u = [5, 5, 5, 5, 5, 0]$, then in 10 iterations we would see $u_{10} = [3.002, 3.002, 3.002, 4, 5, 0]$.

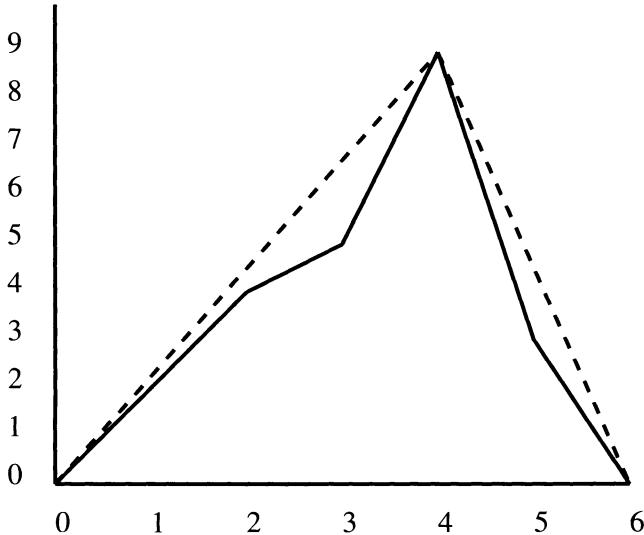
Example 2. Suppose X_n is a simple random walk ($p = 1/2$) with absorbing barriers on $\{0, 1, 2, 3, 4, 5, 6\}$. Let the payoff function f be given by $f = [0, 2, 4, 5, 9, 3, 0]$ (we write the payoff function as a vector in a natural way). We will first determine the optimal strategy. Clearly one stops at state 4 and one has to stop at 0 and 6. From state 5 there is a probability $1/2$ of going to 4 and $1/2$ of going to 6; the expected payoff given that we continue is at least $9/2 > f(5) = 3$, so from 5 we continue. If one starts in state 3, then one can get an expected payoff of $(4 + 9)/2 = 13/2$ by taking one step and then stopping. Since this is greater than $f(3) = 5$, it must be optimal to continue from state 3 and $v(3) \geq 13/2$. Note that from state 2 playing gives an expected payoff of at least $[f(1) + v(3)]/2 \geq 17/4 \geq f(2) = 4$. Hence, we continue on state 2 and $v(2) \geq 17/4$. Similarly, if one continues from state 1 we can obtain an expected payoff of $v(2)/2 \geq 17/8 > f(1) = 2$, so the optimal strategy is to continue. Therefore the stopping set in this case is $S_2 = \{0, 4, 6\}$. The value function can be obtained by

$$v(x) = \mathbb{E}[f(X_T) \mid X_0 = x] = 9 \mathbb{P}\{X_T = 4 \mid X_0 = x\}.$$

The probability has been computed before [see (1.16)] and we get

$$v = \left[0, \frac{9}{4}, \frac{9}{2}, \frac{27}{4}, 9, \frac{9}{2}, 0 \right].$$

In the graph below the solid line represents f and the dotted line represents v . For simple random walk, the superharmonic functions are the concave functions. The function v is the smallest concave function satisfying $v \geq f$.



In this example, if we had started with the function $u_1 = [0, 9, 9, 9, 9, 9, 0]$ and performed the algorithm above we would have gotten within .01 of the actual value of v in about 20 iterations.

In solving the optimal stopping problem we simultaneously compute the value function v and the optimal stopping strategy. Suppose that we knew the strategy that we would choose, i.e., we split the state space into two sets S_1 and S_2 so that we continue on S_1 and stop on S_2 . Let $u(x)$ be the expected payoff using this strategy. Then u satisfies:

$$u(x) = f(x), \quad x \in S_2, \tag{4.4}$$

$$u(x) = \mathbf{P}u(x), \quad x \in S_1. \tag{4.5}$$

This is a discrete analogue of a boundary value problem sometimes called the *Dirichlet problem*. The boundary is the set S_2 where prescribed values are

given. On the “interior” points S_1 , some difference equation holds. As we have seen the probabilistic form of the solution of this system can be given by

$$u(x) = \mathbb{E}[f(X_T) \mid X_0 = x],$$

where

$$T = \min\{j \geq 0 : X_j \in S_2\}.$$

For a finite-state Markov chain, the solution can be found directly because (4.4) and (4.5) give k linear equations in k unknowns, where k is the number of points in S_1 and the unknowns are $u(x), x \in S_1$.

We now verify that the algorithm does converge to the value function v . Let $u(z) = \lim_{n \rightarrow \infty} u_n(z)$. Since u is the decreasing limit of superharmonic functions, u is superharmonic (see Exercise 4.7). Also $u(z) \geq f(z)$ for all z . Hence by the characterization of v , we get

$$u(z) \geq v(z). \quad (4.6)$$

Let the stopping set S_2 be defined by

$$S_2 = \{z : u(z) = f(z)\},$$

$$S_1 = \{z : u(z) > f(z)\}.$$

On S_1 , $\mathbf{P}u(z) = u(z)$ (if $\mathbf{P}u(z) < u(z)$, then for some n , $\mathbf{P}u_n(z) < u(z) \leq u_n(z)$ and hence $u_{n+1}(z) = \max\{\mathbf{P}u_n(z), f(z)\} < u(z)$ which is impossible). Therefore,

$$u(z) = \mathbb{E}[u(X_T) \mid X_0 = z],$$

where T is the strategy associated with the sets S_1, S_2 . Since $v(z)$ is the largest expected value over all choices of stopping sets,

$$u(z) \leq v(z). \quad (4.7)$$

Combining (4.6) and (4.7) we see that $u(z) = v(z)$ for all z .

4.2 Optimal Stopping with Cost

Consider the first example of the previous section, and suppose that there is a charge of \$1 for each additional roll, i.e., on each roll we can either take the payoff associated with that roll or pay \$1 and roll again. In general, we can assume that there is a cost $g(x)$ associated with each state that must be

paid to continue the chain. As before we assume we have a payoff function f and we let $v(x)$ be the expected value of the payoff *minus the cost* assuming a stopping rule is chosen that maximizes this expected value. We can write

$$v(x) = \max_T \mathbb{E} \left[f(X_T) - \sum_{j=0}^{T-1} g(X_j) \mid X_0 = x \right],$$

where again the maximum is over all legal stopping times T . Then $v(x)$ satisfies:

$$v(x) = \max\{f(x), \mathbf{P}v(x) - g(x)\}.$$

Here, the expected payoff minus cost if the chain is continued is $\mathbf{P}v(x) - g(x)$. Again we can divide S into S_1 and S_2 where

$$S_2 = \{x : v(x) = f(x)\},$$

and the optimal stopping rule is to stop when the chain enters a state in S_2 .

Using a similar argument as in Section 4.1, the value function v for this example can be characterized as the smallest function u greater than f that satisfies

$$u(x) \geq \mathbf{P}u(x) - g(x).$$

In other words,

$$v(x) = \inf u(x),$$

where the infimum is over all u satisfying $u(x) \geq f(x)$ and $u(x) \geq \mathbf{P}u(x) - g(x)$. To find the value function, we may use an algorithm similar to that in Section 4.1. We define u_1 to be the function that equals f on all absorbing states and equals the maximum value of f everywhere else. We then define

$$u_n(x) = \max\{f(x), \mathbf{P}u_{n-1}(x) - g(x)\},$$

and then

$$v(x) = \lim_{n \rightarrow \infty} u_n(x).$$

Example 1. Suppose we consider the die game with $f = [1, 2, 3, 4, 5, 0]$ and $g = [1, 1, 1, 1, 1, 1]$. The cost function makes it less likely that we would want to roll again, so it is clear that we should stop if we get a 4 or a 5; similarly, we should stop if we get a 3 since we were indifferent before with this roll and it costs to roll again. If we get a 1, then by rolling again we can get an expected payoff of at least $5/2$ with a cost of 1. Hence we can expect a net gain of at least $3/2$. Therefore we should play if we get a 1.

Suppose we roll again whenever we get a 1 or 2 and stop otherwise. Let $u(k)$ be the expected winnings with this strategy. Then $u(1) = u(2) = u$ and $u(k) = k, k = 3, 4, 5$. Also,

$$u(2) = \frac{1}{6}u(1) + \frac{1}{6}u(2) + \frac{1}{6}u(3) + \frac{1}{6}u(4) + \frac{1}{6}u(5) + \frac{1}{6}u(6) - 1 = \frac{1}{3}u + 1.$$

Solving for u gives $u = 3/2$. Since this is less than $f(2) = 2$, it must be correct to stop at 2. Hence the stopping set is $S_2 = \{2, 3, 4, 5, 6\}$ and the value function is

$$v = [8/5, 2, 3, 4, 5, 0].$$

If we started with the initial $u_1 = [5, 5, 5, 5, 5, 0]$ and performed the algorithm described above, then after only a few iterations we would have

$$u_{10} = [1.6, 2, 3, 4, 5, 0].$$

Example 2. Consider the other example of the previous section where X_n is a simple random walk with absorbing boundary on $\{0, 1, \dots, 6\}$ and $f = [0, 2, 4, 5, 9, 3, 0]$. Suppose we impose a cost of .5 to move from states 0, 1, 2 and a cost of 1 to move from 3, 4, 5, 6, i.e., a cost function

$$g = [.5, .5, .5, 1, 1, 1, 1].$$

If we start with initial $u_1 = [0, 9, 9, 9, 9, 9, 0]$, then in only six iterations we get

$$u_6 = [0, 2, 4, 5.5, 9, 3.5, 0],$$

which gives the value for v . In this case the stopping set is $S_2 = \{0, 1, 2, 4, 6\}$.

Example 3. With a cost function, it is possible to have a nontrivial problem even if the Markov chain is irreducible. Suppose we play the following game: roll two dice; the player may stop at any time and take the roll on the dice or the player may pay 2 units if the roll is less than 5 and 1 unit if the roll is greater than or equal to 5 and roll again. In this case the state space is $\{2, 3, 4, \dots, 12\}$,

$$f = [2, 3, 4, \dots, 12], \quad g = [2, 2, 2, 1, 1, \dots, 1].$$

If we start with the initial guess $u_1 = [12, 12, \dots, 12]$ then within 20 iterations we converge to the value function v ,

$$v = [5\frac{2}{3}, 5\frac{2}{3}, 5\frac{2}{3}, 6\frac{2}{3}, 6\frac{2}{3}, 7, 8, 9, 10, 11, 12].$$

The stopping set is $S_2 = \{7, 8, \dots, 12\}$.

4.3 Optimal Stopping with Discounting

It is often appropriate in modelling financial matters to assume that the value of money decreases with time. Let us assume that a discount factor $\alpha < 1$ is given. By this we mean that 1 dollar received after one time unit is the same as α dollars received in the present. Again suppose we have a Markov chain X_n with transition matrix \mathbf{P} and a payoff function f . It is now the goal to optimize the expected value of the payoff, taking into consideration the decreasing value of the payoff. If we stop after k steps, then the present value of the payoff in k steps is α^k time the actual payoff.

In this case the value function is given by

$$v(x) = \max_T \mathbb{E} [\alpha^T f(X_T) \mid X_0 = x],$$

where again the maximum is over all legal stopping rules. To obtain this value function, we characterize v as the smallest function u satisfying

$$u(x) \geq f(x),$$

$$u(x) \geq \alpha \mathbf{P} u(x).$$

We may obtain v with a similar algorithm as before. Start with an initial function u_1 equal to f at all absorbing states and equal to the maximum value of f at all other states. Then define u_n recursively by

$$u_n(x) = \max\{f(x), \alpha \mathbf{P} u_{n-1}(x)\}.$$

Then

$$v(x) = \lim_{n \rightarrow \infty} u_n(x).$$

Example 1. Consider the die game again. Assume a discounting factor of $\alpha = .8$. Since discounting can only make it more likely to stop it is easy to see that one should stop if the first roll is a 3, 4, or 5. If the first roll is a 1, one can get an expected payoff of at least $.8[(1 + 2 + 3 + 4 + 5)/6] = 2$ by rolling again, so it is best to roll again. Suppose we use the strategy to roll again with a 1, 2 and to stop otherwise and let u be the expected winnings given that one rolls again. Then

$$u = .8 \left(\frac{u}{6} + \frac{u}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} \right).$$

Solving for u we get $u = 24/11 > 2$ so it must be optimal to roll again with a 2. Therefore $S_2 = \{3, 4, 5, 6\}$ and

$$v = \left[\frac{24}{11}, \frac{24}{11}, 3, 4, 5, 0 \right].$$

Example 2. Consider the example of a simple random walk with absorbing boundaries on $\{0, 1, \dots, 6\}$ and $f = [0, 2, 4, 5, 9, 3, 0]$. Suppose that there is no cost function, but the value of money is discounted at rate $\alpha = .9$. If we start with $u_1 = [0, 9, 9, 9, 9, 9, 0]$ then in seven iterations we converge to the value

$$u_7 = [0, 2, 4, 5.85, 9, 4.05, 0].$$

This stopping set is $\{0, 1, 2, 4, 6\}$.

It is possible to include both a cost function and a discounting factor. Suppose in addition to the other assumptions in this section, we have a cost function $g(x)$ that indicates the cost of taking a step given that the chain is in state x . Then the value function v is the smallest function u satisfying

$$u(x) \geq f(x),$$

$$u(x) \geq \alpha \mathbf{P}u(x) - g(x),$$

Example 3. Consider the random walk with absorbing boundaries described before with $f = [0, 2, 4, 5, 9, 3, 0]$ and with both the cost function $g = [.5, .5, .5, 1, 1, 1, 1]$ and the discount factor $\alpha = .9$. If we start with $u_1 = [0, 9, 9, 9, 9, 9, 0]$ then in only three iterations we converge to

$$v = [0, 2, 4, 5, 9, 3.05, 0].$$

The stopping set is $\{0, 1, 2, 3, 4, 6\}$.

Example 4. Consider a random walk with absorbing boundaries on the state space $\{0, 1, \dots, 10\}$. Suppose the payoff function is the square of the site stopped at, i.e.,

$$f = [0, 1, 4, 5, 9, \dots, 100].$$

We assume that there is a constant cost of .6 and a discounting factor of $\alpha = .95$. We then start with the initial guess

$$u_1 = [0, 100, 100, 100, \dots, 100]$$

and after 60 iterations we get

$$u_{60} = [0, 1.51, 4.45, 9.11, 16, 25, 36, 49, 64, 81, 100].$$

The stopping set is $\{0, 4, 5, 6, \dots, 10\}$.

4.4 Exercises

4.1 Consider a simple random walk ($p = 1/2$) with absorbing boundaries on $\{0, 1, 2, \dots, 10\}$. Suppose the following payoff function is given

$$[0, 2, 4, 3, 10, 0, 6, 4, 3, 3, 0].$$

Find the optimal stopping rule and give the expected payoff starting at each site.

4.2 The following game is played: you roll two dice. If you roll a 7, the game is over and you win nothing. Otherwise, you may stop and receive an amount equal to the sum of the two dice. If you continue, you roll again. The game ends whenever you roll a 7 or whenever you say stop. If you say stop before rolling a 7 you receive an amount equal to the sum of the two dice on the last roll. What is your expected winnings: a) if you always stop after the first roll; b) if you play to optimize your expected winnings?

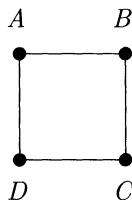
4.3 Consider Exercise 4.1. Do the problem again assuming:

- (a) a constant cost of .75 for each move;
- (b) a discount factor $\alpha = .95$;
- (c) both.

4.4 Consider Exercise 4.2. Do the problem again assuming:

- (a) a cost function of $g = [2, 2, 2, 2, 1, 1, 1, 1, 1, 1]$;
- (b) a discount factor $\alpha = .8$;
- (c) both.

4.5 Consider a simple random walk on the following four-vertex graph.



Assume that the payoff function is: $f(A) = 2, f(B) = 4, f(C) = 5, f(D) = 3$. Assume that there is no cost associated with moving, but there is a discount factor α . What is the largest possible value of α so that the optimal stopping strategy is to stop at every vertex, i.e., so that $S_2 = \{A, B, C, D\}$?

4.6 Consider the following simple game. You roll a single die. If it comes up 1 you lose. If it comes up $k > 1$, you can either take a payoff of k^2 or you can

play again. Hence, the final payoff is either 0 (if you roll a 1) or otherwise the square of the value of your final roll.

(a) What is the optimal strategy in this game and what is the expected winnings if one uses the optimal strategy?

(b) Suppose that it costs r to play the game each time. What is the smallest value of r such that the optimal strategy is to play if one rolls a 2 and to stop if one rolls any other number?

4.7 If $u_1(y), u_2(y), \dots$ are all functions that are superharmonic at x for \mathbf{P} , i.e.,

$$\mathbf{P}u_i(x) \leq u_i(x),$$

and we let u be the function

$$u(y) = \inf_i u_i(y),$$

show that u is superharmonic at x for \mathbf{P} .

4.8 Consider a simple “Wheel of Fortune” game. A wheel is divided into 12 equal-sized wedges. Eleven of the edges are marked with the numbers 100, 200, ..., 1100 denoting an amount of money won if the wheel lands on those numbers. The twelfth wedge is marked “bankrupt.” A player can spin as many times as he or she wants. Each time the wheel lands on a numbered wedge, the player receives that much money which is added to his/her previous winnings. However, if the wheel ever lands on the “bankrupt” wedge, the player loses all of his/her money that has been won up to that point. The player may quit at any time, and take all the money he or she has won (assuming the “bankrupt” wedge has not come up).

Assuming that the goal is to maximize one’s expected winnings in this game, devise an optimal strategy for playing this game and compute one’s expected winnings. You may wish to try a computer simulation first.

4.9 Suppose X_n is random walk with absorbing boundary on $\{0, 1, 2, \dots\}$ with

$$p(n, n+1) = p(n, n-1) = \frac{1}{2}, \quad n \geq 1.$$

Suppose our payoff function is $f(n) = n^2$. Let us try to find a stopping time T that will maximize $\mathbb{E}[f(X_T)]$.

(a) Show that if $X_n > 0$, then

$$\mathbb{E}[f(X_{n+1}) \mid X_n] > f(X_n).$$

Conclude that any optimal strategy does not stop at any integer greater than 0.

(b) Since the random walk is recurrent, we know that we will eventually reach 0 at which point we stop and receive a payoff of 0. Since our “optimal” strategy tells us never to stop before then, our eventual payoff in the optimal strategy is 0. Clearly something is wrong here—any ideas?

4.10 We have been restricting ourselves to stopping rules T that do not look into the future. Suppose we *can* look into the future so that we always know when we reach the site that will have the highest payoff. Explain why the expected payoff is

$$v_{\text{prop}}(x) := \mathbb{E} \left[\max_n f(X_n) \mid X_0 = x \right].$$

The subscript prop stands for “prophet.” Clearly $v_{\text{prop}}(x) \geq v(x)$.

(a) Find v_{prop} for the die game discussed at the beginning of the chapter (where the game stops whenever a 6 is rolled).

(b) Find v_{prop} for the chain and payoff function in Exercise 4.1.

Chapter 5

Martingales

5.1 Conditional Expectation

To understand martingales, which are a model for “fair games,” we first need to understand conditional expectation. We start with some easy examples and build up to a general definition. Suppose Y is a random variable measuring the outcome of some random experiment. If one knows nothing about the outcome of the experiment, then the best guess for the value of Y is $\mathbb{E}(Y)$, the expectation. Of course, if one has complete knowledge of the outcome of the experiment, then one knows the exact value of Y . Conditional expectation deals with making the best guess for Y given some but not all information about the outcome. We will start by discussing the conditional expectation of a random variable Y with respect to a finite number of random variables X_1, \dots, X_n and then finish by discussing the conditional expectation with respect to an infinite number of random variables.

Suppose that X and Y are discrete random variables with joint probability density function

$$f(x, y) = \mathbb{P}\{X = x, Y = y\}$$

and marginal probability density functions

$$f_X(x) = \sum_y f(x, y), \quad f_Y(y) = \sum_x f(x, y).$$

To define the conditional expectation of Y given X , $E(Y | X)$ we need to give the best value of Y for any value of x . A little thought will show that we should define

$$\begin{aligned} E(Y | X)(x) &= \sum_y y \mathbb{P}\{Y = y | X = x\} \\ &= \sum_y y \frac{\mathbb{P}\{X = x, Y = y\}}{\mathbb{P}\{X = x\}} \\ &= \frac{\sum_y y f(x, y)}{f_X(x)}. \end{aligned}$$

This is well defined if $f_X(x) > 0$, and we do not bother to define $E(Y | X)(x)$ for other values of x since such values occur with probability 0. As an example suppose that two independent dice are rolled and X denotes the value of the first roll and Y denotes the sum of the two rolls. Then

$$f(x, y) = \frac{1}{36}, \quad x = 1, 2, \dots, 6, \quad y = x + 1, x + 2, \dots, x + 6,$$

and

$$E(Y | X)(x) = x + \frac{7}{2}.$$

Similarly, if X_1, \dots, X_n, Y are discrete random variables with joint probability density function

$$f(x_1, \dots, x_n, y) = \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n, Y = y\},$$

and the marginal density with respect to X_1, \dots, X_n is given by

$$g(x_1, \dots, x_n) = \sum_y f(x_1, \dots, x_n, y),$$

then the conditional expectation of Y given X_1, \dots, X_n , is given by

$$E(Y | X_1, \dots, X_n)(x_1, \dots, x_n) = \frac{\sum_y y f(x_1, \dots, x_n, y)}{g(x_1, \dots, x_n)}.$$

This is well defined if x_1, \dots, x_n is a possible outcome for the experiment, i.e., if $g(x_1, \dots, x_n) > 0$. Again, we think of $E(Y | X_1, \dots, X_n)$ as being the best guess for the value of Y given the values of X_1, \dots, X_n .

If X and Y are continuous random variables with joint density $f(x, y)$ and marginal densities

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx,$$

then the conditional expectation of Y given X is defined in an analogous way

$$E(Y | X)(x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_X(x)},$$

which is well defined for $f_X(x) > 0$. Similarly if X_1, \dots, X_n, Y have joint density $f(x_1, \dots, x_n, y)$,

$$E(Y | X_1, \dots, X_n)(x_1, \dots, x_n) = \frac{\int_{-\infty}^{\infty} y f(x_1, \dots, x_n, y) dy}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}.$$

The conditional expectation $E(Y | X_1, \dots, X_n)$ is characterized by two properties:

1. The value of the random variable $E(Y | X_1, \dots, X_n)$ depends only on the values of X_1, \dots, X_n , i.e., we can write $E(Y | X_1, \dots, X_n) = \phi(X_1, \dots, X_n)$ for some function ϕ . If a random variable Z can be written as a function of X_1, \dots, X_n it is called *measurable* with respect to X_1, \dots, X_n . (For those who know measure theory, the function must be Borel measurable.)

2. Suppose A is any event that depends only on X_1, \dots, X_n . For example, A might be the event

$$A = \{a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n\}.$$

Let I_A denote the indicator function of A , i.e., the random variable which equals 1 if A occurs and 0 otherwise. Then

$$\mathbb{E}(Y I_A) = \mathbb{E}[E(Y | X_1, \dots, X_n) I_A]. \quad (5.1)$$

Let us derive the last equality in the case where X_1, \dots, X_n, Y are continuous random variables with density $f(x_1, \dots, x_n, y)$ and A is the above event; the derivation for discrete random variables is essentially the same.

$$\begin{aligned} & \mathbb{E}[E(Y | X_1, \dots, X_n) I_A] \\ &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \int_{-\infty}^{\infty} E(Y | X_1 = x_1, \dots, X_n = x_n) \\ & \quad f(x_1, \dots, x_n, y) dy dx_n \cdots dx_1 \\ &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \int_{-\infty}^{\infty} \left[\frac{\int_{-\infty}^{\infty} z f(x_1, \dots, x_n, z) dz}{\int_{-\infty}^{\infty} f(x_1, \dots, x_n, z) dz} \right] \\ & \quad f(x_1, \dots, x_n, y) dy dx_n \cdots dx_1 \\ &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \int_{-\infty}^{\infty} z f(x_1, \dots, x_n, z) dz dx_n \cdots dx_1 \\ &= \mathbb{E}(Y I_A). \end{aligned}$$

Conditions 1 and 2 give a complete characterization of the conditional expectation.

Fact. $E(Y | X_1, \dots, X_n)$ is the unique random variable which depends only on X_1, \dots, X_n and which satisfies (5.1) for every event A that depends only on X_1, \dots, X_n .

In measure theoretic treatments of probability, the conditional expectation is *defined* as the random variable satisfying conditions 1 and 2 and then it is proved that this uniquely defines a random variable (up to an event of probability zero). For our purposes, the characterization will be useful in deriving some properties of conditional expectation.

We will make the notation a little more compact. If X_1, X_2, \dots is a sequence of random variables we will use \mathcal{F}_n to denote the “information contained in

X_1, \dots, X_n ." We will write $E(Y | \mathcal{F}_n)$ for $E(Y | X_1, \dots, X_n)$. If we apply (5.1) to the event A consisting of the entire sample space (so that $I_A = 1$) we get

$$\mathbb{E}[E(Y | \mathcal{F}_n)] = \mathbb{E}(Y). \quad (5.2)$$

Conditional expectation is a linear operation: if a, b are constants, then

$$E(aY_1 + bY_2 | \mathcal{F}_n) = aE(Y_1 | \mathcal{F}_n) + bE(Y_2 | \mathcal{F}_n). \quad (5.3)$$

To prove this, we need only note that the right-hand side is measurable with respect to X_1, \dots, X_n and satisfies (5.1). The next two properties can be derived similarly. If Y is already a function of X_1, \dots, X_n then

$$E(Y | \mathcal{F}_n) = Y. \quad (5.4)$$

For any Y , if $m < n$, then

$$E(E(Y | \mathcal{F}_n) | \mathcal{F}_m) = E(Y | \mathcal{F}_m). \quad (5.5)$$

If Y is independent of X_1, \dots, X_n , then information about X_1, \dots, X_n should not be useful in determining Y and

$$E(Y | \mathcal{F}_n) = \mathbb{E}(Y). \quad (5.6)$$

This can be derived easily from (5.1) since in this case Y and I_A are independent random variables. The last property we will need is a little trickier: if Y is any random variable and Z is a random variable that is measurable with respect to X_1, \dots, X_n , then

$$E(YZ | \mathcal{F}_n) = ZE(Y | \mathcal{F}_n). \quad (5.7)$$

It is clear that the right-hand side is measurable with respect to X_1, \dots, X_n , so it suffices to show that it satisfies (5.1). We will not prove it here; the basic idea is to approximate Z by simple functions, for which (5.1) can be derived easily, and pass to the limit.

Example 1. Suppose X_1, X_2, \dots are independent, identically distributed random variables with mean μ . Let S_n denote the partial sum

$$S_n = X_1 + \dots + X_n.$$

Let \mathcal{F}_n denote the information in X_1, \dots, X_n . Suppose $m < n$. Then by (5.3),

$$E(S_n | \mathcal{F}_m) = E(X_1 + \dots + X_m | \mathcal{F}_m) + E(X_{m+1} + \dots + X_n | \mathcal{F}_m).$$

Since $X_1 + \dots + X_m$ is measurable with respect to X_1, \dots, X_m , by (5.4),

$$E(X_1 + \dots + X_m | \mathcal{F}_m) = X_1 + \dots + X_m = S_m.$$

Since $X_{m+1} + \dots + X_n$ is independent of X_1, \dots, X_m , by (5.6),

$$E(X_{m+1} + \dots + X_n | \mathcal{F}_m) = \mathbb{E}(X_{m+1} + \dots + X_n) = (n - m)\mu.$$

Therefore,

$$E(S_n | \mathcal{F}_m) = S_m + (n - m)\mu. \quad (5.8)$$

Example 2. Suppose X_1, X_2, \dots and S_n are as in Example 1. Suppose $\mu = 0$ and $\text{Var}(X_i) = \mathbb{E}(X_i^2) = \sigma^2$. Let $m < n$. Then, by (5.3),

$$\begin{aligned} E(S_n^2 | \mathcal{F}_m) &= E([S_m + (S_n - S_m)]^2 | \mathcal{F}_m) \\ &= E(S_m^2 | \mathcal{F}_m) + 2E(S_m(S_n - S_m) | \mathcal{F}_m) \\ &\quad + E((S_n - S_m)^2 | \mathcal{F}_m). \end{aligned}$$

Since S_m depends only on X_1, \dots, X_m and $S_n - S_m$ is independent of X_1, \dots, X_m , we have as in the previous example

$$E(S_m^2 | \mathcal{F}_m) = S_m^2,$$

$$E((S_n - S_m)^2 | \mathcal{F}_m) = \mathbb{E}((S_n - S_m)^2) = \text{Var}(S_n - S_m) = (n - m)\sigma^2.$$

Finally, by (5.7),

$$E(S_m(S_n - S_m) | \mathcal{F}_m) = S_m E(S_n - S_m | \mathcal{F}_m) = S_m \mathbb{E}(S_n - S_m) = 0.$$

Therefore,

$$E(S_n^2 | \mathcal{F}_m) = S_m^2 + (n - m)\sigma^2.$$

Example 3. Consider a special case of Example 1 where the random variable X_i has a Bernoulli distribution, $\mathbb{P}\{X_i = 1\} = p$, $\mathbb{P}\{X_i = 0\} = 1 - p$. Again assume that $m < n$. For any $i \leq m$, consider $E(X_i | S_n)$. If $S_n = k$ then there are k 1s in the first n trials. Given $S_n = k$ it is an easy exercise in conditional probability to show that $\mathbb{P}\{X_i = 1 | S_n = k\} = k/n$. Hence,

$$E(X_i | S_n) = \frac{S_n}{n},$$

and

$$E(S_m | S_n) = E(X_1 | S_n) + \dots + E(X_m | S_n) = \frac{m}{n} S_n.$$

We will need to consider conditional expectations with respect to an infinite collection of random variables,

$$E(Y | X_\alpha, \alpha \in \mathcal{A}).$$

Let \mathcal{F} denote the information in $\{X_\alpha\}$. A random variable Z is \mathcal{F} -measurable if knowledge of all the $\{X_\alpha\}$ determines Z . Essentially, Z is \mathcal{F} -measurable if $Z = \phi(X_{\alpha_1}, \dots, X_{\alpha_n})$ for some function ϕ and some finite subcollection $X_{\alpha_1}, \dots, X_{\alpha_n}$, or if Z is a limit of such random variables. As an example, suppose Y, X_1, X_2, \dots are independent random variables with X_1, X_2, \dots normal mean zero, variance one and Y having some unknown nontrivial distribution. Let

$$Z_j = X_j + Y.$$

Let \mathcal{F}_n denote the information in Z_1, \dots, Z_n and let \mathcal{F}_∞ denote the information in Z_1, Z_2, \dots . One cannot determine the value of Y given Z_1, \dots, Z_n , so Y is not \mathcal{F}_n -measurable. However, Y is \mathcal{F}_∞ -measurable since the law of large numbers implies

$$Y = \lim_{n \rightarrow \infty} \frac{Z_1 + \dots + Z_n}{n}.$$

If \mathcal{F} denotes the information contained in $\{X_\alpha\}$, we will say that an event A is \mathcal{F} -measurable if one can determine whether or not the event has occurred if one knows the values of $\{X_\alpha\}$. This is equivalent to saying that the indicator function I_A is an \mathcal{F} -measurable random variable. The conditional expectation $E(Y | \mathcal{F})$ is defined to be the unique \mathcal{F} -measurable random variable Z such that for every \mathcal{F} -measurable event A ,

$$\mathbb{E}(Y I_A) = \mathbb{E}(Z I_A).$$

All of the properties (5.2) through (5.7) hold for this more general conditional expectation.

5.2 Definition and Examples

A martingale is a model of a fair game. We will let $\{\mathcal{F}_n\}$ denote an increasing collection of information. By this we mean for each n , we have a collection of random variables \mathcal{A}_n such that $\mathcal{A}_m \subset \mathcal{A}_n$ if $m < n$. The information that we have at time n is the value of all of the variables in \mathcal{A}_n . The assumption $\mathcal{A}_m \subset \mathcal{A}_n$ means that we do not lose information. A random variable X is \mathcal{F}_n -measurable if we can determine the value of X if we know the value of all the random variables in \mathcal{A}_n . The increasing sequence of information \mathcal{F}_n is often called a *filtration*.

We say that a sequence of random variables M_0, M_1, M_2, \dots with $\mathbb{E}(|M_i|) < \infty$ is a *martingale* with respect to $\{\mathcal{F}_n\}$ if each M_n is measurable with respect to \mathcal{F}_n , and for each $m < n$,

$$\mathbb{E}(M_n | \mathcal{F}_m) = M_m, \tag{5.9}$$

or equivalently,

$$E(M_n - M_m \mid \mathcal{F}_m) = 0.$$

The condition $\mathbb{E}(|M_i|) < \infty$ is needed to guarantee that the conditional expectations are well defined. If \mathcal{F}_n is the information in random variables X_1, \dots, X_n , then we will also say that M_0, M_1, \dots is a martingale with respect to X_0, X_1, \dots . Sometimes we will just say M_0, M_1, \dots is a martingale without making reference to the filtration \mathcal{F}_n . In this case it will mean that the sequence M_n is a martingale with respect to itself (in which case the first condition is trivially true). In order to verify (5.9) it suffices to prove that for all n ,

$$E(M_{n+1} \mid \mathcal{F}_n) = M_n,$$

since if this holds, by (5.5),

$$\begin{aligned} E(M_{n+2} \mid \mathcal{F}_n) &= E(E(M_{n+2} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) \\ &= E(M_{n+1} \mid \mathcal{F}_n) = M_n, \end{aligned}$$

and so on.

Example 1. Let X_1, X_2, \dots be independent random variables each with mean μ . Let $S_0 = 0$ and for $n > 0$ let S_n be the partial sum

$$S_n = X_1 + \cdots + X_n.$$

Then $M_n = S_n - n\mu$ is a martingale with respect to \mathcal{F}_n , the information contained in X_0, \dots, X_n . This can easily be checked by using Example 1 of Section 5.1,

$$\begin{aligned} E(M_{n+1} \mid \mathcal{F}_n) &= E(S_{n+1} - (n+1)\mu \mid \mathcal{F}_n) = E(S_{n+1} \mid \mathcal{F}_n) - (n+1)\mu \\ &= (S_n + \mu) - (n+1)\mu = M_n. \end{aligned}$$

In particular, if $\mu = 0$, then S_n is a martingale with respect to \mathcal{F}_n .

Example 2. The following is a version of the “martingale betting strategy” which is a way to beat a fair game. Suppose X_1, X_2, \dots are independent random variables with

$$\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = \frac{1}{2}.$$

We can think of the random variables X_i as the results of a game where one flips a coin and wins \$1 if it comes up heads and loses \$1 if it comes up tails. One way to beat the game is to keep doubling our bet until we eventually win. At this point we stop. Let W_n denote the winnings (or losses) up through

n flips of the coin using this strategy. $W_0 = 0$. Whenever we win we stop playing, so our winnings stop changing and

$$\mathbb{P}\{W_{n+1} = 1 \mid W_n = 1\} = 1.$$

Now suppose the first n flips of the coin have turned up tails. After each flip we have doubled our bet, so we have lost $1 + 2 + 4 + \dots + 2^{n-1} = 2^n - 1$ dollars and $W_n = -(2^n - 1)$. At this time we double our bet again and wager 2^n on the next flip. This gives

$$\mathbb{P}\{W_{n+1} = 1 \mid W_n = -(2^n - 1)\} = \frac{1}{2},$$

$$\mathbb{P}\{W_{n+1} = -(2^{n+1} - 1) \mid W_n = -(2^n - 1)\} = \frac{1}{2}.$$

It is then easy to verify that

$$E(W_{n+1} \mid \mathcal{F}_n) = W_n,$$

and hence W_n is a martingale with respect to \mathcal{F}_n .

Example 3. We can generalize the previous example. Suppose X_1, X_2, \dots are as in Example 2. Suppose that on the n th flip we make a bet equal to B_n . In determining the amount of the bet, we may look at the results of the first $(n-1)$ st flips but cannot look beyond that. In other words, B_n is a random variable measurable with respect to \mathcal{F}_{n-1} (we assume that B_1 is just a constant). The winnings after n flips, W_n , are given by $W_0 = 0$ and

$$W_n = \sum_{j=1}^n B_j X_j.$$

We allow B_n to be negative; this corresponds to betting that the coin will come up tails. Assume that $\mathbb{E}(|B_n|) < \infty$ (which will be guaranteed if the bet at time n is required to be less than some constant C_n). Then W_n is a martingale with respect to \mathcal{F}_n . To see this, we first note that $\mathbb{E}(|W_n|) < \infty$ follows from the fact that $\mathbb{E}(|B_n|) < \infty$ for each n . It is clear that W_n is \mathcal{F}_n -measurable. Finally,

$$\begin{aligned} E(W_{n+1} \mid \mathcal{F}_n) &= E\left(\sum_{j=1}^{n+1} B_j X_j \mid \mathcal{F}_n\right) \\ &= E\left(\sum_{j=1}^n B_j X_j \mid \mathcal{F}_n\right) + E(B_{n+1} X_{n+1} \mid \mathcal{F}_n). \end{aligned}$$

By (5.4),

$$E\left(\sum_{j=1}^n B_j X_j \mid \mathcal{F}_n\right) = \sum_{j=1}^n B_j X_j = W_n.$$

Since B_{n+1} is \mathcal{F}_n -measurable, it follows from (5.7) and (5.6) that

$$E(B_{n+1}X_{n+1} \mid \mathcal{F}_n) = B_{n+1}E(X_{n+1} \mid \mathcal{F}_n) = B_{n+1}\mathbb{E}(X_{n+1}) = 0.$$

Therefore,

$$E(W_{n+1} \mid \mathcal{F}_n) = W_n.$$

Example 4. Polya's Urn. Consider an urn with balls of two colors, red and green. Assume that initially there is one ball of each color in the urn. At each time step, a ball is chosen at random from the urn. If a red ball is chosen, it is returned and in addition another red ball is added to the urn. Similarly, if a green ball is chosen, it is returned together with another green ball. Let X_n denote the number of red balls in the urn after n draws. Then $X_0 = 1$ and X_n is a (time-inhomogeneous) Markov chain with transitions

$$\mathbb{P}\{X_{n+1} = k+1 \mid X_n = k\} = \frac{k}{n+2},$$

$$\mathbb{P}\{X_{n+1} = k \mid X_n = k\} = \frac{n+2-k}{n+2}.$$

Let $M_n = X_n/(n+2)$ be the fraction of red balls after n draws. Then M_n is a martingale. To see this, note that

$$E(X_{n+1} \mid X_n) = X_n + \frac{X_n}{n+2}.$$

Since this is a Markov chain, all the relevant information in \mathcal{F}_n for determining X_{n+1} is contained in X_n . Therefore,

$$\begin{aligned} E(M_{n+1} \mid \mathcal{F}_n) &= E((n+3)^{-1}X_{n+1} \mid X_n) \\ &= \frac{1}{n+3} \left[X_n + \frac{X_n}{n+2} \right] \\ &= \frac{X_n}{n+2} = M_n. \end{aligned}$$

A process M_n with $\mathbb{E}(|M_n|) < \infty$ is called a *submartingale* (*supermartingale*) with respect to $\{\mathcal{F}_n\}$ if for each $m < n$, $E(M_n \mid \mathcal{F}_m) \geq (\leq) M_m$. In other words, a submartingale is a game in one's favor and a supermartingale is an unfair game. Note that M_n is a martingale if and only if it is both a submartingale and a supermartingale.

Example 5. Let X_n be a Markov chain with finite state space. Suppose a payoff function f is given as in Chapter 4. Let v be the value function associated to the payoff functions, $v(x) = E(f(X_T) \mid X_0 = x)$, where T is the stopping rule associated with the optimal strategy. Then $M_n = v(X_n)$ is a supermartingale with respect to X_0, X_1, \dots .

5.3 Optional Sampling Theorem

The optional sampling theorem states in effect, “You cannot beat a fair game.” However, it is easy to see that this statement is false in complete generality. For example, suppose one plays the fair game of flipping a coin, winning one’s bet if the coin comes up heads and losing one’s bet if it is tails. Then using the “martingale betting strategy” described in Example 2 of Section 5.2, one can guarantee that one finishes the game ahead.

A stopping time T with respect to a filtration $\{\mathcal{F}_n\}$ is a random variable taking values in the nonnegative integers (we allow $T = \infty$ as a possible value) that gives the time at which some procedure is stopped ($T = \infty$ corresponds to never stopping), such that the decision whether to stop at time n must be made using only the information available at time n . More precisely, we say that T is a stopping time (with respect to $\{\mathcal{F}_n\}$) if for each n , the event $\{T = n\}$ is measurable with respect to \mathcal{F}_n .

Example 1. Let k be an integer and let $T = k$.

Example 2. Let A be a set and let $T_A = \min\{j : X_j \in A\}$.

Example 3. If T and U are stopping times, then so are $\min\{T, U\}$ and $\max\{T, U\}$. In particular, if T is a stopping time and $T_n = \min\{T, n\}$, then each T_n is a stopping time, $T_0 \leq T_1 \leq T_2 \leq \dots$, and $T_n \leq n$.

The optional sampling (or optional stopping) theorem states that (under certain conditions) if M_n is a martingale and T is a stopping time then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$. This will not hold under all conditions since if we consider the martingale betting strategy and let T be the first time that the coin comes up heads, then $1 = \mathbb{E}(M_T) \neq \mathbb{E}(M_0) = 0$. The first thing we would like to show is that there is no way to beat a fair game if one has only a finite amount of time.

Fact. Suppose M_0, M_1, \dots is a martingale with respect to $\{\mathcal{F}_n\}$ and suppose T is a stopping time. Suppose that T is bounded, $T \leq K$. Then

$$\mathbb{E}(M_T | \mathcal{F}_0) = M_0.$$

In particular, $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

To prove this fact, we first note that the event $\{T > n\}$ is measurable with respect to \mathcal{F}_n (since we need only the information up through time n to determine if we have stopped by time n). Since M_T is the random variable which equals M_j if $T = j$ we can write

$$M_T = \sum_{j=0}^K M_j I\{T = j\}.$$

Let us take the conditional expectation with respect to \mathcal{F}_{K-1} ,

$$E(M_T | \mathcal{F}_{K-1}) = E(M_K I\{T = K\} | \mathcal{F}_{K-1})$$

$$+ \sum_{j=0}^{K-1} E(M_j I\{T = j\} | \mathcal{F}_{K-1}).$$

For $j \leq K-1$, $M_j I\{T = j\}$ is \mathcal{F}_{K-1} -measurable; hence

$$E(M_j I\{T = j\} | \mathcal{F}_{K-1}) = M_j I\{T = j\}.$$

Since T is known to be no more than K , the event $\{T = K\}$ is the same as the event $\{T > K-1\}$. The latter event is measurable with respect to \mathcal{F}_{K-1} . Hence, using (5.7),

$$\begin{aligned} E(M_K I\{T = K\} | \mathcal{F}_{K-1}) &= E(M_K I\{T > K-1\} | \mathcal{F}_{K-1}) \\ &= I\{T > K-1\} E(M_K | \mathcal{F}_{K-1}) \\ &= I\{T > K-1\} M_{K-1}. \end{aligned}$$

The last equality follows from the fact the M_n is a martingale. Therefore,

$$\begin{aligned} E(M_T | \mathcal{F}_{K-1}) &= I\{T > K-1\} M_{K-1} + \sum_{j=0}^{K-1} M_j I\{T = j\} \\ &= I\{T > K-2\} M_{K-1} + \sum_{j=0}^{K-2} M_j I\{T = j\}. \end{aligned}$$

If we work through this argument again, this time conditioning with respect to \mathcal{F}_{K-2} , we get

$$\begin{aligned} E(M_T | \mathcal{F}_{K-2}) &= E(E(M_T | \mathcal{F}_{K-1}) | \mathcal{F}_{K-2}) \\ &= I\{T > K-3\} M_{K-2} + \sum_{j=0}^{K-3} M_j I\{T = j\}. \end{aligned}$$

We can continue this process until we get

$$E(M_T | \mathcal{F}_0) = M_0.$$

There are many examples of interest where the stopping time T is not bounded. Suppose T is a stopping time with $\mathbb{P}\{T < \infty\} = 1$, i.e., a rule that guarantees that one stops eventually. (Note that the time associated to the martingale betting strategy satisfies this condition.) When can we

conclude that $\mathbb{E}(M_T) = \mathbb{E}(M_0)$? To investigate this consider the stopping times $T_n = \min\{T, n\}$. Note that

$$M_T = M_{T_n} + M_T I\{T > n\} - M_n I\{T > n\}.$$

Hence,

$$\mathbb{E}(M_T) = \mathbb{E}(M_{T_n}) + \mathbb{E}(M_T I\{T > n\}) - \mathbb{E}(M_n I\{T > n\}).$$

Since T_n is a bounded stopping time, it follows from the above that $\mathbb{E}(M_{T_n}) = \mathbb{E}(M_0)$. We would like to be able to say that the other two terms do not contribute as $n \rightarrow \infty$. The second term is not much of a problem. Since the probability of the event $\{T > n\}$ goes to 0 as $n \rightarrow \infty$, we are taking the expectation of the random variable M_T restricted to a smaller and smaller set. One can show (see Section 5.4) that if $\mathbb{E}(|M_T|) < \infty$ then $\mathbb{E}(|M_T|I\{T > n\}) \rightarrow 0$.

The third term is more troublesome. Consider Example 2 of Section 5.2 again. In this example, the event $\{T > n\}$ is the event that the first n flips are tails and has probability 2^{-n} . If this event occurs, the bettor has lost a total of $2^n - 1$ dollars, i.e., $M_n = 1 - 2^n$. Hence

$$\mathbb{E}(M_n I\{T > n\}) = 2^{-n}(1 - 2^n),$$

which does not go to 0 as $n \rightarrow \infty$. This is why the desired result fails in this case. However, if M_n and T are given satisfying

$$\lim_{n \rightarrow \infty} \mathbb{E}(|M_n| I\{T > n\}) = 0,$$

then we will be able to conclude that $\mathbb{E}(M_T) = \mathbb{E}(M_0)$. We summarize this as follows.

Optional Sampling Theorem. *Suppose M_0, M_1, \dots is a martingale with respect to $\{\mathcal{F}_n\}$ and T is a stopping time satisfying $\mathbb{P}\{T < \infty\} = 1$,*

$$\mathbb{E}(|M_T|) < \infty, \tag{5.10}$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E}(|M_n| I\{T > n\}) = 0. \tag{5.11}$$

Then, $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

Example 1. Let X_n be a simple random walk ($p = 1/2$) on $\{0, \dots, N\}$ with absorbing boundaries. Suppose $X_0 = a$. Then, X_n is a martingale. Let $T = \min\{j : X_j = 0 \text{ or } N\}$. T is a stopping time, and since X_n is bounded, (5.10) and (5.11) are satisfied [note that (5.10) and (5.11) are always satisfied if the martingale is bounded and $\mathbb{P}\{T < \infty\} = 1$]. Therefore

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = a.$$

But in this case $\mathbb{E}(M_T) = N \mathbb{P}\{X_T = N\}$. Therefore,

$$\mathbb{P}\{X_T = N\} = \frac{a}{N}.$$

This gives another derivation of the gambler's ruin result for simple random walk.

Example 2. Let X_n be as in Example 1 and let $M_n = X_n^2 - n$. Then M_n is a martingale with respect to X_n . To see this, note that by Example 2, Section 5.1,

$$E(M_{n+1} | \mathcal{F}_n) = E(X_{n+1}^2 - (n+1) | \mathcal{F}_n) = X_n^2 + 1 - (n+1) = M_n.$$

Again, let $T = \min\{j : X_j = 0 \text{ or } N\}$. In this case, M_n is not a bounded martingale so it is not immediate that (5.10) and (5.11) hold. However, one can prove (Exercise 1.7) that there exist $C < \infty$ and $\rho < 1$ such that

$$\mathbb{P}\{T > n\} \leq C\rho^n.$$

Since $|M_n| \leq N^2 + n$, one can then show that $\mathbb{E}(|M_T|) < \infty$ and

$$\mathbb{E}(|M_n| I\{T > n\}) \leq C\rho^n(N^2 + n) \rightarrow 0.$$

Hence the conditions of the optional sampling theorem hold and we can conclude

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = a^2.$$

Note that

$$\mathbb{E}(M_T) = \mathbb{E}(X_T^2) - \mathbb{E}(T) = N^2 \mathbb{P}\{X_T = N\} - \mathbb{E}(T) = aN - \mathbb{E}(T).$$

Hence,

$$\mathbb{E}(T) = aN - a^2 = a(N - a).$$

Example 3. Let X_n be a simple random walk ($p = 1/2$) on the integers $\{\dots, -1, 0, 1, \dots\}$ with $X_0 = 0$. We have seen that this is a martingale. Let $T = \min\{j : X_j = 1\}$. Since simple random walk is recurrent, $\mathbb{P}\{T < \infty\} = 1$. Note that $X_T = 1$ and hence

$$1 = \mathbb{E}(X_T) \neq \mathbb{E}(X_0) = 0.$$

Therefore, the conditions of the optional sampling theorem must not hold. We will not give the details here but it can be shown in this case that $\mathbb{P}\{T > n\} \sim cn^{-1/2}$ for some constant c . By the central limit theorem, the random walk tends to go a distance of order \sqrt{n} in n steps. In this case $\mathbb{E}(|X_n| I\{T > n\})$ does not go to 0.

Example 4. Example 1 can be extended to general Markov chains. Let \mathbf{P} be the transition matrix for the irreducible Markov chain X_n on the finite state space S . Let A be a subset of S and let F be a function from A to \mathbb{R} . Then we claim that there is a unique function f on S satisfying

$$f(x) = F(x), \quad x \in A,$$

$$\mathbf{P}f(x) := \sum_{y \in A} p(x, y) f(y) = 0, \quad x \in S \setminus A.$$

This is not surprising if one realizes that the last equation gives k equations in k unknowns where k is the number of elements in $S \setminus A$. Suppose f satisfies these conditions. Let $T = \min\{n \geq 0 : X_n \in A\}$ and $T_n = \min\{n, T\}$. Let $M_n = f(X_{T_n})$. Then M_n is a bounded martingale, and hence,

$$f(x) = \mathbb{E}[M_0 | X_0 = x] = \mathbb{E}[M_T | X_0 = x] = \mathbb{E}[f(X_T) | X_0 = x].$$

5.4 Uniform Integrability

Condition (5.11) is often hard to verify. For this reason we would like to give conditions that may be easier to check from which we can conclude (5.11). We start by considering one random variable X with $\mathbb{E}(|X|) < \infty$. Let F denote the distribution function for $|X|$. Then it follows that

$$\lim_{K \rightarrow \infty} \mathbb{E}(|X| I\{|X| > K\}) = \lim_{K \rightarrow \infty} \int_K^\infty |x| dF(x) = 0.$$

Now suppose we have a sequence of random variables X_1, X_2, \dots . We say that the sequence is *uniformly integrable* if for every $\epsilon > 0$ there exists a K such that for each n ,

$$\mathbb{E}[|X_n| I\{|X_n| > K\}] < \epsilon.$$

It is important that K does not depend on n . If X_1, X_2, \dots are uniformly integrable, then the following holds: for every $\epsilon > 0$, there is a $\delta > 0$ such that if $\mathbb{P}(A) < \delta$, then

$$\mathbb{E}(|X_n| I_A) < \epsilon \tag{5.12}$$

for each n . Again, δ may not depend on n and (5.12) must hold for all values of n . To see that uniform integrability implies this, let $\epsilon > 0$ and choose K sufficiently large so that $\mathbb{E}[|X_n| I\{|X_n| > K\}] < \epsilon/2$ for all n . If we let $\delta = \epsilon/(2K)$, then if $\mathbb{P}(A) < \delta$,

$$\begin{aligned} \mathbb{E}(|X_n| I_A) &\leq \mathbb{E}(|X_n| I_A; |X_n| \leq K) + \mathbb{E}(|X_n|; |X_n| > K) \\ &< K\mathbb{P}(A) + (\epsilon/2) < \epsilon. \end{aligned}$$

To develop a feeling for the definition, we will start by giving an example of random variables that are not uniformly integrable. Consider Example 2 of Section 5.2, the martingale betting strategy, and consider the random variables W_0, W_1, W_2, \dots . If we let A_n be the event $\{X_1 = X_2 = \dots = X_n = -1\}$ then $\mathbb{P}(A_n) = 2^{-n}$ and $\mathbb{E}(|W_n|I_{A_n}) = 2^{-n}(2^n - 1) \rightarrow 1$. We clearly cannot satisfy the conditions for uniform integrability for any $\epsilon < 1$.

Now suppose that M_0, M_1, \dots is a uniformly integrable martingale with respect to X_0, X_1, \dots and T is a stopping time with $\mathbb{P}\{T < \infty\} = 1$. Then

$$\lim_{n \rightarrow \infty} \mathbb{P}\{T > n\} = 0,$$

and hence by uniformly integrability

$$\lim_{n \rightarrow \infty} \mathbb{E}(|M_n|I\{T > n\}) = 0;$$

that is, (5.11) holds. We can therefore give another statement of the optional sampling theorem.

Optional Sampling Theorem. *Suppose M_0, M_1, \dots is a uniformly integrable martingale with respect to $\{\mathcal{F}_n\}$ and T is a stopping time satisfying $\mathbb{P}\{T < \infty\} = 1$ and $\mathbb{E}(|M_T|) < \infty$. Then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.*

The condition of uniform integrability can be difficult to verify. There are a number of easier conditions that imply uniform integrability. We mention one here and give another in the exercises (Exercise 5.15).

Fact. *If X_1, X_2, \dots is a sequence of random variables and there exists a $C < \infty$ such that $\mathbb{E}(X_n^2) < C$ for each n , then the sequence is uniformly integrable.*

To prove the fact, let $\epsilon > 0$ be given and let $\delta = \epsilon^2/4C$. Suppose $\mathbb{P}(A) < \delta$. Then

$$\begin{aligned} \mathbb{E}(|X_n|I_A) &= \mathbb{E}[|X_n|I(A \cap \{|X_n| \geq 2C/\epsilon\})] \\ &\quad + \mathbb{E}[|X_n|I(A \cap \{|X_n| < 2C/\epsilon\})] \\ &\leq (\epsilon/2C) \mathbb{E}[|X_n|^2 I(A \cap \{|X_n| \geq 2C/\epsilon\})] \\ &\quad + (2C/\epsilon) \mathbb{P}(A \cap \{|X_n| < 2C/\epsilon\}) \\ &\leq (\epsilon/2C) \mathbb{E}(|X_n|^2) + (2C/\epsilon) \mathbb{P}(A) < \epsilon. \end{aligned}$$

Example 1. Random Harmonic Series. It is well known that the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \dots$ diverges while the alternating harmonic series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$ converges. What if the pluses and minuses are chosen at random? To study this, let X_1, X_2, \dots be independent random variables with $\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = 1/2$. Let $M_0 = 0$ and for $n > 0$,

$$M_n = \sum_{j=1}^n \frac{1}{j} X_j.$$

By Example 1, Section 5.2, M_n is a martingale. Since $\mathbb{E}(M_n) = 0$,

$$\mathbb{E}(M_n^2) = \text{Var}(M_n^2) = \sum_{j=1}^n \text{Var}\left(\frac{1}{j}X_j\right) = \sum_{j=1}^n \frac{1}{j^2} \leq \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty.$$

Hence M_n is a uniformly integrable martingale. The question of convergence is discussed in the next section.

Example 2. Branching Process. Let X_n denote the number of offspring in the n th generation of a branching process (see Section 2.4) whose offspring distribution has mean μ and variance σ^2 . Then (Exercise 5.5) $M_n = \mu^{-n}X_n$ is a martingale with respect to X_1, X_2, \dots . Suppose $\mu > 1$. Then (Exercise 5.11) there exists a constant such that for all n , $\mathbb{E}(M_n^2) < \infty$ and hence M_n is a uniformly integrable martingale for $\mu > 1$.

5.5 Martingale Convergence Theorem

The martingale convergence theorem states that under very general conditions a martingale M_n converges to a limiting random variable M_∞ . We start by considering a particular example, Polya's urn (Example 4, Section 5.2). In this case M_n is the proportion of red balls in the urn after n draws. What happens as n gets large? In Exercise 5.12 it is shown that the distribution of M_n is approximately a uniform distribution on $[0, 1]$ for large values of n . This leads to a question: Does the proportion of red balls fluctuate between 0 and 1 infinitely often or does it eventually settle down to a particular value? We will show now that the latter is true.

Let $0 < a < b < \infty$ and suppose that $M_n < a$. Let T be the stopping time

$$T = \min\{j : j \geq n \text{ and } M_j \geq b\},$$

and let $T_m = \min\{T, m\}$. Then for $m > n$, the optional sampling theorem states that

$$\mathbb{E}(M_{T_m}) = M_n < a.$$

But

$$\mathbb{E}(M_{T_m}) \geq \mathbb{E}(M_{T_m} I\{T \leq m\}) = \mathbb{E}(M_T I\{T \leq m\}) \geq b \mathbb{P}\{T \leq m\}.$$

Hence,

$$\mathbb{P}\{T \leq m\} < \frac{a}{b}.$$

Since this is true for all m ,

$$\mathbb{P}\{T < \infty\} \leq \frac{a}{b}.$$

This says that with probability of at least $1 - (a/b)$ the proportion of red balls never gets as high as b . Now suppose the proportion of red balls does get higher than b . What then is the probability that the proportion goes down below a again? By the same argument applied to the proportion of green balls we can say that the probability of dropping below a is at most $(1-b)/(1-a)$. By continuing this argument, we can see that, starting at a , the probability of going above b , then below a again, then above b , then below a , a total of n times, can be bounded above by

$$\left(\frac{a}{b}\right) \left(\frac{1-b}{1-a}\right) \left(\frac{a}{b}\right) \left(\frac{1-b}{1-a}\right) \cdots \left(\frac{a}{b}\right) \left(\frac{1-b}{1-a}\right) = \left(\frac{a}{b}\right)^n \left(\frac{1-b}{1-a}\right)^n,$$

which tends to 0 as $n \rightarrow \infty$. Hence, the proportion does not fluctuate infinitely often between a and b . Since a and b are arbitrary, this shows that it is impossible for the proportion to fluctuate infinitely often between any two numbers, or, in other words, the limit

$$M_\infty = \lim_{n \rightarrow \infty} M_n$$

exists. The limit M_∞ is a random variable; it is not difficult to show (see Exercise 5.12) that M_∞ has a uniform distribution on $[0, 1]$.

We now state a general result.

Martingale Convergence Theorem. *Suppose M_0, M_1, \dots is a martingale with respect to $\{\mathcal{F}_n\}$ such that there exists a $C < \infty$ with $\mathbb{E}(|M_n|) < C$ for all n . Then there exists a random variable M_∞ such that*

$$M_n \longrightarrow M_\infty.$$

Note that the limiting random variable M_∞ is measurable with respect to M_0, M_1, \dots . The proof of the theorem is similar to the argument above. What we will show is that for every $0 < a < b < \infty$ the probability that the martingale fluctuates infinitely often between a and b is 0. Since this will be true for every $a < b$, it must be the case that the martingale M_n converges to some value M_∞ .

Fix $a < b$. We will consider the following betting strategy which is reminiscent of the martingale betting strategy. We think of M_n as giving the cumulative results of some fair game and $M_{n+1} - M_n$ as being the result of the game at time $n + 1$. Whenever $M_n < a$, bet 1 on the martingale. Continue this procedure until the martingale gets above b . Then stop betting until the martingale drops below a again and return to betting 1. Continue

this process, changing the bet to 0 when M_n goes above b and changing back to 1 when M_n drops below a . Note that if the martingale fluctuates infinitely often between a and b this gives a strategy that produces a long-term gain from the fair game.

After n steps the winnings in this strategy are given by

$$W_n = \sum_{j=1}^n B_j (M_j - M_{j-1}),$$

where B_j is the bet which equals 1 or 0 depending on whether the martingale was most recently below a or above b . One can verify as in Example 3, Section 5.2 that W_n is a martingale with respect to M_0, M_1, \dots . We note that

$$W_n \geq (b - a) U_n - |M_n - a|,$$

where U_n denotes the number of times that the martingale goes between a and b (this is often called the number of *upcrossings*) and $|M_n - a|$ gives an estimate for the amount lost in the last interval (this is relevant if the bettor is betting 1 at time n). Since W_n is a martingale we have

$$\mathbb{E}(W_0) = \mathbb{E}(W_n) \geq (b - a) \mathbb{E}(U_n) - \mathbb{E}(|M_n - a|).$$

Since $\mathbb{E}(|M_n - a|) \leq \mathbb{E}(|M_n|) + a \leq C + a$, we get

$$\mathbb{E}(U_n) \leq \frac{\mathbb{E}(W_0) + C + a}{b - a}.$$

Since this holds for every n , the expected number of upcrossings up to infinity is bounded and hence with probability one the number of upcrossings is finite. This proves the theorem.

The martingale property implies that for every n , $\mathbb{E}(M_n) = \mathbb{E}(M_0)$. It is not necessarily true, however, that $\mathbb{E}(M_\infty) = \mathbb{E}(M_0)$. For a counterexample, we return to the martingale betting strategy. In this case

$$W_\infty = \lim_{n \rightarrow \infty} W_n = 1,$$

and hence $\mathbb{E}(W_\infty) \neq \mathbb{E}(W_0) = 0$. If the martingale is uniformly integrable, it is true that the limiting random variable has the same expectation (see Exercise 5.13).

Fact. *If M_n is a uniformly integrable martingale with respect to X_0, X_1, \dots , then*

$$M_\infty = \lim_{n \rightarrow \infty} M_n,$$

exists and $\mathbb{E}(M_\infty) = \mathbb{E}(M_0)$.

Example 1. Let X_n be the number of individuals in the n th generation of a branching process whose offspring distribution has mean μ and variance σ^2 . Assume $X_0 = 1$ and let $M_n = \mu^{-n} X_n$ be the associated martingale. If $\mu \leq 1$, we know that extinction occurs with probability one and hence $M_n \rightarrow M_\infty = 0$. In this case $\mathbb{E}(M_\infty) \neq \mathbb{E}(M_0)$. In Section 5.4, we noted that M_n is uniformly integrable if $\mu > 1$, and hence M_∞ is a nontrivial random variable with $\mathbb{E}(M_\infty) = 1$.

Example 2. Let X_1, X_2, \dots be independent random variables with $\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = 1/2$ and let M_n be the random harmonic series

$$M_n = \sum_{j=1}^n \frac{1}{j} X_j.$$

It was noted in Section 5.4 that M_n is a uniformly integrable martingale. Hence M_n approaches a random variable M_∞ . This says that the random harmonic series converges with probability one.

Example 3. Let M_n be the proportion of red balls in Polya's urn. In this case, suppose that at time $n = 0$ there are k red balls and m green balls (so after n draws there are $n + k + m$ balls). Since M_n is bounded it is easy to see that M_n is a uniformly integrable martingale and M_n approaches a random variable M_∞ with $\mathbb{E}(M_\infty) = \mathbb{E}(M_0) = k/(k+m)$. It can be shown (see Example 7 below) that the distribution of M_∞ is a beta distribution with parameters k and m , i.e., it has density

$$\frac{(k+m-1)!}{(k-1)!(m-1)!} x^{k-1} (1-x)^{m-1}, \quad 0 < x < 1.$$

Example 4. Let M_n be a martingale with respect to X_0, X_1, \dots , and let T be a stopping time with $\mathbb{P}\{T < \infty\} = 1$. Let $T_n = \min\{n, T\}$ and $Y_n = M_{T_n}$. Then $Y_n \rightarrow Y_\infty$ where $Y_\infty = M_T$. As we saw in the optional sampling theorem, it is not always the case that $\mathbb{E}(Y_\infty) = \mathbb{E}(Y_0)$. However, this is true if M_n is uniformly integrable.

Example 5. Let X_n be an irreducible Markov chain on a countably infinite state space S with transition function $p(x, y)$. A function f is called *harmonic* at x if

$$f(x) = \sum_{y \in S} p(x, y) f(y).$$

In Chapter 2 we considered the problem of determining whether or not the chain was recurrent. We now prove one of the assertions we made there. Suppose z is a fixed state in S and let $u(x)$ denote the probability starting at state x that the chain ever reaches state z . In other words, if

$$T = \min\{j \geq 0 : X_j = z\},$$

then

$$u(x) = \mathbb{P}\{T < \infty \mid X_0 = x\}.$$

As we noted then, $u(z) = 1$ and $u(x)$ is harmonic at any $x \neq z$.

Suppose now that we can find some function v that satisfies:

$$v(z) = 1, \quad (5.13)$$

$$0 \leq v(x) \leq 1, \quad (5.14)$$

$$v(x) = \sum_{y \in S} p(x, y) v(y), \quad x \neq z. \quad (5.15)$$

If T is defined as above, and $T_n = \min\{n, T\}$, one can check that $M_n = v(X_{T_n})$ is a martingale with respect to X_0, X_1, \dots . Since v is bounded, M_n is uniformly integrable and

$$\lim_{n \rightarrow \infty} M_n = M_\infty,$$

exists with $\mathbb{E}(M_\infty) = \mathbb{E}(M_0)$.

If the chain is recurrent, then $\mathbb{P}\{T < \infty\} = 1$ and $M_\infty = v(z) = 1$. Hence if $X_0 = x$, $1 = \mathbb{E}(M_0) = v(x)$. Thus, if the chain is recurrent there is no nontrivial solution to equations (5.13) through (5.15).

Example 6. Let X_1, X_2, \dots be independent random variables with

$$\mathbb{P}\left\{X_i = \frac{3}{2}\right\} = \mathbb{P}\left\{X_i = \frac{1}{2}\right\} = \frac{1}{2}.$$

Let $M_0 = 1$ and for $n > 0$, let $M_n = X_1 \cdots X_n$. Note that $\mathbb{E}(M_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n) = 1$, and in fact, if \mathcal{F}_n denotes the information contained in X_1, \dots, X_n ,

$$\begin{aligned} E(M_{n+1} \mid \mathcal{F}_n) &= E(X_1 \cdots X_{n+1} \mid \mathcal{F}_n) \\ &= X_1 \cdots X_n E(X_{n+1} \mid \mathcal{F}_n) \\ &= X_1 \cdots X_n \mathbb{E}(X_{n+1}) = M_n. \end{aligned}$$

Hence M_n is a martingale with respect to X_1, X_2, \dots . Since $\mathbb{E}(|M_n|) = \mathbb{E}(M_n) = 1$, the conditions of the martingale convergence theorem hold and hence

$$M_n \rightarrow M_\infty$$

for some random variable M_∞ . Is M_n uniformly integrable? The answer is no; in fact, the limiting random variable $M_\infty = 0$ [and hence $\mathbb{E}(M_\infty) \neq \mathbb{E}(M_0)$]. To see this, consider the logarithm of the martingale,

$$\ln M_n = \sum_{j=1}^n \ln X_j.$$

The right-hand side is the sum of independent identically distributed random variables with mean

$$\mathbb{E}(\ln X_i) = \frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{3}{2} < 0.$$

By the law of large numbers, $\ln M_n \rightarrow -\infty$ and hence $M_n \rightarrow 0$.

Note in this case

$$\mathbb{E}(M_n^2) = \mathbb{E}(X_1^2) \cdots \mathbb{E}(X_n^2) = (5/4)^n,$$

so the second moment is not uniformly bounded.

Example 7. A typical problem in statistics is to estimate the mean θ of a distribution given independent samples

$$Y_1, Y_2, Y_3, \dots$$

from the distribution. In Bayesian statistics, the parameter θ is taken to be a random variable with a certain distribution, called the *prior distribution*. Assume that $\mathbb{E}[\theta] = \mu$ under the prior distribution. Let $M_0 = \mu$ and

$$M_n = E[\theta | Y_1, \dots, Y_n].$$

Then M_n is a martingale. The conditional distribution on M_n given $Y_1 = y_1, \dots, Y_n = y_n$ is called the *posterior distribution*. The martingale convergence theorem tells us that

$$\lim_{n \rightarrow \infty} M_n = M_\infty,$$

for some random variable which depends on the infinite sequence of values $\{Y_1, Y_2, \dots\}$. Moreover, it can be shown that $M_n = E[M_\infty | Y_1, \dots, Y_n]$. The strong law of large numbers tells us that for fixed θ ,

$$\lim_{n \rightarrow \infty} \frac{Y_1 + \dots + Y_n}{n} = \theta.$$

That is, the random variable θ can be determined from the infinite sequence of values. This gives $M_\infty = \theta$,

$$\lim_{n \rightarrow \infty} E[\theta | Y_1, \dots, Y_n] = \theta.$$

As an example assume that Y_1, Y_2, \dots are independent samples from a Bernoulli distribution with $\mathbb{P}\{Y_j = 1\} = 1 - \mathbb{P}\{Y_j = 0\} = \theta$. If we have no a priori knowledge about θ we might assume that θ is a random variable uniformly distributed on $[0, 1]$. For fixed θ ,

$$\mathbb{P}\{Y_1 + \dots + Y_n = k\} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

Let $f_n(\theta | k)$ denote the conditional density on θ given $Y_1 + \dots + Y_k = n$. Bayes theorem shows that

$$f_n(\theta | k) = \frac{\binom{n}{k} \theta^k (1 - \theta)^{n-k}}{\int_0^1 \binom{n}{k} \theta^k (1 - \theta)^{n-k} d\theta} = \frac{(n+1)!}{k!(n-k)!} \theta^k (1 - \theta)^{n-k}.$$

This is called the beta distribution with parameters $k+1$ and $n-k+1$. A straightforward calculation shows that the mean of this distribution is $(k+1)/(n+2)$. Note that

$$\begin{aligned} \mathbb{P}\{Y_{n+1} = k+1 | Y_n = k\} &= \int_0^1 \mathbb{P}\{Y_{n+1} = 1 | \theta\} f_n(\theta | k) d\theta \\ &= \int_0^1 \theta f_n(\theta | k) d\theta = \frac{k+1}{n+2}. \end{aligned}$$

If we let $Y_n + 1$ represent the number of red balls in an urn and $(n - Y_n) + 1$ the number of green balls in the urn, we have exactly the transitions for Polya's urn.

5.6 Maximal Inequalities

If M_0, M_1, M_2, \dots is a sequence of random variables, define the *maximum processes* by

$$\overline{M}_n = \max\{M_0, \dots, M_n\}, \quad M_n^* = \max\{|M_0|, \dots, |M_n|\}.$$

Maximal inequalities relate probabilities or expectations for \overline{M}_n, M_n^* to those for M_n or $|M_n|$. We give two examples here, the *reflection principle* and the *Doob maximal inequality*. The basic ideas of the proofs is the following: if M_n is a martingale or a submartingale and M_j is large for some $j \leq n$, then there is a good chance that M_n will be large as well. Stopping times are used to make these arguments precise.

Reflection Principle. Suppose X_1, X_2, \dots are independent random variables whose distribution is symmetric about the origin. Let $M_0 = 0, M_n = X_1 + \dots + X_n$. Then for every $a > 0$,

$$\mathbb{P}\{\overline{M}_n \geq a\} \leq 2 \mathbb{P}\{M_n \geq a\}.$$

To prove this, let T be the smallest j such that $M_j \geq a$ and note that

$$\mathbb{P}\{\overline{M}_n \geq a\} = \sum_{j=0}^n \mathbb{P}\{T = j\},$$

and

$$\begin{aligned} \mathbb{P}\{M_n \geq a\} &= \sum_{j=0}^n \mathbb{P}\{T = j, M_n \geq a\} \\ &= \sum_{j=0}^n \mathbb{P}\{T = j\} \mathbb{P}\{M_n \geq a \mid T = j\}. \end{aligned}$$

However, independence and symmetry of the distribution of X_1, X_2, \dots show that

$$\begin{aligned} \mathbb{P}\{M_n \geq a \mid T = j\} &\geq \mathbb{P}\{M_n - M_j \geq 0 \mid T = j\} \\ &= \mathbb{P}\{M_n - M_j \geq 0\} \geq \frac{1}{2}. \end{aligned}$$

Doob's Maximal Inequality. Suppose M_0, M_1, M_2, \dots is a nonnegative submartingale with respect to \mathcal{F}_n . Then for every $a > 0$,

$$\mathbb{P}\{\overline{M}_n \geq a\} \leq \frac{\mathbb{E}[M_n]}{a}.$$

This inequality can be considered as a generalization of the inequality

$$\mathbb{P}\{M_n \geq a\} \leq \frac{\mathbb{E}[M_n]}{a}.$$

To prove the maximal inequality, we again let T be the smallest j with $M_j \geq a$ and let A_j denote the \mathcal{F}_j -measurable event $\{T = j\}$. Since M_n is nonnegative we can write

$$\mathbb{E}[M_n] \geq \mathbb{E}[M_n I\{T \leq n\}] = \sum_{j=0}^n \mathbb{E}[M_n I_{A_j}],$$

where I denotes the indicator function. However, since A_j is \mathcal{F}_j -measurable, properties of conditional expectation can be used to see that

$$\begin{aligned} \mathbb{E}[M_n I_{A_j}] &= \mathbb{E}[E(M_n I_{A_j} \mid \mathcal{F}_j)] = \mathbb{E}[E(M_n \mid \mathcal{F}_j) I_{A_j}] \\ &\geq \mathbb{E}[M_j I_{A_j}] \\ &\geq \mathbb{E}[a I_{A_j}] = a \mathbb{P}(A_j). \end{aligned}$$

Therefore,

$$\mathbb{E}[M_n] \geq \sum_{j=0}^n a \mathbb{P}(A_j) = a \mathbb{P}\{\bar{M}_n \geq a\}.$$

If M_0, M_1, \dots is a martingale with respect to \mathcal{F}_n , not necessarily nonnegative, we cannot apply this inequality immediately. However, if $r \geq 1$, and $\mathbb{E}[|M_n|^r] < \infty$ for all n , then $|M_n|^r$ is a submartingale. To check this we need only establish the following fact about conditional expectations: if $r \geq 1$,

$$E[|Y|^r | \mathcal{F}_n] \geq |E[Y | \mathcal{F}_n]|^r, \quad (5.16)$$

for then

$$E[|M_{n+1}|^r | \mathcal{F}_n] \geq |E(M_{n+1} | \mathcal{F}_n)|^r = |M_n|^r.$$

Also, if $\mathbb{E}[e^Y] < \infty$, then

$$E[e^Y | \mathcal{F}] \geq e^{E(Y|\mathcal{F})}, \quad (5.17)$$

and hence for every b ,

$$E[e^{bM_{n+1}} | \mathcal{F}_n] \geq e^{E(bM_{n+1} | \mathcal{F}_n)} = e^{bM_n}.$$

This shows that e^{bM_n} is a submartingale, assuming $\mathbb{E}[e^{bM_n}] < \infty$. We leave the derivation of (5.16) and (5.17) to Exercise 5.3, but we state the conclusion here.

Doob's Maximal Inequality. Suppose M_0, M_1, M_2, \dots is a martingale with respect to \mathcal{F}_n . Then for every $a, b > 0$ and $r \geq 1$.

$$\mathbb{P}\{|\bar{M}_n| \geq a\} \leq \frac{\mathbb{E}[|M_n|^r]}{a^r},$$

$$\mathbb{P}\{\bar{M}_n \geq a\} \leq \frac{\mathbb{E}[e^{bM_n}]}{e^{ba}}.$$

Example. Let $S_n = X_1 + \dots + X_n$ denote simple random walk in \mathbb{Z} , and let $b = 1/\sqrt{n}$. Since S_n is a martingale, we get

$$\mathbb{P}\{\max\{S_1, \dots, S_n\} \geq a\sqrt{n}\} \leq e^{-a} \mathbb{E}[e^{S_n/\sqrt{n}}].$$

But,

$$\begin{aligned} \mathbb{E}[e^{S_n/\sqrt{n}}] &= \mathbb{E}[e^{(X_1+\dots+X_n)/\sqrt{n}}] \\ &= (\mathbb{E}[e^{X_1/\sqrt{n}}])^n = \left(\frac{e^{1/\sqrt{n}} + e^{-1/\sqrt{n}}}{2}\right)^n. \end{aligned}$$

Taylor series shows that

$$\frac{e^{1/\sqrt{n}} + e^{-1/\sqrt{n}}}{2} = 1 + \frac{1}{2n} + O\left(\frac{1}{n^2}\right).$$

Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{E}[e^{S_n/\sqrt{n}}] = \lim_{n \rightarrow \infty} \left[1 + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)\right]^n = e^{1/2}.$$

Hence, there is a $C < \infty$ such that for all n sufficiently large and for all $a > 0$,

$$\mathbb{P}\{\max\{S_1, \dots, S_n\} \geq a\sqrt{n}\} \leq C e^{-a}. \quad (5.18)$$

5.7 Exercises

5.1 Consider the experiment of rolling two dice. Let X be the value of the first roll and Y the sum of the two dice. Find $E(X | Y)$, i.e., give the value of $E(X | Y)(y)$ for all y .

5.2 Suppose that X_t is a Poisson process with parameter $\lambda = 1$. Find $E(X_1 | X_2)$ and $E(X_2 | X_1)$.

5.3 A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is *convex* if for every $0 \leq p \leq 1$ and $x < y$,

$$f(px + (1 - p)y) \leq p f(x) + (1 - p) f(y).$$

- (a) Show that if $f''(x) \geq 0$ for all x , then f is convex.
- (b) Show that if $r \geq 1$, then $f(x) = |x|^r$ is convex.
- (c) Show that if b is a real number, then $f(x) = e^{bx}$ is convex.
- (d) Show that if f is convex; p_1, \dots, p_n are nonnegative numbers summing to 1; and x_1, \dots, x_n are real numbers, then

$$f\left(\sum_{j=1}^n p_j x_j\right) \leq \sum_{j=1}^n p_j f(x_j).$$

(e) Establish *Jensen's inequality*: for any random variable X , $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$, assuming the expectations exist.

(f) Show that if Y is a discrete random variable and X is as in (e), then $E(f(X) | Y) \geq f(E(X) | Y)$. (Note: this fact can then be established for Y that are not discrete by a limit process.)

5.4 Let X_1, X_2, X_3, \dots be independent identically distributed random variables. Let $m(t) = \mathbb{E}(e^{tX_1})$ be the moment generating function of X_1 (and hence of each X_i). Fix t and assume $m(t) < \infty$. Let $S_0 = 0$ and for $n > 0$,

$$S_n = X_1 + \cdots + X_n.$$

Let $M_n = m(t)^{-n} e^{tS_n}$. Show that M_n is a martingale with respect to X_1, X_2, \dots .

5.5 Let X_0, X_1, \dots be the values of a branching process as in Chapter 2, Section 2.4, i.e., X_n gives the number of individuals in the n th generation. Suppose that the mean number of offspring per individual is μ . Show that $M_n = \mu^{-n} X_n$ is a martingale with respect to X_0, X_1, \dots

5.6 COMPUTER SIMULATION

(a) Consider the Polya urn model. Simulate this model with a computer by starting with one red and one green ball and continuing until the number of balls in the urn is 1000. Note the fraction of red balls in the 1000 balls. Do this simulation at least 2000 times and note how many times the fraction of red balls is in the intervals $[0, .05], [.05, .1], \dots, [.95, 1]$. From the simulation data, make a conjecture as to what the distribution of the fraction of red balls looks like.

(b) Do another simulation of the Polya urn model. Again, start with one red and one green ball and continue until there are 1000 balls in the urn. Note the proportion of red balls at this time and then continue until there are 2000 balls. Compare these two numbers (i.e., compare M_{998} and M_{1998}). Do this at least 100 times.

5.7 Consider a biased random walk on the integers with probability $p < 1/2$ of moving to the right and probability $1 - p$ of moving to the left. Let S_n be the value at time n and assume that $S_0 = a$, where $0 < a < N$.

(a) Show that $M_n = [(1 - p)/p]^{S_n}$ is a martingale.

(b) Let T be the first time that the random walk reaches 0 or N , i.e.,

$$T = \min\{n : S_n = 0 \text{ or } N\}.$$

Use optional sampling on the martingale M_n to compute $\mathbb{P}\{S(T) = 0\}$.

5.8 Let S_n be as in Exercise 5.7.

(a) Show that $M_n = S_n + (1 - 2p)n$ is a martingale.

(b) Let T be the first time that the random walk reaches 0 or N , i.e.,

$$T = \min\{n : S_n = 0 \text{ or } N\}.$$

Let $T_n = \min\{n, T\}$ and let Z_n be the martingale $Z_n = M_{T_n}$. Show that there exists a $C < \infty$ such that $\mathbb{E}(Z_n^2) < C$ for all n . You may wish to use Exercise 1.7.

(c) Apply the optional sampling theorem to $\mathbb{E}(M_T)$ and use this and the result from Exercise 5.7 to find the expected number of steps until absorption, $\mathbb{E}(T)$.

5.9 Suppose X_n is an irreducible Markov chain on finite state space S with transition matrix \mathbf{P} . Suppose A is a subset of S and $F : A \rightarrow \mathbb{R}$ and $g : S \setminus A \rightarrow \mathbb{R}$ are given functions. Let $T = \min\{n : X_n \in A\}$ and $T_n = \min\{n, T\}$. Suppose $f : S \rightarrow \mathbb{R}$ is a function satisfying:

$$f(x) = F(x), \quad x \in A$$

$$\mathbf{P}f(x) = g(x), \quad x \in S \setminus A.$$

(a) Show that

$$M_n = f(X_{T_n}) - \sum_{j=0}^{T_n-1} g(X_{T_j}),$$

is a martingale.

(b) Use optional sampling to conclude that

$$f(x) = \mathbb{E} \left[F(X_T) - \sum_{j=0}^{T-1} g(X_j) \mid X_0 = x \right].$$

(Hint: Exercise 1.7 could be useful.)

5.10 Let S_n be as in Exercise 5.7 and let \mathcal{F}_n denote the information in S_0, \dots, S_n . Let

$$M_n = \frac{1}{[4p(1-p)]^{n/2}} \left(\frac{1-p}{p} \right)^{S_n/2}.$$

- (a) Show that M_n is a martingale with respect to \mathcal{F}_n .
- (b) Show that $M_n S_n$ is a martingale with respect to \mathcal{F}_n .
- (c) Suppose that R_n is a process such that $R_0 = M_0$ and both R_n and $R_n S_n$ are martingales with respect to \mathcal{F}_n . Show that $R_n = M_n$ for all n .

5.11 Let X_n be the number of individuals in the n th generation of a branching process in which each individual produces offspring from a distribution with mean μ and variance σ^2 . We have seen previously that $M_n = \mu^{-n} X_n$ is a martingale.

(a) Let \mathcal{F}_n denote the information contained in X_0, \dots, X_n . Show that

$$E(X_{n+1}^2 \mid \mathcal{F}_n) = \mu^2 X_n^2 + \sigma^2 X_n.$$

(b) Suppose $\mu > 1$. Show that there exists a $C < \infty$ such that for all n

$$\mathbb{E}(M_n^2) < C.$$

(c) Show that this is not the case if $\mu \leq 1$.

5.12 Consider the Polya urn problem. Let M_n be the proportion of red balls after n draws (starting with one red and one green ball). Prove by induction on n that

$$\mathbb{P}\left\{M_n = \frac{k}{n+2}\right\} = \frac{1}{n+1}, \quad k = 1, 2, \dots, n+1.$$

5.13 Suppose X_1, X_2, \dots are uniformly integrable with $X_n \rightarrow Y$ with probability one. Show that $\mathbb{E}(X_n) \rightarrow \mathbb{E}(Y)$.

5.14 Let X_1, X_2, \dots be independent, identically distributed random variables taking values in $\{-1, 0, 1, \dots\}$ with mean $\mu < 0$. Let $S_0 = 1$ and for $n > 0$,

$$S_n = 1 + X_1 + \dots + X_n.$$

Let $T = \min\{n : S_n = 0\}$. By the law of large numbers, we know that $\mathbb{P}\{T < \infty\} = 1$. Show that $\mathbb{E}(T) \leq 1/|\mu|$. [Hint: it suffices to prove for each n , if $T_n = \min\{n, T\}$, then $\mathbb{E}(T_n) \leq 1/|\mu|$. Consider the martingale $M_n = S_n - n\mu$.] Exercise 5.16 below can be used to prove that $\mathbb{E}(T) = 1/|\mu|$.

5.15 Let M_n be a martingale with respect to \mathcal{F}_n . Assume there exists a nonnegative random variable Y with $\mathbb{E}(Y) < \infty$ and $|M_n| \leq Y$ for all n . Show that M_n is a uniformly integrable martingale.

5.16 Let X_1, X_2, \dots be independent, identically distributed random variables with mean μ . Let T be a stopping time with respect to X_1, X_2, \dots with $\mathbb{E}(T) < \infty$.

(a) Let

$$Y = \sum_{n=1}^{\infty} |X_n| I\{T \geq n\},$$

where I denotes the indicator function. Show that $\mathbb{E}(Y) < \infty$.

(b) Let $T_n = \min\{n, T\}$ and

$$M_n = X_1 + \dots + X_{T_n} - \mu T_n.$$

Explain why M_n is a uniformly integrable martingale (see Exercise 5.15).

(c) Prove Wald's equation,

$$\mathbb{E} \left(\sum_{n=1}^T X_n \right) = \mu \mathbb{E}(T). \quad (5.19)$$

(d) Suppose $\{\mathcal{F}_n\}$ is a filtration such that X_n is \mathcal{F}_n -measurable and for $m > n$, X_m is independent of \mathcal{F}_n (i.e., X_m is independent of every \mathcal{F}_n -measurable random variable). Suppose that T is a stopping time with respect to $\{\mathcal{F}_n\}$. (In other words, more information than X_1, \dots, X_n is used to determine whether to stop at time n . However, any additional information used is independent of X_{n+1}, X_{n+2}, \dots). Show that (a) through (c) still hold.

5.17 Let S_n be simple random walk in \mathbb{Z} .

(a) Show that for every $\beta > 0$ there is a $C_\beta < \infty$ such that for all positive integers n and all $a > 0$

$$\mathbb{P}\{\max\{S_1, \dots, S_n\} \geq a\sqrt{n}\} \leq C_\beta e^{-a\beta}.$$

(Hint: follow the derivation of (5.18) using $b = \beta/\sqrt{n}$.)

(b) Show that for every $c > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\{S_n \geq c\sqrt{n} \log n\} < \infty.$$

(c) Use this to show that with probability one,

$$\lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}(\log n)} = 0.$$

Chapter 6

Renewal Processes

6.1 Introduction

Let T_1, T_2, \dots be independent, identically distributed, nonnegative random variables with distribution function $F(x) = \mathbb{P}\{T_i \leq x\}$. We will think of the random variables T_i as being the lifetimes of a component or as the times between occurrences of some event. The *renewal process* associated with T_i is the process that counts the number of events that have occurred by time t . More precisely, the renewal process N_t is defined by $N_t = 0$ for $t < T_1$ and otherwise

$$N_t = \max\{n : T_1 + \cdots + T_n \leq t\}.$$

We are assuming that at time 0 we are at the beginning of a lifetime. Sometimes we will consider a slightly more general process where the process at time 0 is in the middle of a lifetime. We let Y be a nonnegative random variable independent of T_1, T_2, \dots , with perhaps a different distribution. We think of Y as the time until the first event, and then the waiting times for later events are given by the T_i . More precisely, we set $N_t = 0$ for $t < Y$; and for $t \geq Y$,

$$N_t = \min\{n : Y + T_1 + \cdots + T_n > t\}. \quad (6.1)$$

We will assume that the random variables T_i have finite, positive mean and we set

$$\mu = \mathbb{E}(T_i).$$

Example 1. Poisson Process. Consider the Poisson process with rate parameter λ . The waiting times T_1, T_2, \dots are independent, exponential random variables with parameter λ and N_t is the Poisson process. In this case $\mu = 1/\lambda$.

Example 2. Let X_n be an irreducible, positive recurrent, discrete-time Markov chain starting in state x . Let

$$T_1 = \min\{n > 0 : X_n = x\},$$

and for $i > 1$ let

$$T_i = \min\{n > 0 : X_{T_1 + \dots + T_{i-1} + n} = x\}.$$

In other words, T_i measures the amount of time between the $(i - 1)$ st return and the i th return to state x . In general it is difficult to determine the distribution function F for T_i given the transition matrix for the chain. We noted previously [see (1.11)] that

$$\mathbb{E}(T_i) = \frac{1}{\pi(x)},$$

where π denotes the invariant probability measure for the chain. If we instead start the chain at some state $y \neq x$ we can define

$$Y = \min\{n > 0 : X_n = x\},$$

$$T_1 = \min\{n > 0 : X_{Y+n} = x\},$$

and recursively,

$$T_i = \min\{n > 0 : X_{Y+T_1 + \dots + T_{i-1} + n} = x\}.$$

Example 3. Let X_t be an irreducible, positive recurrent, continuous-time Markov chain starting in state x . Define

$$R_1 = \inf\{t > 0 : X_t \neq x\},$$

$$S_1 = \inf\{t > R_1 : X_t = x\},$$

$$T_1 = R_1 + S_1,$$

and in general

$$R_i = \inf\{t > 0 : X_{T_1 + \dots + T_{i-1} + t} \neq x\},$$

$$S_i = \inf\{t > 0 : X_{T_1 + \dots + T_{i-1} + R_i + t} = x\},$$

$$T_i = R_i + S_i.$$

The random variables R_i are exponential with parameter $\alpha(x)$, the rate at which the chain is changing from state x . The distribution of the S_i , and hence the T_i , is not so easy to determine.

Example 4. $M/G/1$ Queue. Suppose we have a queue with a single server. Customers arrive according to a Poisson process with rate λ , i.e., the waiting times between customer arrivals are independent exponential random variables with parameter λ . We will assume that the service times for customers are independent, identically distributed random variables with mean μ . However, we will not assume that the service times are exponential (in most cases of interest one does not expect that the service time should have the “loss of memory” property so an exponential distribution is not appropriate). The G in $M/G/1$ stands for “general” (service distribution).

If we let Y_t denote the number of people in the queue at time t , then Y_t is not a Markov process. However there is a natural renewal process one can associate with the queue. Suppose $Y_0 = 0$. Let

$$R_1 = \inf\{t > 0 : Y_t = 1\},$$

$$S_1 = \inf\{t > 0 : Y_{R_1+t} = 0\},$$

$$T_1 = R_1 + S_1.$$

Similarly, we define for $i > 1$,

$$R_i = \inf\{t > 0 : Y_{T_1+\dots+T_{i-1}+t} = 1\},$$

$$S_i = \inf\{t > 0 : Y_{T_1+\dots+T_{i-1}+R_i+t} = 0\},$$

$$T_i = R_i + S_i.$$

Note that the variables R_i are exponential with rate λ , but the distribution of the S_i can be very complicated. Nevertheless, under the assumption that $\mathbb{E}(T_i) < \infty$, we can see that T_1, T_2, \dots satisfy the conditions for a renewal process. We can think of the time represented by the R_i as the “idle times” and the time represented by the S_i as the “busy times.”

Suppose we have a renewal process N_t corresponding to the random variables T_1, T_2, \dots . In general, N_t is not a Markov process; in order to predict when the next occurrence will happen we need to know when the last occurrence took place. For this reason it is natural to consider the “age process”

$$A_t = \begin{cases} t, & \text{if } N_t = 0, \\ t - [T_1 + \dots + T_{N_t}], & \text{if } N_t > 0. \end{cases}$$

The process (N_t, A_t) can be thought of as a Markov process. The Poisson process is a special example of a renewal process that *is* a Markov process; for the Poisson process the probability of an event occurring in the interval

$[t, t + \Delta t]$ is independent of A_t . This follows from the “loss of memory” property associated with the exponential distribution.

Our first result for renewal processes will be the analogue of the (strong) law of large numbers. Recall that the law of large numbers states that with probability 1,

$$\lim_{n \rightarrow \infty} \frac{T_1 + \cdots + T_n}{n} = \mu.$$

In terms of the renewal process N_t this states that for all $\epsilon > 0$, if n is sufficiently large,

$$N_{\mu n(1-\epsilon)} \leq n,$$

$$N_{\mu n(1+\epsilon)} \geq n.$$

Equivalently, for all $\epsilon > 0$, if t is sufficiently large,

$$N_t \leq \frac{t}{\mu(1-\epsilon)},$$

$$N_t \geq \frac{t}{\mu(1+\epsilon)}.$$

This gives the following.

Law of Large Numbers. *With probability one,*

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mu}. \quad (6.2)$$

We now derive a central limit theorem for renewal processes. Assume that the variance of each T_i is $\sigma^2 < \infty$. Then the usual central limit theorem states that the distribution of

$$\frac{T_1 + \cdots + T_n - n\mu}{\sigma\sqrt{n}}$$

approaches a unit normal (i.e., a normal random variable with mean 0, variance 1). Slightly more informally we can say that for large n

$$T_1 + \cdots + T_n \approx n\mu + \sigma\sqrt{n}B,$$

where B is a unit normal. This states that the number of occurrences in time $n\mu + \sigma\sqrt{n}B$ is n . From (6.2), we would expect the number of occurrences in

the time interval of size $\sigma\sqrt{n}|B|$ to be about $\sigma\sqrt{n}|B|/\mu$. Hence we have the number of occurrences in time $n\mu$ is about

$$n - \frac{\sigma}{\mu} \sqrt{n} B.$$

If we write t for $n\mu$ and note that $-B$ is also a unit normal random variable we see that

$$N_t \approx \frac{t}{\mu} + \frac{\sigma}{\mu^{3/2}} \sqrt{t} B,$$

where B is a unit normal. While this is only a rough sketch, this argument can be made rigorous, giving a central limit theorem for renewal processes.

Central Limit Theorem. *If the waiting times T_i have mean μ and variance σ^2 , then as $t \rightarrow \infty$ the distribution of*

$$\frac{N_t - \mu^{-1}t}{\sigma\mu^{3/2}\sqrt{t}}$$

approaches a standard normal distribution.

Example 5. This kind of informal reasoning can be applied to more complicated examples. Suppose we have a continuous-time Markov chain X_t on state space $\{1, 2\}$ with $\alpha(1, 2) = \alpha_1$ and $\alpha(2, 1) = \alpha_2$. Assume $X_0 = 1$ and let Y_t denote the amount of time spent in state 1 up to time t ,

$$Y_t = \int_0^t I\{X_s = 1\} ds.$$

Define R_i and S_i as in Example 3 above (with $x = 1$). The random variables R_i are exponential with rate α_1 and hence have mean $\mu_1 = 1/\alpha_1$ and variance $\sigma_1^2 = 1/\alpha_1^2$. Similarly the random variables S_i are exponential with mean $\mu_2 = 1/\alpha_2$ and variance $\sigma_2^2 = 1/\alpha_2^2$. For large n the central limit theorem states that

$$R_1 + \cdots + R_n \approx n\mu_1 + \sigma_1\sqrt{n}B_1,$$

$$S_1 + \cdots + S_n \approx n\mu_2 + \sigma_2\sqrt{n}B_2,$$

where B_1 and B_2 are independent unit normals. In other words, in time $n(\mu_1 + \mu_2) + \sqrt{n}(\sigma_1 B_1 + \sigma_2 B_2)$, the amount of time spent in state 1 is approximately $n\mu_1 + \sqrt{n}\sigma_1 B_1$. For large t , the amount of time spent in state 1 in an interval $[t, t + \Delta t]$ is about $\Delta t[\mu_1/(\mu_1 + \mu_2)]$. Hence the amount of time spent in state 1 up through time $(\mu_1 + \mu_2)n$ is approximately

$$n\mu_1 + \sqrt{n}\sigma_1 B_1 - \frac{\mu_1}{\mu_1 + \mu_2} \sqrt{n}(\sigma_1 B_1 + \sigma_2 B_2)$$

$$= n\mu_1 + \frac{\sqrt{n}}{\mu_1 + \mu_2} [\mu_2\sigma_1 B_1 - \mu_1\sigma_2 B_2].$$

Since B_1 and B_2 are independent, we can write this as

$$n\mu_1 + \sqrt{n} \sqrt{(\frac{\sigma_1\mu_2}{\mu_1 + \mu_2})^2 + (\frac{\sigma_2\mu_1}{\mu_1 + \mu_2})^2} B = \frac{1}{\alpha_1} n + \sqrt{2n} \frac{1}{\alpha_1 + \alpha_2} B,$$

where B is a unit normal. If we let $t = (\mu_1 + \mu_2)n$ we see that the distribution of

$$\frac{Y_t - \frac{\alpha_2}{\alpha_1 + \alpha_2} t}{\bar{\sigma}\sqrt{t}}$$

approaches a unit normal where

$$\bar{\sigma}^2 = \frac{2\alpha_1\alpha_2}{(\alpha_1 + \alpha_2)^3}.$$

6.2 Renewal Equation

We are interested in the large-time behavior of renewal processes. Assume we have a renewal process with waiting times T_1, T_2, \dots with mean μ as defined in the previous section. For $T \geq 0$, we let $U(t)$ be the expected number of occurrences up through time t , where for convenience we will say that an event occurs at time 0. In other words,

$$U(t) = \mathbb{E}(N_t + 1).$$

Renewal Theorem I

$$\lim_{t \rightarrow \infty} \frac{U(t)}{t} = \frac{1}{\mu}. \quad (6.3)$$

This is almost a consequence of (6.2); one does need to be a little careful, however, because it is possible for random variables to converge without the expectations converging. We leave the derivation of (6.3) from (6.2) to the exercises (Exercise 6.5).

To analyze the large-time behavior of renewal processes we will need a second, stronger version of the renewal theorem. The second renewal theorem can be thought of as a “derivative” form of (6.3) or as a statement that the renewal process converges to a steady state. The second renewal theorem states that under appropriate hypotheses, for every $r > 0$,

$$\lim_{t \rightarrow \infty} U(t + r) - U(t) = \frac{r}{\mu}, \quad (6.4)$$

i.e., for large t , the expected number of renewals in any interval of length r is about r/μ . It is not too difficult to see that some restrictions must be put on the distribution for (6.4) to hold. For example, if the waiting times T_i take on only integer values, then for every integer n ,

$$U(n) = U\left(n + \frac{1}{2}\right),$$

since renewals occur only at integer times. It turns out that this is really the only thing that can go wrong. We say that a nonnegative random variable X has a lattice distribution if there exists a number a such that with probability one the value of X lies in

$$\{ak : k = 0, 1, 2, \dots\},$$

and we call the smallest such a the *period* of the distribution. Otherwise we say the X has a nonlattice distribution. We now state the second renewal theorem.

Renewal Theorem II. *If T_1, T_2, \dots have a nonlattice distribution, then for every $r > 0$,*

$$\lim_{t \rightarrow \infty} U(t + r) - U(t) = \frac{r}{\mu}.$$

If the T_1, T_2, \dots have a lattice distribution with period a , then

$$\lim_{n \rightarrow \infty} U((n + 1)a) - U(na) = \frac{a}{\mu}.$$

We will not give a proof of the nonlattice form of this theorem, but rather will concentrate on showing how it is used. In the next section we will relate the lattice form of this theorem to known results about positive recurrent Markov chains. Let F denote the distribution of T_i . Recall that the convolution of two distributions F, G of nonnegative random variables is defined by

$$F * G(t) = \int_0^t F(t - s) dG(s) = \int_0^t G(t - s) dF(s).$$

The convolution $F * G$ gives the distribution function of the sum of two independent random variables with distribution functions F and G respectively. Let F be the distribution function for the T_i . We will write $F^{(n)}$ for the convolution of F n times, i.e., for the distribution function of $T_1 + \dots + T_n$. For convenience we will write $F^{(0)}$ for the trivial distribution function associated to the random variable which is identically 0. Recall [see (1.13)] that if Y is a random variable taking values in the nonnegative integers, then

$$\mathbb{E}(Y) = \sum_{n=1}^{\infty} \mathbb{P}\{Y \geq n\}.$$

Using this, we can write the renewal function $U(t)$ as

$$\begin{aligned} U(t) &= \mathbb{E}(N_t + 1) = 1 + \sum_{n=1}^{\infty} \mathbb{P}\{N_t \geq n\} \\ &= 1 + \sum_{n=1}^{\infty} \mathbb{P}\{T_1 + \cdots + T_n \leq t\} \\ &= \sum_{n=0}^{\infty} F^{(n)}(t). \end{aligned}$$

Let A_t denote the time elapsed since the last renewal,

$$A_t = \begin{cases} t & \text{if } N_t = 0, \\ t - (T_1 + \cdots + T_n), & \text{if } N_t = n. \end{cases}$$

If we think of the times T_i as being lifetimes of some component, then A_t represents the age of the current component. We would like to determine the steady-state distribution of A_t , i.e., we would like to determine for each x ,

$$\Psi_A(x) = \lim_{t \rightarrow \infty} \mathbb{P}\{A_t \leq x\}.$$

We will condition on the first renewal. One way for A_t to be less than x is for no event to have occurred up through time t and $t \leq x$. This corresponds to $t < T_1$ and has probability $1 - F(t)$ if $t \leq x$. If the first renewal has occurred before time t , at time s say, then the renewal process starts over and there is time $t - s$ left until time t . From this we get the equation

$$\mathbb{P}\{A_t \leq x\} = 1_{[0,x]}(t) [1 - F(t)] + \int_0^t \mathbb{P}\{A_{t-s} \leq x\} dF(s). \quad (6.5)$$

Here $1_{[0,x]}(t)$ denotes the function that equals 1 for $0 \leq t \leq x$ and equals zero otherwise. If we let $\phi(t) = \phi(t, x) = \mathbb{P}\{A_t \leq x\}$, then this becomes

$$\phi(t) = 1_{[0,x]}(t) [1 - F(t)] + \int_0^t \phi(t-s) dF(s).$$

This is an example of a *renewal equation*. We will now consider solutions to renewal equations of the form

$$\phi(t) = h(t) + \int_0^t \phi(t-s) dF(s), \quad (6.6)$$

or in the language of convolutions,

$$\phi(t) = h(t) + \phi * F(t).$$

We will need the associativity property for convolutions: if F and G are distribution functions

$$(\phi * F) * G(t) = \phi * (F * G)(t). \quad (6.7)$$

Let us derive this in the case where F and G have densities, so that $dF(t) = f(t) dt$ and $dG(t) = g(t) dt$. In this case

$$\begin{aligned} (\phi * F) * G(t) &= \int_0^t (\phi * F)(t-s)g(s) ds \\ &= \int_0^t \left[\int_0^{t-s} \phi(t-s-r)f(r) dr \right] g(s) ds \\ &= \int_0^t \left[\int_s^t \phi(t-y)f(y-s) dy \right] g(s) ds \\ &= \int_0^t \phi(t-y) \left[\int_0^y f(y-s)g(s) ds \right] dy \\ &= \int_0^t \phi(t-y)(f * g)(y) dy \\ &= \phi * (F * G)(t). \end{aligned}$$

Here $(f * g)(y) = (d/dy)(F * G)(y)$ denotes the density of the sum of two independent random variables with density f and g , respectively.

We will first show that there is only one solution to (6.6) in the sense that there is at most one $\phi(t)$ that satisfies (6.6) with $\phi(t) = 0$ for $t < 0$ and such that for each t there is a number $M = M_t < \infty$ with $|\phi(s)| \leq M$ for all $0 \leq s \leq t$. Assume there were two such solutions, $\phi_1(t)$ and $\phi_2(t)$, for a given h . Then $\psi(t) = \phi_1(t) - \phi_2(t)$ satisfies $|\psi(s)| \leq 2M$, $0 \leq s \leq t$, and

$$\psi(t) = \int_0^t \psi(t-s) dF(s).$$

If we iterate (6.7) we see for each n ,

$$\psi(t) = \int_0^t \psi(t-s) dF^{(n)}(s).$$

But,

$$|\psi(t)| = \left| \int_0^t \psi(t-s) dF^{(n)}(s) \right| \leq 2MF^{(n)}(t).$$

For fixed t , $F^{(n)}(t) \rightarrow 0$ as $n \rightarrow \infty$. This shows that $\psi(t) = 0$.

Now that we know there is only one solution, we need only produce a solution. Let

$$\begin{aligned}\phi(t) &= \int_0^t h(t-s) dU(s) = \sum_{n=0}^{\infty} \int_0^t h(t-s) dF^{(n)}(s) \\ &= h(t) + \sum_{n=1}^{\infty} \int_0^t h(t-s) dF^{(n)}(s).\end{aligned}$$

Then one can see, using (6.7), that this satisfies (6.6). This therefore gives the unique solution.

Let us now assume that the F is a nonlattice distribution. Another way of stating the second renewal theorem is to say that for large s ,

$$dU(s) \approx \mu^{-1} ds.$$

If $h(t)$ is a bounded function with $\int_0^\infty |h(t)| dt < \infty$, then this implies that

$$\lim_{t \rightarrow \infty} \int_0^t h(t-s) dU(s) = \lim_{t \rightarrow \infty} - \int_0^t h(s) dU(t-s) = \frac{1}{\mu} \int_0^\infty h(s) ds. \quad (6.8)$$

Since the age distribution A_t satisfies (6.5), we can conclude that the large-time age distribution function $\Psi_A(x)$ is given by

$$\begin{aligned}\Psi_A(x) &= \lim_{t \rightarrow \infty} \mathbb{P}\{A_t \leq x\} = \frac{1}{\mu} \int_0^\infty 1_{[0,x]}(s) [1 - F(s)] ds \\ &= \frac{1}{\mu} \int_0^x [1 - F(s)] ds.\end{aligned}$$

Note that

$$\begin{aligned}\lim_{x \rightarrow \infty} \Psi_A(x) &= \frac{1}{\mu} \int_0^\infty [1 - F(s)] ds \\ &= \frac{1}{\mu} \int_0^\infty \int_s^\infty dF(r) ds \\ &= \frac{1}{\mu} \int_0^\infty \left[\int_0^r ds \right] dF(r) \\ &= \frac{1}{\mu} \int_0^\infty r dF(r) = 1,\end{aligned}$$

so this gives a valid distribution function. It has density

$$\psi_A(x) = \Psi'_A(x) = \frac{1}{\mu} [1 - F(x)], \quad 0 < x < \infty.$$

Example 1. Suppose that the waiting times are exponential with rate λ , so that $F(t) = 1 - e^{-\lambda t}$, $\mu = 1/\lambda$. Then

$$\Psi_A(x) = \lim_{t \rightarrow \infty} \mathbb{P}\{A_t \leq x\} = \frac{1}{\mu} \int_0^x e^{-\lambda s} ds = 1 - e^{-\lambda x}.$$

Hence the large-time age distribution for a Poisson process with rate λ is an exponential distribution with rate λ . This is very plausible: at a large time t , the age A_t is the amount of time in the past one must go to see an event. This reverse process also looks like a Poisson process, so the time until an event should be exponential.

Example 2. Suppose that the waiting time distribution is uniform on $[0, 10]$ so that $F(t) = (t/10)$, $0 \leq t \leq 10$, and $\mu = 5$. Then the age A_t is always less than 10 and for $x < 10$,

$$\Psi_A(x) = \lim_{t \rightarrow \infty} \mathbb{P}\{A_t \leq x\} = \frac{1}{\mu} \int_0^x \left[1 - \frac{t}{10}\right] dt = \frac{x}{5} - \frac{x^2}{100}. \quad (6.9)$$

Note in this case (as in essentially all cases but for exponential waiting times) the large-time age distribution is not the same as the waiting time distribution.

We will now consider two other processes, the *residual life*

$$B_t = \inf\{s : N_{t+s} > N_t\},$$

and the *total lifetime*

$$C_t = A_t + B_t.$$

The residual life gives the amount of time until the current component in a system fails. Consider $\mathbb{P}\{B_t \leq x\}$. There are two ways for B_t to be less than x . One way is for there to be no renewals up to time t and $B_t \leq x$. This corresponds to $t < T_1 \leq t + x$ which has probability $F(t + x) - F(t)$. The other possibility is that there is a first renewal at time $s < t$ in which case we need to consider $\{B_{t-s} \leq x\}$. This gives the renewal equation

$$\mathbb{P}\{B_t \leq x\} = [F(t + x) - F(t)] + \int_0^t \mathbb{P}\{B_{t-s} \leq x\} dF(s).$$

The solution to this renewal equation is

$$\mathbb{P}\{B_t \leq x\} = \int_0^t [F(t - s + x) - F(t - s)] dU(s).$$

From (6.8), we can determine the large-time residual life distribution function $\Psi_B(x)$,

$$\begin{aligned}
\Psi_B(x) &= \lim_{t \rightarrow \infty} \int_0^t [F(t-s+x) - F(t-s)] dU(s) \\
&= \lim_{t \rightarrow \infty} - \int_0^t [F(s+x) - F(s)] dU(t-s) \\
&= \frac{1}{\mu} \int_0^\infty [F(s+x) - F(s)] ds \\
&= \frac{1}{\mu} \left[\int_0^\infty [1 - F(s)] ds - \int_0^\infty [1 - F(s+x)] ds \right] \\
&= \frac{1}{\mu} \left[\int_0^\infty [1 - F(s)] ds - \int_x^\infty [1 - F(r)] dr \right] \\
&= \frac{1}{\mu} \int_0^x [1 - F(s)] ds.
\end{aligned}$$

What we see is that the large-time distribution function for the residual life is the same as that for the age distribution. If one thinks about this, it is reasonable. Consider every lifetime T_i . For every r, s with $r+s = T_i$, there will correspond one time t when $A_t = r$, $B_t = s$ and another time u when $A_u = s$, $B_u = r$. By this symmetry, we would expect A_t and B_t to have the same limiting distribution.

Now consider the total lifetime C_t and $\mathbb{P}\{C_t \leq x\}$. One way for C_t to be less than x is for there to be no renewals up through time t and the total lifetime less than x . This corresponds to $t < T_1 \leq x$ which has probability $F(x) - F(t)$. The other possibility is that the first event occurs at some $s < t$ in which case we need to consider $\mathbb{P}\{C_{t-s} \leq x\}$. This gives the renewal equation

$$\mathbb{P}\{C_t \leq x\} = 1_{[0,x]}(t) [F(x) - F(t)] + \int_0^t \mathbb{P}\{C_{t-s} \leq x\} dF(s).$$

By solving the renewal equation and using (6.8), we see that the limiting distribution for the lifetime, $\Psi_C(x)$ is given by

$$\begin{aligned}
\Psi_C(x) &= \lim_{t \rightarrow \infty} \int_0^t 1_{[0,x]}(t-s) [F(x) - F(t-s)] dU(s) \\
&= \lim_{t \rightarrow \infty} - \int_0^t 1_{[0,x]}(s) [F(x) - F(s)] dU(t-s) \\
&= \frac{1}{\mu} \int_0^\infty 1_{[0,x]}(s) [F(x) - F(s)] ds \\
&= \frac{1}{\mu} \int_0^x [F(x) - F(s)] ds \\
&= \frac{1}{\mu} \left[xF(x) - \int_0^x F(s) ds \right].
\end{aligned}$$

This formula is best understood in the case where F has a density $f(t)$. In

this case $\Psi_C(x)$ has density

$$\psi_C(x) = \Psi'_C(x) = \frac{1}{\mu} x f(x). \quad (6.10)$$

This can be understood intuitively. Suppose $x < y$. Then the relative “probability” of waiting times of size x and size y is $f(x)/f(y)$. However, every waiting time of size y uses up y units of time while a waiting time of size x uses up x units of time. So the ratio of times in an interval of size x to an interval of size y should be $xf(x)/yf(y)$. The $1/\mu$ can easily be seen to be the appropriate normalization factor to make this a probability density.

Example 3. If the waiting times are exponential with rate λ , then $\mu = 1/\lambda$ and Ψ_A and Ψ_B have exponential distributions with rate λ . Note that Ψ_C has density

$$\psi_C(x) = \lambda^2 x e^{-\lambda x}.$$

This is the density of a Gamma distribution with parameters 2 and λ and is the density of the sum of two independent exponential random variables with rate λ . For large times, the age and the residual life are independent random variables.

Example 4. If F is uniform on $[0, 10]$, then $\mu = 5$, and Ψ_A and Ψ_B are given by (6.9) with densities

$$\psi_A(x) = \psi_B(x) = \frac{1}{5} - \frac{x}{50}, \quad 0 < x < 10.$$

Note that the expected age or the expected residual life in the long run is given by

$$\int_0^{10} x \left[\frac{1}{5} - \frac{x}{50} \right] dx = \frac{10}{3}.$$

The density of Ψ_C is given by

$$\psi_C(x) = \frac{1}{\mu} x f(x) = \frac{x}{50}, \quad 0 < x < 10.$$

It is easy to check that the age and residual life are not asymptotically independent in this case, e.g., there is a positive probability that the age is over 8 and a positive probability that the residual life is over 8, but it is impossible for both of them to be over 8 since the total lifetime is bounded by 10.

Suppose one is replacing components as they fail and the lifetimes are independent with distribution F . Suppose we consider the system at some large t , and ask how long the present component is expected to last. This is equivalent to finding the expected value of the residual life. This is given by

$$\int_0^\infty x \psi_B(x) dx = \frac{1}{\mu} \int_0^\infty x [1 - F(x)] dx = \frac{1}{2\mu} \int_0^\infty x^2 dF(x).$$

The last equality is obtained by integrating by parts. It is easy to give examples (see Exercise 6.6) of distributions of densities $f(x)$ such that

$$\mu < \frac{1}{2\mu} \int_0^\infty x^2 f(x) dx.$$

In fact, it is possible for $\mu < \infty$ and the expected residual lifetime to be infinite. This may be surprising at first; however, a little thought will show that this is not so unreasonable.

We finish this section by describing how to create a “stationary renewal process.” Suppose T_1, T_2, \dots are independent with nonlattice distribution F . Let Ψ_B be the large-time residual life distribution and let Y be a random variable independent of T_1, T_2, \dots with distribution function Ψ_B . Define N_t as in (6.1). Then N_t looks like a renewal process in steady state. It has the property that for every $s < t$, $N_t - N_s$ has the same distribution as N_{t-s} .

6.3 Discrete Renewal Processes

In this section we will suppose that the random variables T_1, T_2, \dots are lattice random variables. Without loss of generality we will assume that the period a as defined in Section 6.2 is equal to 1 (the period is always equal to 1 in some choice of time units). Let F be the distribution function for the T_i and let

$$p_n = \mathbb{P}\{T_i = n\} = F(n) - F(n-1).$$

We will assume for ease that $p_0 = 0$; if $p_0 > 0$ we can make a slight adjustment of the methods in this section (see Exercise 6.10). Since the period is 1, the greatest common divisor of the set

$$\{n : p_n > 0\}$$

is 1. As before set

$$\mu = \mathbb{E}(T_i) = \sum_{n=1}^{\infty} n p_n,$$

and we assume $\mu < \infty$.

Let N_j denote the number of events that have occurred up through and including time j , i.e., $N_j = 0$ if $j < T_1$ and otherwise

$$N_j = \max\{n : T_1 + \cdots + T_n \leq j\}.$$

We can also define the age process A_j by $A_j = j$ if $j < T_1$ and otherwise

$$A_j = j - (T_1 + \cdots + T_{N_j}).$$

The key fact is that A_j is a Markov chain. Let

$$\lambda_n = \mathbb{P}\{T_i = n \mid T_i > n - 1\} = \frac{p_n}{1 - F(n - 1)}.$$

Then A_j is a discrete-time Markov chain with transition probabilities

$$p(n, 0) = \lambda_{n+1}, \quad p(n, n + 1) = 1 - \lambda_{n+1}.$$

Let K be the largest number k such that $p_k > 0$ (where $K = \infty$ if $p_k > 0$ for infinitely many k). Then A_j is an irreducible Markov chain with state space $\{0, 1, \dots, K - 1\}$ if $K < \infty$ and state space $\{0, 1, \dots\}$ if $K = \infty$. The chain is also aperiodic since we assumed the period of F is 1. We start with $A_0 = 0$ and note that the n th return to state 0 occurs at time $T_1 + \dots + T_n$. The condition $\mathbb{E}(T_i) < \infty$ implies that A_j is a positive recurrent chain.

The invariant probability π for this chain can be obtained by solving the equations

$$\begin{aligned} \pi(n + 1) &= p(n, n + 1) \pi(n) = (1 - \lambda_{n+1}) \pi(n) \\ &= \frac{1 - F(n + 1)}{1 - F(n)} \pi(n), \quad n > 0, \end{aligned}$$

$$\pi(0) = \sum_{n=0}^{\infty} p(n, 0) \pi(n) = \sum_{n=0}^{\infty} \lambda_{n+1} \pi(n).$$

The first equations can be solved recursively to yield

$$\pi(n) = [1 - F(n)] \pi(0).$$

To find the value for $\pi(0)$ for which $\sum \pi(n) = 1$, we check that

$$\begin{aligned} \sum_{n=0}^{\infty} [1 - F(n)] &= \sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} p_m \\ &= \sum_{m=1}^{\infty} p_m \sum_{n=0}^{m-1} 1 \\ &= \sum_{m=1}^{\infty} m p_m = \mu. \end{aligned}$$

In particular,

$$\pi(0) = \frac{1}{\mu}.$$

Note that

$$\mathbb{P}\{\text{an event at time } j\} = \mathbb{P}\{N_j > N_{j-1}\} = \mathbb{P}\{A_j = 0\}.$$

Since A_j is an aperiodic, irreducible, positive recurrent Markov chain we know that

$$\lim_{j \rightarrow \infty} \mathbb{P}\{A_j = 0\} = \pi(0) = \frac{1}{\mu}.$$

This gives the second renewal theorem for discrete renewal processes.

We have also derived the large-time age distribution,

$$\psi_A(n) = \lim_{j \rightarrow \infty} \mathbb{P}\{A_j = n\} = \pi(n) = \frac{1 - F(n)}{\mu}.$$

Consider the residual life,

$$B_j = \min\{k > 0 : N_{j+k} > N_j\}.$$

We can compute the large-time distribution of B_j ,

$$\begin{aligned} \psi_B(n) &= \lim_{j \rightarrow \infty} \mathbb{P}\{B_j = n\} \\ &= \lim_{j \rightarrow \infty} \sum_{m=0}^{\infty} \mathbb{P}\{A_j = m\} \mathbb{P}\{B_j = n \mid A_j = m\} \\ &= \sum_{m=0}^{\infty} \pi(m) \mathbb{P}\{B_j = n \mid A_j = m\} \\ &= \sum_{m=0}^{\infty} \frac{1 - F(m)}{\mu} \frac{p_{n+m}}{1 - F(m)} \\ &= \frac{1}{\mu} \sum_{m=0}^{\infty} p_{n+m} \\ &= \frac{1 - F(n-1)}{\mu}. \end{aligned}$$

In other words,

$$\psi_B(n) = \lim_{j \rightarrow \infty} \mathbb{P}\{B_j = n\} = \lim_{j \rightarrow \infty} \mathbb{P}\{A_j = n-1\} = \psi_A(n-1).$$

The residual life has the same large-time distribution as the age except for a difference of 1 which comes from the fact that the smallest value for the residual life is 1 while the smallest value for the age is 0. For the total lifetime of the component at time j ,

$$C_j = A_j + B_j,$$

we can compute

$$\begin{aligned}
\psi_C(n) &= \lim_{j \rightarrow \infty} \mathbb{P}\{C_j = n\} \\
&= \lim_{j \rightarrow \infty} \sum_{m=0}^{n-1} \mathbb{P}\{A_j = m\} \mathbb{P}\{C_j = n \mid A_j = m\} \\
&= \sum_{m=0}^{n-1} \pi(m) \mathbb{P}\{C_j = n \mid A_j = m\} \\
&= \sum_{m=0}^{n-1} \frac{1 - F(m)}{\mu} \frac{p_n}{1 - F(m)} \\
&= \frac{1}{\mu} \sum_{m=0}^{n-1} p_n \\
&= \frac{n p_n}{\mu}.
\end{aligned}$$

This is the discrete analogue of (6.10).

Example 1. Bernoulli Process. The discrete analogue of the Poisson process is the Bernoulli process. Let $0 < p < 1$ and let X_1, X_2, \dots be independent random variables with $\mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = 0\} = p$. $N_j = X_1 + \dots + X_j$ represents the number of “successes” in the first j trials of an experiment with probability p of success. The waiting times T_i have a geometric distribution

$$\mathbb{P}\{T_i = n\} = (1 - p)^{n-1} p, \quad n \geq 1,$$

with $\mu = 1/p$. The asymptotic age distribution is given by

$$\psi_A(n) = \frac{1 - F(n)}{\mu} = p \sum_{j=m+1}^{\infty} (1 - p)^{m-1} p = p(1 - p)^n,$$

i.e., the age is one less than a random variable with a geometric distribution. The residual life distribution is geometric with parameter p . The asymptotic lifetime distribution is given by

$$\phi_C(n) = n p^2 (1 - p)^{n-1},$$

which is the distribution of the sum of two independent random variables with distributions ϕ_A and ϕ_B , respectively. The age and the residual life are asymptotically independent.

Example 2. Suppose F is uniformly distributed on $\{1, \dots, 10\}$ with $\mu = 11/2$. Then

$$F(n) = \frac{n}{10}, \quad n = 1, 2, \dots, 10.$$

The asymptotic age distribution is given by

$$\psi_A(n) = \frac{1 - F(n)}{\mu} = \frac{10 - n}{55}, \quad n = 0, \dots, 9.$$

and for large time the residual life distribution is given by

$$\psi_B(n) = \frac{1 - F(n-1)}{\mu} = \frac{11 - n}{55}, \quad n = 1, \dots, 10.$$

The asymptotic lifetime distribution is given by

$$\psi_C(n) = \frac{n}{55}, \quad n = 1, 2, \dots, 10.$$

In this case, the age and residual life are not asymptotically independent.

6.4 M/G/1 and G/M/1 Queues

We will consider Example 4 from Section 6.1. Customers arrive into a single-server queue from a Poisson Process with rate λ . Customers are served (first come, first served) and the service time is a random variable with distribution function F and mean $\mu < \infty$. We will call the service rate $1/\mu$, even though the service times are not exponential. The service times and the arrival times are independent. As mentioned before there is a natural renewal process involved where R_1, R_2, \dots denote the amount of time spent in “idle times” while S_1, S_2, \dots denote the amount of time spent in “busy times.” If the queue starts idle, i.e., if $X_0 = 0$ where X_t denotes the size of the queue (including the person being served) at time t , then the time until the start of the next idle time is given by $T_1 = R_1 + S_1$ and the time until the start of the $(n+1)$ st idle time is given

$$T_1 + \dots + T_n,$$

where $T_i = R_i + S_i$.

The times R_i are exponential with rate λ , i.e., with mean $1/\lambda$. The distribution of the S_i is more difficult to determine. However, we will be able to determine $\mathbb{E}(S_i)$. Assume that the service rate is greater than the arrival rate, i.e.,

$$\mu\lambda < 1.$$

Consider the start of a busy time, so that $X_t = 1$. We will consider a discrete-time Markov chain Y_n that gives the number of people in the queue immediately after the n th person has been served. We start with $Y_0 = 1$. The value

Y_1 is obtained by considering the number of people who entered the queue during the first service time and subtracting 1 (for the person who has left the queue). For $i > 1$, Y_i is obtained by adding to Y_{i-1} the number of people who entered the queue while the i th person was being serviced and subtracting one. Let

$$\tau = \min\{n : Y_n = 0\}.$$

If U_1, U_2, \dots denote the service times of the customers, then the length of the first busy time is given by

$$S_1 = U_1 + U_2 + \dots + U_\tau.$$

The U_1, U_2, \dots are independent random variables, each with distribution function F , but the U_i are not independent of τ . If we let \mathcal{F}_n denote the information in Y_1, \dots, Y_n and U_1, \dots, U_n , then τ is a stopping time with respect to $\{\mathcal{F}_n\}$ and U_{n+1}, U_{n+2}, \dots are independent of \mathcal{F}_n . If $\mathbb{E}(\tau) < \infty$, then Wald's equation (5.19) implies that

$$\mathbb{E}(S_1) = \mathbb{E}(U_i) \mathbb{E}(\tau). \quad (6.11)$$

It was shown in Exercise 5.14 that $\mathbb{E}(\tau) < \infty$ if $\mathbb{E}(Y_i) < 0$ and in this case another application of Wald's equation can be made to show that

$$\mathbb{E}(\tau) = -\frac{1}{\mathbb{E}(Y_i)}.$$

Let us compute $\mathbb{E}(Y_i)$. The probability that k people arrive in the queue during a service time U_i is

$$\begin{aligned} q_k &= \int_0^\infty \mathbb{P}\{k \text{ arrive} \mid U_i = s\} dF(s) \\ &= \int_0^\infty \frac{e^{-s\lambda} (s\lambda)^k}{k!} dF(s). \end{aligned}$$

The expected number of arrivals is therefore

$$\begin{aligned} \sum_{k=0}^{\infty} k q_k &= \int_0^\infty \sum_{k=0}^{\infty} k \frac{e^{-s\lambda} (s\lambda)^k}{k!} dF(s) \\ &= \int_0^\infty s\lambda dF(s) = \lambda\mu. \end{aligned}$$

Hence $\mathbb{E}(Y_i) = \lambda\mu - 1$ and

$$\mathbb{E}(\tau) = \frac{1}{1 - \lambda\mu} = \frac{\rho}{\rho - \lambda},$$

where we write $\rho = 1/\mu$ for the service rate. The expected length of a busy time is given by

$$\mathbb{E}(S_1) = \mathbb{E}(U_i) \mathbb{E}(\tau) = \frac{1}{\rho - \lambda}.$$

The fraction of time that the queue is busy is given by

$$\frac{\mathbb{E}(S_1)}{\mathbb{E}(R_1) + \mathbb{E}(S_1)} = \frac{\lambda}{\rho}.$$

Note that this ratio tends to 1 as $\lambda \rightarrow \rho$.

If $\lambda = \rho$, the chain Y_n can be shown to be recurrent (see Exercise 2.15) so that the queue size returns to 0 infinitely often. However, in the long run the fraction of time spent with the queue empty goes to 0. If $\lambda > \rho$, the chain Y_n is transient, and hence the queue size goes to infinity.

Now let us consider the somewhat less realistic $G/M/1$ queue. Here customers arrive one at a time with waiting times T_1, T_2, \dots having common distribution function F with mean $1/\lambda$. There is one server and the service times are exponential with rate ρ . We will assume that the service rate is greater than the arrival rate, $\rho > \lambda$.

There exists a natural Markov chain embedded in the $G/M/1$ queue. Consider Y_n the number of customers in the system immediately before the n th customer arrives. (We will assume that the queue starts out empty and we set $Y_0 = 0$.) Then Y_n can easily be checked to be a Markov chain with state space $\{0, 1, 2, \dots\}$.

To compute the transition probability for this chain we first for ease consider what happens if there are an infinite number of people in the queue. Let q_k be the probability that exactly k individuals are served between the arrival times of two successive customers. If the arrival time is t , then the number of customers served has a Poisson distribution with parameter ρt . Hence

$$\begin{aligned} q_k &= \int_0^\infty \mathbb{P}\{k \text{ served} \mid T_i = t\} dF(t) \\ &= \int_0^\infty e^{-\rho t} \frac{(\rho t)^k}{k!} dF(t). \end{aligned}$$

The expected number served is

$$\begin{aligned} \sum_{k=0}^{\infty} k q_k &= \sum_{k=0}^{\infty} k \int_0^\infty e^{-\rho t} \frac{(\rho t)^k}{k!} dF(t) \\ &= \int_0^\infty \left[\sum_{k=0}^{\infty} k e^{-\rho t} \frac{(\rho t)^k}{k!} \right] dF(t) \\ &= \int_0^\infty \rho t dF(t) \\ &= \rho/\lambda > 1. \end{aligned}$$

Now if $Y_n = j$, then after the n th customer arrives there will be $j + 1$ customers in the queue. The queue will serve customers until the queue empties. It is easy to see then that

$$\mathbb{P}\{Y_{n+1} = k \mid Y_n = j\} = q_{(j+1)-k}, \quad k = 1, \dots, j+1,$$

$$\mathbb{P}\{Y_{n+1} = 0 \mid Y_n = j\} = \sum_{k \leq 0} q_{(j+1)-k} = \sum_{i \geq j+1} q_i.$$

If we set $p_l = q_{1-l}$, $l = 1, 0, -1, \dots$, then we see that Y_n has transition probabilities

$$p(j, k) = p_{k-j}, \quad k = 1, \dots, j+1,$$

$$p(j, 0) = \sum_{k \leq 0} p_{k-j}.$$

It can be shown (see Exercise 2.16) that this is a positive recurrent Markov chain. Its invariant probability is of the form

$$\pi(j) = \beta^j (1 - \beta),$$

where β is the unique solution to

$$\beta = \sum_{j=0}^{\infty} q_j \beta^j,$$

with $\beta \in (0, 1)$. It is hard to evaluate β analytically but it can be computed numerically.

6.5 Exercises

6.1 Suppose the lifetime of a component T_i in hours is uniformly distributed on $[100, 200]$. Components are replaced as soon as one fails and assume that this process has been going on long enough to reach equilibrium.

(a) What is the probability that the current component has been in operation for at least 50 hours?

(b) What is the probability that the current component will last for at least 50 more hours?

(c) What is the probability that the total lifetime of the current component will be at least 150 hours?

(d) Suppose it is known that the current component has been in operation for exactly 90 hours. What is the probability that it will last at least 50 more hours?

6.2 Repeat Exercise 6.1 with the T_i exponentially distributed with mean 150.

6.3 Repeat Exercise 6.1 with the T_i having density

$$f(t) = \frac{1}{t \ln 2}, \quad 100 < t < 200.$$

6.4 Repeat Exercise 6.1 with the T_i having distribution

$$\mathbb{P}\{T_i = 100\} = \mathbb{P}\{T_i = 200\} = 1/2.$$

6.5 Let N_t denote the renewal process associated with independent, identically distributed random variables T_1, T_2, \dots with mean μ .

(a) Explain why for any positive integers j, k and any t , the following inequality holds

$$\mathbb{P}\{N_t \geq jk\} \leq [\mathbb{P}\{N_t \geq j\}]^k.$$

(b) The law of large numbers for renewal processes, (6.2), states that for every $\epsilon > 0$

$$\lim_{t \rightarrow \infty} \mathbb{P} \left\{ \frac{t(1-\epsilon)}{\mu} \leq N_t \leq \frac{t(1+\epsilon)}{\mu} \right\} = 1. \quad (6.12)$$

Use (a) and (6.12) to conclude that for every $\epsilon > 0$,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E} \left[N_t I \left\{ N_t > \frac{t(1+\epsilon)}{\mu} \right\} \right] = 0.$$

(c) Derive the first renewal theorem, (6.3).

6.6 Assume that the waiting times T_i have distribution

$$\mathbb{P}\{T_i = 1\} = \frac{9}{10}, \quad \mathbb{P}\{T_i = 10\pi\} = \frac{1}{10}.$$

Note that the times T_i have a nonlattice distribution.

(a) What is the age distribution $\Psi_C(n)$?

(b) For large times, what is the expected residual life? Compare to $\mathbb{E}(T_i)$.

6.7 Suppose that there are two brands of replacement components, Brand X and Brand Y, and that for political reasons a company buys replacements of both types. When a Brand X component fails it is replaced with a new Brand Y component and when a Brand Y component fails it is replaced with a Brand X component. The lifetimes (measured in thousands of hours) of Brand X components are uniform on $[1, 2]$ and the Brand Y components have

lifetimes that are uniform on $[1, 3]$. Answer the following questions for large time t .

- (a) What is the probability that the current component is Brand X?
- (b) What is the distribution of the age of the current component?
- (c) What is the distribution of the total lifetime of the current component?
- (d) Would these answers be different if instead of alternating the brands, they used the rule that when a component fails they randomly choose a Brand X or Brand Y component with probability $1/2$ for each?

6.8 Suppose customers arrive in a one-server queue according to a Poisson distribution with rate $\lambda = 1$ (in hours). Suppose that the service times equal $1/4$ hour, $1/2$ hour, or one hour each with probability $1/3$.

- (a) Assume that the queue is empty and a customer arrives. What is the expected amount of time until that customer leaves?
- (b) Assume that the queue is empty and a customer arrives. What is the expected amount of time until the queue is empty again?
- (c) At a large time t what is the probability that there are no customers in the queue?

6.9 Give an example of a renewal process with $\mathbb{E}[T_i] < \infty$ such that the large time residual life distribution has infinite mean.

6.10 Assume T_1, T_2, \dots are independent identically distributed nonnegative random variables with $\mathbb{P}\{T_i = 0\} = q \in (0, 1)$. Suppose the distribution function of the T_i is F with mean μ , and let G be the conditional distribution function of the T_i given that the $T_i > 0$,

$$G(x) = \mathbb{P}\{T_i \leq x \mid T_i > 0\} = \frac{F(x) - F(0)}{1 - q}.$$

Let $\tilde{T}_1, \tilde{T}_2, \dots$ be independent, identically distributed random variables with distribution function G and let $U(t)$ and $\tilde{U}(t)$ be the renewal functions associated with the T_i and the \tilde{T}_i respectively. Show that

$$\tilde{U}(t) = (1 - q)U(t).$$

Chapter 7

Reversible Markov Chains

7.1 Reversible Processes

In this chapter we will study a particular class of Markov chains, reversible chains. A large number of important chains are reversible, and we can take advantage of this fact in trying to understand their behavior.

Suppose we have a continuous-time Markov chain X_t taking values in state space S (finite or countably infinite) with transition rates $\alpha(x, y)$. If π is any measure on S , i.e., a nonnegative function on S , then the chain is said to be *reversible with respect to the measure π* if for all $x, y \in S$,

$$\pi(x) \alpha(x, y) = \pi(y) \alpha(y, x).$$

We will say that the chain is *symmetric* if for every x, y

$$\alpha(x, y) = \alpha(y, x).$$

Note that a chain is symmetric if and only if it is reversible with respect to the uniform measure $\pi(x) = 1, x \in S$. Similarly, a discrete-time Markov chain with transition matrix \mathbf{P} is said to be reversible with respect to π if

$$\pi(x) \mathbf{P}(x, y) = \pi(y) \mathbf{P}(y, x),$$

for all $x, y \in S$ and symmetric if $\mathbf{P}(x, y) = \mathbf{P}(y, x)$. In the next two sections we will discuss continuous-time chains, but analogous statements hold for discrete-time chains.

Example 1. Let $G = (V, E)$ be a graph as in Example 5, Section 1.1. Let $S = V$ and

$$\alpha(x, y) = \frac{1}{d(x)}, \quad (x, y) \in E,$$

where $d(x)$ is the number of vertices adjacent to x . This is a continuous-time analogue of Example 5. Then this chain is reversible with respect to the measure $\pi(x) = d(x)$. If instead we choose

$$\alpha(x, y) = 1, \quad (x, y) \in E,$$

then the chain is symmetric and hence reversible with respect to the uniform measure.

Example 2. Let $G = (V, E)$ be any graph and let $g : E \rightarrow [0, \infty)$. Such a configuration is often called a *network*. A network gives rise to a symmetric chain with transitions

$$\alpha(x, y) = \alpha(y, x) = g(e),$$

if e denotes the edge connecting x and y . In the study of electrical networks the rates $g(e)$ are called conductances and their reciprocals are called resistances.

Example 3. Suppose we have a birth-and-death chain on $S = \{0, 1, 2, \dots\}$ with birth rates λ_n and death rates μ_n . In other words, the transition rates are

$$\alpha(n, n+1) = \lambda_n, \quad \alpha(n, n-1) = \mu_n.$$

Let $\pi(0) = 1$ and for $n > 0$,

$$\pi(n) = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}.$$

Then the chain is reversible with respect to the measure π .

Example 4. Let $G = (V, E)$ be any graph and suppose $\pi : V \rightarrow (0, \infty)$ is a positive measure on G . Suppose each vertex is adjacent to only a finite number of other vertices. Define $\alpha(x, y) = 0$ if (x, y) is not an edge of G and for $(x, y) \in E$,

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}.$$

Then α generates a chain that is reversible with respect to π .

If a chain is reversible with respect to π , then

$$\sum_{y \in S} \pi(y) \alpha(y, x) = \pi(x) \sum_{y \in S} \alpha(x, y) = \pi(x) \alpha(x),$$

i.e., π is an invariant measure for α . If the state space is finite, or if the state space is infinite with $\sum \pi(x) < \infty$, then we can normalize π so that it is an invariant probability for α . In particular, if α is irreducible, we know that if α is reversible with respect to a probability measure π then π is the (unique) invariant measure. Conversely, if an irreducible chain is reversible with respect to a π with $\sum \pi(x) = \infty$, we can conclude that there is no invariant probability measure and hence the chain is null recurrent or transient.

The reversibility condition is a way of stating that the system in equilibrium looks the same whether time goes forward or backward. To give an easy

example of a nonreversible chain consider the three-state chain on $S = \{0, 1, 2\}$ with rates

$$\alpha(0, 1) = \alpha(1, 2) = \alpha(2, 0) = 1,$$

$$\alpha(1, 0) = \alpha(2, 1) = \alpha(0, 2) = 2.$$

This is clearly irreducible with invariant probability measure $\pi(0) = \pi(1) = \pi(2) = 1/3$. If the chain were to be reversible, it would need to be reversible with respect to π , but clearly

$$\pi(0)\alpha(0, 1) \neq \pi(1)\alpha(1, 0).$$

7.2 Convergence to Equilibrium

It is often useful to give estimates for the amount of time needed for the chain to reach a measure close to the invariant probability measure. Let X_t be an irreducible continuous-time Markov chain with rates $\alpha(x, y)$, reversible with respect to the probability measure π . We will assume that the state space is finite, $S = \{1, \dots, N\}$, but one can generalize these ideas to positive recurrent chains on an infinite state space. For ease, we will only consider the case where \mathbf{A} is symmetric (reversible with respect to the uniform measure), but these ideas hold for all reversible chains.

There are a number of ways to measure the “distance” between two probability measures π and ν on S . One very natural definition is the *total variation* distance defined by

$$\|\pi - \nu\|_{\text{TV}} = \max\{|\pi(A) - \nu(A)| : A \subset S\}.$$

It is easy to see that the maximum is obtained on the set $A = \{x : \pi(x) \geq \nu(x)\}$. Therefore,

$$\begin{aligned} \|\pi - \nu\|_{\text{TV}} &= \sum_{\pi(x) \geq \nu(x)} (\pi(x) - \nu(x)) \\ &= \frac{1}{2} \left[\sum_{\pi(x) \geq \nu(x)} (\pi(x) - \nu(x)) + \sum_{\pi(x) < \nu(x)} (\nu(x) - \pi(x)) \right] \\ &= \frac{1}{2} \sum_{x \in S} |\pi(x) - \nu(x)| \\ &= \frac{1}{2} \sum_{x \in S} \frac{1}{N} |N\pi(x) - N\nu(x)|. \end{aligned}$$

In the last expression, the $1/N$ represents the uniform measure on S and $N\pi, N\nu$ are the “derivatives” of π, ν with respect to this measure.

Another measure of distance which is not quite as natural but is sometimes easier to analyze is the L^2 or *mean-squared* distance,

$$\|\pi - \nu\|_{L^2} = \left[\sum_{x \in S} \frac{1}{N} |N\pi(x) - N\nu(x)|^2 \right]^{1/2}.$$

Note that $\|\pi - \nu\|_{L^2} = N^{1/2} \|\pi - \nu\|$ where $\|\cdot\|$ denotes the usual Euclidean norm in \mathbb{R}^N . The Cauchy-Schwartz inequality

$$|\bar{v} \cdot \bar{w}| \leq \|\bar{v}\|^{1/2} \|\bar{w}\|^{1/2},$$

gives the inequality

$$\|\pi - \nu\|_{L^2} \geq 2 \|\pi - \nu\|_{TV}.$$

Example 1. Consider the chain with rates $\alpha(i, j) = b/N, i \neq j$ where $b > 0$. For any i the vector \bar{v} with

$$v^j = \begin{cases} N-1, & i = j \\ -1, & i \neq j, \end{cases}$$

is a right eigenvector with eigenvalue $-b$. There is an $N-1$ dimensional subspace of such eigenvectors; hence the eigenvalues for A are 0 with multiplicity 1 and $-b$ with multiplicity $N-1$. If ν is any probability vector, we can give an exact expression for $e^{t\mathbf{A}}\nu$. Suppose we start in state x . This chain starts with distribution ν , waits for an exponential “alarm clock” with rate b (mean $1/b$) to ring, and then chooses one of the N sites from the uniform distribution. If we let π denote the uniform distribution, then

$$e^{t\mathbf{A}}\nu = e^{-tb} \nu + (1 - e^{-tb}) \pi.$$

The e^{-tb} term denotes the probability that the alarm clock has not gone off. Therefore,

$$\|e^{t\mathbf{A}}\nu - \pi\|_{TV} = e^{-tb} \|\nu - \pi\|_{TV} \leq e^{-tb},$$

$$\|e^{t\mathbf{A}}\nu - \pi\|_{L^2} = e^{-tb} \|\nu - \pi\|_{L^2}.$$

If the chain starts at x , so that $\nu(x) = 1$, then $\|\nu - \pi\|_{L^2} \sim \sqrt{N}$, so the L^2 distance is still large.

Despite its limitations, we will focus on bounding the rate of convergence in the L^2 -distance, because techniques of linear algebra can be used. If \mathbf{A} is a symmetric matrix, then it can be shown (see an advanced book on linear

algebra) that there is a complete set of eigenvalues and eigenvectors. Moreover, all the eigenvalues are real so we can write the eigenvalues in decreasing order,

$$0 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N.$$

We know $\lambda_2 < 0$ because the chain is irreducible. By symmetry, we see that if $\langle \cdot, \cdot \rangle$ denotes inner product,

$$\langle \mathbf{A}\bar{v}, \bar{w} \rangle = \langle \bar{v}, \mathbf{A}\bar{w} \rangle = \sum_{i=1}^N \sum_{j=1}^N v^i w^j \mathbf{A}(i, j). \quad (7.1)$$

A matrix satisfying the first equality is said to be self-adjoint (with respect to the uniform measure) and the expression on the right is often called the quadratic form associated with the matrix.

Let

$$\bar{1} = \bar{v}_1, \bar{v}_2, \dots, \bar{v}_N,$$

be the eigenvectors for \mathbf{A} , which are both right and left eigenvectors since \mathbf{A} is symmetric. Using (7.1) we can see that

$$\lambda_j \langle \bar{v}_j, \bar{v}_k \rangle = \langle \mathbf{A}\bar{v}_j, \bar{v}_k \rangle = \langle \bar{v}_j, \mathbf{A}\bar{v}_k \rangle = \lambda_k \langle \bar{v}_j, \bar{v}_k \rangle,$$

and hence eigenvectors for different eigenvalues are orthogonal ($\langle \bar{v}_j, \bar{v}_k \rangle = 0$). We can therefore choose the eigenvectors so they are all orthogonal. These eigenvectors are also the eigenvectors for the matrix $e^{t\mathbf{A}}$ with corresponding eigenvalues $e^{t\lambda_j}$,

$$e^{t\mathbf{A}}\bar{v}_j = e^{t\lambda_j}\bar{v}_j.$$

Let $U \subset R^N$ denote the $N - 1$ dimensional subspace generated by the vectors $\{\bar{v}_2, \dots, \bar{v}_N\}$, or equivalently, the set of vectors \bar{w} satisfying

$$\sum_{i=1}^N w^i = 0.$$

By writing any $\bar{w} \in U$ as a linear combination of left eigenvectors, we can easily see that

$$\|\bar{w}e^{t\mathbf{A}}\| \leq e^{t\lambda_2}\|\bar{w}\|,$$

where $\|w\|^2 = \sum_{i=1}^N [w^i]^2$. Now suppose we start the chain with any probability vector $\bar{\nu}$. We can write

$$\bar{\nu} = \bar{\pi} + \bar{w},$$

where $\bar{\pi} = (1/N)\bar{1}$ is the invariant probability and $\bar{w} = \bar{\nu} - \bar{\pi} \in U$. Since $\bar{\pi}e^{t\mathbf{A}} = \bar{\pi}$, we can conclude

$$\|\bar{\nu}e^{t\mathbf{A}} - \bar{\pi}\|_{L^2} = \|(\bar{\nu} - \bar{\pi})e^{t\mathbf{A}}\|_{L^2} \leq e^{t\lambda_2} \|\bar{\nu} - \bar{\pi}\|_{L^2}.$$

What we see is that the rate of convergence is essentially controlled by the size of λ_2 , and if we can get lower bounds on $|\lambda_2|$, we can bound the rate of convergence.

Example 2. Consider simple random walk on a circle, i.e., the chain with state space $S = \{1, \dots, N\}$ and rates $\alpha(x, y) = 1/2$ if $|x - y| = 1 \pmod{N}$. This is reversible with respect to the uniform measure on S . The eigenvalues for \mathbf{A} can be found exactly in this case (see Exercise 7.9),

$$\lambda_j = \cos\left(\frac{(j-1)2\pi}{N}\right) - 1, \quad j = 1, 2, \dots, N.$$

In particular, $\lambda_2 = \cos(2\pi/N) - 1$ which for large N (by the Taylor series for cosine) looks like $-2\pi^2 N^{-2}$. This says that it takes on the order of about N^2 time units in order for the distribution to be within e^{-1} of the uniform distribution. It makes sense that it takes on order N^2 steps to get close to equilibrium, if we remember that it takes a random walker on the order of N^2 steps to go a distance of about N .

Example 3. Let the state space S be all binary sequences of length N , i.e., all N -tuples (a_1, \dots, a_N) , $a_i \in \{0, 1\}$. Note that the state space has 2^N elements. Consider the chain with $\alpha(x, y) = 1$ if x and y are two sequences that differ in exactly one component and $\alpha(x, y) = 0$ otherwise. This is sometimes called random walk on the N -dimensional hypercube. Clearly this is reversible with respect to the uniform measure. It can be shown that $-2j/N$ is an eigenvalue with multiplicity $\binom{N}{j}$. In this case, $\lambda_2 = -2/N$ and it takes on order N steps to get close to equilibrium. This can be understood intuitively by noting that if the number of steps is of order N , most components have had an opportunity to change at least once.

Now let U be the set of vectors that are orthogonal to $\bar{1}$, i.e., the set of vectors \bar{w} satisfying

$$\sum_{i=1}^N w^i = 0.$$

If $\bar{w} \in U$, then $\mathbf{A}\bar{w} \in U$. If we write

$$\bar{w} = a_2\bar{v}_2 + \cdots + a_n\bar{v}_n,$$

with $a_i = \langle \bar{v}_i, \bar{w} \rangle$, we see that

$$\begin{aligned}\langle \bar{w}, \mathbf{A}\bar{w} \rangle &= \sum_{i=2}^N \sum_{j=2}^N \langle a_i \bar{v}_i, a_j \mathbf{A} \bar{v}_j \rangle \\ &= \sum_{i=2}^N \sum_{j=2}^N a_i a_j \lambda_j \langle \bar{v}_i, \bar{v}_j \rangle \\ &= \sum_{i=2}^N a_i^2 \lambda_i \langle \bar{v}_i, \bar{v}_i \rangle \\ &\leq \lambda_2 \sum_{i=2}^n \langle a_i \bar{v}_i, a_i \bar{v}_i \rangle = \lambda_2 \langle \bar{w}, \bar{w} \rangle.\end{aligned}$$

Also, we get equality in the above expression if we choose $\bar{w} = \bar{v}_2$. What we have derived is the Rayleigh–Ritz variational formulation for the second eigenvalue,

$$\lambda_2 = \sup \frac{\langle \bar{w}, \mathbf{A}\bar{w} \rangle}{\langle \bar{w}, \bar{w} \rangle},$$

where the supremum is taken over all vectors \bar{w} with

$$\langle \bar{1}, \bar{w} \rangle = \sum_{i=1}^N w^i = 0.$$

Lower bounds for λ_2 (i.e., upper bounds of $|\lambda_2|$) can be obtained by considering particular $\bar{w} \in U$. If $T \subset S$, let $\bar{w} \in U$ with components

$$w^i = \begin{cases} 1 - \pi(T), & i \in T \\ -\pi(T), & i \notin T, \end{cases}$$

$$\pi(T) = \frac{\text{number of elements in } T}{N}.$$

Note that $\langle \bar{1}, \bar{w} \rangle = 0$ and

$$\begin{aligned}\langle \bar{w}, \bar{w} \rangle &= \sum_{i \in T} [1 - \pi(T)]^2 + \sum_{i \notin T} \pi(T)^2 \\ &= [1 - \pi(T)]^2 N \pi(T) + \pi(T)^2 N [1 - \pi(T)] = N \pi(T) [1 - \pi(T)].\end{aligned}$$

If $i \in T$,

$$\begin{aligned}
(\mathbf{A}\bar{w})^i &= \sum_j \mathbf{A}_{ij} w^j \\
&= -\alpha(i)[1 - \pi(T)] + \sum_{j \in T, j \neq i} \alpha(j, i)[1 - \pi(T)] - \sum_{j \notin T} \alpha(j, i)\pi(T) \\
&= -\sum_{j \notin T} \alpha(j, i)[1 - \pi(T)] - \sum_{j \notin T} \alpha(j, i)\pi(T) \\
&= -\sum_{j \notin T} \alpha(j, i).
\end{aligned}$$

Similarly, if $i \notin T$,

$$(\mathbf{A}w)^i = \sum_{j \in T} \alpha(j, i).$$

Therefore,

$$\begin{aligned}
\langle \bar{w}, \mathbf{A}\bar{w} \rangle &= \sum_i w^i (\mathbf{A}\bar{w})^i \\
&= \sum_{i \in T} [1 - \pi(T)] \sum_{j \notin T} [-\alpha(j, i)] + \sum_{i \notin T} [-\pi(T)] \sum_{j \in T} \alpha(j, i) \\
&= -\sum_{i \in T} \sum_{j \notin T} \alpha(j, i).
\end{aligned}$$

Define κ by

$$\kappa = \inf_{T \subset S} \frac{\sum_{i \in T} \sum_{j \notin T} \alpha(i, j)\pi(i)}{\pi(T)[1 - \pi(T)].}$$

Then by considering this choice of \bar{w} in the Rayleigh–Ritz formulation, we have

$$|\lambda_2| \leq \kappa.$$

Unfortunately this bound is often not very good. A large area of research is concerned with finding better ways to estimate λ_2 ; we do not discuss this any further in this book.

7.3 Markov Chain Algorithms

A recent application of Markov chain theory has been in Monte Carlo simulations of random systems. The idea of Monte Carlo simulations is simple: to

understand a random system one does many trials on a computer and sees how it behaves. These simulations always use a random number generator, generally a function that produces independent numbers distributed uniformly between 0 and 1. (Actually, a computer can only produce pseudo-random numbers and there are important questions as to whether pseudo-random number generators are “random” enough. We will not worry about that question here and will just assume that we have a means to generate independent identically distributed numbers U_1, U_2, \dots distributed uniformly on $[0, 1]$.)

As an example, suppose we were interested in studying properties of “random” matrices whose entries are 0s and 1s. As a probability space we could choose the set S of $N \times N$ matrices \mathbf{M} , with

$$\mathbf{M}(i, j) = 0 \text{ or } 1, \quad 1 \leq i, j \leq N.$$

A natural probability measure would be the uniform measure on all 2^{N^2} such matrices. Writing an algorithm to produce a random matrix from this distribution is easy—choose N^2 uniform random numbers $U(i, j)$, $1 \leq i, j \leq N$, and set

$$\mathbf{M}(i, j) = \begin{cases} 0, & \text{if } U(i, j) < .5 \\ 1, & \text{if } U(i, j) \geq .5 \end{cases}.$$

It takes on the order of N^2 operations to produce one $N \times N$ matrix, and clearly every matrix in S has the same chance of being produced.

Now suppose we change our probability space and say we are only interested in matrices in S that have no two 1s together. Let T be the matrices in S with no two 1s together, i.e., the matrices $\mathbf{M} \in S$ such that

$$\mathbf{M}(i - 1, j) = \mathbf{M}(i + 1, j) = \mathbf{M}(i, j - 1) = \mathbf{M}(i, j + 1) = 0,$$

if $\mathbf{M}(i, j) = 1$. Suppose also we want to put the uniform probability measure on T (this is a natural measure from the perspective of statistical physics where 1s can denote particles and there is a repulsive interaction that keeps particles from getting too close together). While it is easy to define this measure, it is a hard problem to determine $c(N)$, the number of elements of T . It can be shown that there is a constant $\beta \in (1, 2)$ such that

$$\lim_{N \rightarrow \infty} c(N)^{1/N^2} = \beta$$

(so that the number of elements in T is approximately β^{N^2}) but the exact value of β is not known. Still we might be interested in the properties of such matrices and hence would like to sample from the uniform distribution on T .

While it is very difficult to give an efficient algorithm that exactly samples from the uniform distribution (and even if we had one, the errors in the random number generation would keep it from being an exact sampling), we can give a very efficient algorithm that produces samples from an almost

uniform distribution. What we do is run an irreducible Markov chain with state space T whose invariant measure is the uniform distribution. We can then start with any matrix in T ; run the chain long enough so that the chain is near equilibrium; and then choose the matrix we have at that point.

For this example, one algorithm is as follows: 1) start with any matrix $\mathbf{M} \in T$, e.g., the matrix with all zero entries; 2) choose one of the entries at random, i.e., choose an ordered pair (i, j) from the uniform distribution on the N^2 ordered pairs; and 3) consider the matrix gotten by changing only the (i, j) entry of \mathbf{M} . If this new matrix is in T , we let this be the new value of the chain; if the new matrix is not in T , we make no change in the value of the chain; return to 2). This algorithm is a simulation of the discrete-time Markov chain with state space T and transition probabilities

$$\mathbf{P}(\mathbf{M}, \mathbf{M}') = N^{-2},$$

if $\mathbf{M}, \mathbf{M}' \in T$ differ in exactly one entry; $\mathbf{P}(\mathbf{M}, \mathbf{M}') = 0$ if \mathbf{M} and \mathbf{M}' differ by more than one entry; and $\mathbf{P}(\mathbf{M}, \mathbf{M})$ is whatever is necessary so that the rows add up to 1. Clearly, \mathbf{P} is a symmetric matrix and it is not too difficult to see that it is irreducible. Hence \mathbf{P} is a reversible Markov chain with state space T and its invariant distribution is the uniform measure.

Of course, we need to know how long to run the chain in order to guarantee that one is close to the invariant distribution. As noted in the previous section, this boils down to estimating the second eigenvalue for the Markov chain. Unfortunately, estimating this eigenvalue is often much more difficult than showing that the chain has the right invariant measure (which is quite easy in this example). In this example, we clearly need at least N^2 steps to get close, since each of the entries should have a good chance to be changed.

We will give some other examples of where these kinds of algorithms have been used. In all of these cases the algorithms are fairly efficient, although in some cases only partial rigorous analysis has been given.

Example 1. Ising Model. Let S be the set of $N \times N$ matrices with entries 1 or -1 . For any $\mathbf{M} \in S$ we define the “energy” of the matrix by

$$H(\mathbf{M}) = - \sum_{(i,j) \sim (i',j')} \mathbf{M}(i,j) \mathbf{M}(i',j'),$$

where $(i, j) \sim (i', j')$ if the entries are “nearest neighbors,”

$$|i - i'| + |j - j'| = 1.$$

The value $\mathbf{M}(i, j)$ is called the “spin” at site (i, j) and the energy is minimized when all the spins are the same. The Ising model gives a probability distribution on S that weights matrices of low energy the highest. For any $a > 0$ we let

$$\pi_a(\mathbf{M}) = \frac{\exp\{-aH(\mathbf{M})\}}{\sum_{\mathbf{M}' \in S} \exp\{-aH(\mathbf{M}')\}}.$$

This is a well-defined probability measure, although it is difficult to calculate the normalization factor

$$Z(a) = \sum_{\mathbf{M}' \in S} \exp\{-aH(\mathbf{M}')\}.$$

If \mathbf{M} and \mathbf{M}' are two matrices that agree in all but one entry, we can calculate $\pi_a(\mathbf{M})/\pi_a(\mathbf{M}')$ easily without calculating $Z(a)$.

Write $\mathbf{M} \sim \mathbf{M}'$ if \mathbf{M} and \mathbf{M}' differ in exactly one entry. We define \mathbf{P}_a by

$$\mathbf{P}_a(\mathbf{M}, \mathbf{M}') = \frac{1}{N^2} \min \left\{ 1, \frac{\pi_a(\mathbf{M}')}{\pi_a(\mathbf{M})} \right\}, \quad \mathbf{M} \sim \mathbf{M}'$$

and

$$\mathbf{P}_a(\mathbf{M}, \mathbf{M}) = 1 - \frac{1}{N^2} \sum_{\mathbf{M}' \sim \mathbf{M}} \mathbf{P}_a(\mathbf{M}, \mathbf{M}').$$

In other words, one runs an algorithm as follows: 1) start with a matrix \mathbf{M} ; 2) choose an entry of the matrix at random and let \mathbf{M}' be the matrix which agrees with \mathbf{M} everywhere except at that entry; 3) move to matrix \mathbf{M}' with probability $\min\{1, \pi_a(\mathbf{M}')/\pi_a(\mathbf{M})\}$ and otherwise stay at the matrix \mathbf{M} . It is easy to check that this is an irreducible Markov chain reversible with respect to π_a .

Example 2. The above example is a specific case of a general algorithm. Suppose $G = (V, E)$ is a connected graph such that each vertex is adjacent to at most K other vertices. Suppose a positive function f on V is given, and let π be the probability measure

$$\pi(v) = \frac{f(v)}{\sum_{w \in V} f(w)}.$$

Write $v \sim w$ if $(v, w) \in E$ and set

$$\mathbf{P}(v, w) = \frac{1}{K} \min \left\{ 1, \frac{f(w)}{f(v)} \right\}, \quad v \sim w,$$

and

$$\mathbf{P}(v, v) = 1 - \sum_{w \sim v} \mathbf{P}(v, w).$$

Then \mathbf{P} is an irreducible Markov chain, reversible with respect to π . Algorithms of this type are often referred to as Metropolis algorithms.

Example 3. There is another class of algorithms, called Gibbs samplers, which are similar. Suppose we have n variables (x_1, \dots, x_n) each of which can take on one of K values say $\{a_1, \dots, a_K\}$. Let S be the set of K^n possible

n -tuples and assume we have a positive function f on S . We want to sample from the distribution

$$\pi(x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n) \in S} f(y_1, \dots, y_n)}. \quad (7.2)$$

Our algorithm is to choose a $j \in \{1, \dots, n\}$ at random and then change x_j to z according to the conditional probability

$$\frac{f(x_1, \dots, x_{j-1}, z, x_{j+1}, \dots, x_n)}{\sum_{k=1}^K f(x_1, \dots, x_{j-1}, a_k, x_{j+1}, \dots, x_n)}.$$

This gives the transition probability

$$\mathbf{P}((x_1, \dots, x_n), (y_1, \dots, y_n)) =$$

$$\frac{1}{n} \frac{f(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)}{\sum_{k=1}^K f(x_1, \dots, x_{j-1}, a_k, x_{j+1}, \dots, x_n)}, \quad y_j \neq x_j; \quad y_i = x_i, i \neq j,$$

and $\mathbf{P}((x_1, \dots, x_n), (x_1, \dots, x_n))$ equal to whatever is necessary to make the rows sum to 1. Again it is straightforward to check that this is an irreducible Markov chain, reversible with respect to π . Note also that to run the chain we never need to calculate the denominator in (7.2).

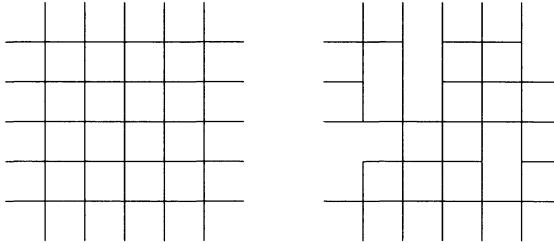
The Ising model can be considered one example with $n = N^2$, $K = 2$, and the possible values $-1, 1$. In this case we get

$$\mathbf{P}(\mathbf{M}, \mathbf{M}') = \frac{1}{N^2} \frac{\exp\{-aH(\mathbf{M}')\}}{\exp\{-aH(\mathbf{M})\} + \exp\{-aH(\mathbf{M}')\}},$$

if \mathbf{M} and \mathbf{M}' differ in exactly one entry.

7.4 A Criterion for Recurrence

In this section we develop a useful monotonicity result for random walks with symmetric rates. To illustrate the usefulness of the result consider two possible rates on \mathbb{Z}^2 . The first is $\alpha(x, y) = 1$ if $|x - y| = 1$ and 0 otherwise. This corresponds to simple random walk which we have already seen is recurrent in two dimensions. For the other rate, suppose we remove some edges from the integer lattice as illustrated below. More precisely, suppose we have a subset B of the edges of the lattice and state that $\alpha(x, y) = 1$ only if the edge (x, y) is contained in B .



What our result will say is that for *any* such subset B the corresponding chain is still recurrent. Assume we have a graph $G = (V, E)$ and two symmetric rate functions α and β on E .

Fact. *If α produces a recurrent chain and $\beta(x, y) \leq \alpha(x, y)$ for all (x, y) , then β also produces a recurrent chain.*

The proof of this statement takes a little work. We start with some preliminary remarks. Suppose we write the elements of V as $\{x_0, x_1, x_2, \dots\}$ (we will assume V is infinite, for otherwise the chains are always recurrent). Let $A_n = \{x_0, x_n, x_{n+1}, \dots\}$. Let us start the chain at x_0 , wait until it leaves x_0 for the first time, and then see what point in A_n is hit first by the chain. Let $h_n(x_0) = h_n(x_0; \alpha)$ be the probability that the first such point hit is not x_0 (using transition rates α). Then it is not too difficult to convince oneself that the chain is recurrent if and only if

$$\lim_{n \rightarrow \infty} h_n(x_0) = 0. \quad (7.3)$$

It is the goal of this section to give a formulation of $h_n(x_0)$ that will allow us to conclude the monotonicity result.

For this section we will assume that a graph $G = (V, E)$ is given as well as a symmetric transition rate $\alpha : E \rightarrow [0, \infty)$. Let A be a subset of V and fix $x_0 \in A$. Let X_t be a continuous time Markov chain with rates α and let τ be the infimum of all $t \geq 0$ such that $X_t \in A$. Define $f(y)$ to be the probability starting at y that the first visit to A occurs at the point x_0 ,

$$f(y) = \mathbb{P}\{X_\tau = x_0 \mid X_0 = y\}.$$

It is easy to see that $f(x_0) = 1$ and $f(y) = 0$ for $y \in A$, $y \neq x_0$. Suppose $y \notin A$. Then the probability that the first new site that y visits is z is

$\alpha(y, z)/\alpha(y)$, where again we write $\alpha(y) = \sum_{z \in V} \alpha(y, z)$. By concentrating on this first move, we see that

$$\begin{aligned} f(y) &= \sum_{z \in V} \mathbb{P}\{\text{first new site is } z\} f(z) \\ &= \sum_{z \in V} \frac{\alpha(y, z)}{\alpha(y)} f(z), \end{aligned}$$

or

$$\alpha(y)f(y) = \sum_{z \in V} \alpha(y, z)f(z). \quad (7.4)$$

A function f satisfying (7.4) is called α -harmonic at y . We have shown that our given f is α -harmonic at all $y \notin A$, and one can show with a little more work that f is the unique function that is α -harmonic at $y \notin A$ and that satisfies the boundary condition $f(x_0) = 1$, $f(y) = 0$, $y \in A, y \neq x$.

We will now characterize f as the function that minimizes a particular functional (a functional is a real-valued function of a function). For any function g let

$$Q_\alpha(g) = \sum_{x \in V} \sum_{y \in V} \alpha(x, y)(g(x) - g(y))^2.$$

Suppose we consider only those functions g that satisfy the boundary condition $g(x_0) = 1$, $g(y) = 0$, $y \in A, y \neq x$. Let \bar{g} be the function satisfying this boundary condition which minimizes Q_α . Then at any $y \notin A$, perturbations of \bar{g} at y , leaving all other values fixed, should increase Q_α . In other words if we define $\bar{g}_\epsilon(z)$ by

$$\bar{g}_\epsilon(z) = \begin{cases} \bar{g}(z), & z \neq y, \\ \bar{g}(y) + \epsilon, & z = y, \end{cases}$$

Then

$$\left. \frac{d}{d\epsilon} Q_\alpha(\bar{g}_\epsilon) \right|_{\epsilon=0} = 0.$$

A simple calculation shows that this holds if and only if for every $y \notin A$,

$$\sum_{z \in V} \bar{g}(z)\alpha(y, z) = \sum_{z \in V} \bar{g}(y)\alpha(y, z) = \bar{g}(y)\alpha(y).$$

In other words \bar{g} is the function that is α -harmonic at each $y \notin A$ and satisfies the boundary conditions. Since f is the only such function, $\bar{g} = f$. Summarizing, f , as defined above, is also the function that minimizes $Q_\alpha(g)$ subject to the boundary condition, $g(x_0) = 1$, $g(y) = 0$, $y \in A, y \neq x$.

We now use “summation by parts” to give another expression for $Q_\alpha(f)$. We start by writing

$$\begin{aligned} Q_\alpha(f) &= \sum_{x \in V} \sum_{y \in V} \alpha(x, y)(f(x) - f(y))^2 \\ &= \sum_x \sum_y \alpha(x, y)f(x)(f(x) - f(y)) \\ &\quad - \sum_x \sum_y \alpha(x, y)f(y)(f(x) - f(y)) \\ &= 2 \sum_x \sum_y \alpha(x, y)f(x)(f(x) - f(y)). \end{aligned}$$

The last equality uses the symmetry of α . Since $f(x_0) = 1$ and $f(y) = 0$, $y \in A, y \neq x$ we can write this as

$$2 \sum_y \alpha(x_0, y)(1 - f(y)) + 2 \sum_{x \notin A} f(x) \sum_y \alpha(x, y)(f(x) - f(y)).$$

But, if $x \notin A$, then f is α -harmonic at x ,

$$\sum_y \alpha(x, y)f(y) = \sum_y \alpha(x, y)f(x) = \alpha(x)f(x).$$

Hence the second term in the sum is 0 and we get

$$Q_\alpha(f) = 2 \sum_{y \in V} \alpha(x_0, y)(1 - f(y)) = 2\alpha(x_0) \sum_{y \in V} \frac{\alpha(x_0, y)}{\alpha(x_0)}(1 - f(y)).$$

Now let $h(x_0)$ be the probability that the chain starting at x_0 makes its first visit to A , after leaving x_0 for the first time, at some point other than x_0 . By considering the first step, we see that

$$h(x_0) = \sum_{y \in V} \frac{\alpha(x_0, y)}{\alpha(x_0)}(1 - f(y)) = \frac{Q_\alpha(x_0, A)}{2\alpha(x_0)},$$

where

$$Q_\alpha(x_0, A) = \inf \sum_{x \in V} \sum_{y \in V} \alpha(x, y)(g(x) - g(y))^2,$$

and the infimum is taken over all functions g satisfying $g(x_0) = 0$ and $g(y) = 1$, $y \in A, y \neq x$. The beauty of the formula comes in the realization that if $\beta(x, y)$ is another collection of rates with $\beta(x, y) \leq \alpha(x, y)$ for all $x, y \in V$, then

$$Q_\beta(x_0, A) \leq Q_\alpha(x_0, A).$$

If we now write $h_n(x_0; \alpha)$ and $h_n(x_0, \beta)$ as in the beginning of this section, we see that we have shown that if $\beta(x, y) \leq \alpha(x, y)$ for all x, y ,

$$h_n(x_0; \beta) \leq \frac{\beta(x_0)}{\alpha(x_0)} h_n(x_0; \alpha).$$

In particular, if we use the criterion given in (7.3), we see that if the chain with rates α is recurrent, then the chain with rates β is also recurrent.

7.5 Exercises

7.1 Show that every irreducible, discrete-time, two-state Markov chain is reversible with respect to its invariant probability.

7.2 Suppose X_t is a continuous-time Markov chain with state space $S = \{1, \dots, N\}$ and symmetric rates α .

(a) Show that for all t and all x ,

$$\mathbb{P}\{S_t = x \mid S_0 = x\} \geq \frac{1}{N}. \quad (7.5)$$

(Hint: write

$$\mathbb{P}\{S_t = x \mid S_0 = x\} = \sum_{y \in S} \{S_{t/2} = y \mid S_0 = x\}^2.)$$

(b) Give an example of a non-symmetric chain whose invariant probability distribution is uniform such that (7.5) does not hold for some $x \in S, t > 0$.

7.3 Let X_n be an aperiodic, discrete-time Markov chain on $S = \{1, \dots, N\}$ whose transition probability is symmetric. Show that for all $x \in S$ and all integers n ,

$$\mathbb{P}\{S_{2n} = x \mid S_0 = x\} \geq \frac{1}{N}.$$

Does this hold if $2n$ is replaced with $2n + 1$?

7.4 Let X_t be the continuous-time simple random walk on a circle as in Example 2, Section 7.2. Show that there exists a $c > 0$, independent of N , such that for all $x, y \in \{1, \dots, N\}$ and all $t \geq N^2$,

$$\mathbb{P}\{S_t = x \mid S_0 = y\} \geq \frac{c}{N^2}.$$

(Hint: (7.5) may be helpful.)

7.5 Let X_t be an aperiodic Markov chain with state space $S = \{1, \dots, N\}$ with rates α and invariant probability π . For every $0 < \epsilon < 1$, let T_ϵ be the infimum of all $t > 0$ such that for every $x, y \in S$,

$$\mathbb{P}\{X_t = x \mid X_0 = y\} \geq \epsilon \pi(x). \quad (7.6)$$

- (a) Explain why $T_\epsilon < \infty$ for every $0 < \epsilon < 1$.
- (b) Show that (7.6) holds for all $t > T_\epsilon$.
- (c) Show that if $0 < \epsilon < 1$ and k is a positive integer,

$$T_{1-(1-\epsilon)^k} \leq k T_\epsilon.$$

(d) Let X_t be the continuous-time simple random walk on a circle as in Example 2, Section 7.2. Show that there exist a $c, \beta > 0$, independent of N such that for all initial probability distributions ν and all $t > 0$,

$$\|e^{t\mathbf{A}}\nu - \pi\|_{\text{TV}} \leq c e^{-\beta t/N^2},$$

where π denotes the uniform distribution.

7.6 COMPUTER SIMULATION. Let \mathbf{M} be a matrix chosen uniformly from the set of 50×50 matrices with entries 0 and 1 such that no two 1s are together (see Section 7.3). Use a Markov chain simulation as described in Section 7.3 to estimate the probability that the $\mathbf{M}(25, 25)$ entry of this matrix is a 1.

7.7 COMPUTER SIMULATION. Let S_n be the set of finite sequences of numbers (k_0, k_1, \dots, k_n) where each $k_j \in \{0, 1\}$ and no two 1s are adjacent, i.e., $k_j + k_{j-1} \leq 1$ for $j = 1, \dots, n$. Let $p_n(j)$ denote the fraction of such sequences with $k_j = 1$. Do a Markov chain simulation similar to the previous exercise to estimate $p_{200}(0), p_{200}(100)$.

7.8 In this exercise, we will calculate the values of $p_n(j)$ in Exercise 7.7 exactly. Let $r_n(ij)$ denote the number of sequences in S_n with $k_0 = i, k_n = j$.

- (a) Explain why

$$r_{n+1}(00) = r_n(00) + r_n(01), \quad r_{n+1}(01) = r_n(00),$$

and give similar equations for $r_{n+1}(10), r_{n+1}(11)$.

(b) Use these equations to find $r_n(00), r_n(01), r_n(10), r_n(11)$. (Hint: see Exercise 0.3.)

- (c) Find $p_n(j)$.

7.9 Find the eigenvalues of the $N \times N$ matrix \mathbf{A} from Example 2, Section 7.2,

$$\mathbf{A}(i, j) = \begin{cases} -1, & i = j, \\ 1/2, & |i - j| = 1(\text{mod } N), \\ 0, & \text{otherwise.} \end{cases}$$

[Hint: any eigenvector with eigenvalue λ can be considered as a function $f(n)$ on the integers satisfying

$$\lambda f(n) = \frac{1}{2}f(n+1) + \frac{1}{2}f(n-1) - f(n),$$

$$f(n) = f(n+N),$$

for each n . Find the general solution of the difference equation and then use the periodicity condition to put restrictions on the λ .]

7.10 Let $\alpha(x, y)$ be a symmetric rate function on the edges of the integer lattice \mathbb{Z}^d , i.e., a nonnegative function defined for all $x, y \in \mathbb{Z}^d$ with $|x-y|=1$ that satisfies $\alpha(x, y) = \alpha(y, x)$. Suppose there exist numbers $0 < c_1 < c_2 < \infty$ such that for all x, y with $|x-y|=1$,

$$c_1 \leq \alpha(x, y) \leq c_2.$$

Let X_t be a continuous-time Markov chain with rates $\alpha(x, y)$.

- (a) If $d = 1, 2$, show that the chain is recurrent.
- (b) If $d \geq 3$, show that the chain is transient.

Chapter 8

Brownian Motion

8.1 Introduction

Brownian motion is a stochastic process that models random continuous motion. In order to model “random continuous motion,” we start by writing down the physical assumptions that we will make. Let X_t represent the position of a particle at time t . In this case t takes on values in the nonnegative real numbers and X_t takes on values in the real line (or perhaps the plane or space). This will be an example of a stochastic process with both continuous time and continuous state space.

For ease we will start with the assumption $X_0 = 0$. The next assumption is that the motion is “completely random.” Consider two times $s < t$. We do not wish to say that the positions X_s and X_t are independent, but rather that the motion after time s , $X_t - X_s$, is independent of X_s . We will need this assumption for any finite number of times: for any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, the random variables $X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \dots, X_{t_n} - X_{s_n}$ are independent. Also the distribution of the random movements should not change with time. Hence we will assume that the distribution of $X_t - X_s$ depends only on $t - s$. For the time being, we will also assume that there is no “drift” to the process, i.e., $\mathbb{E}(X_t) = 0$.

The above assumptions are not sufficient to describe the model we want. In fact, if Y_t is the Poisson process and $X_t = Y_t - t$ [so that $\mathbb{E}(X_t) = 0$], X_t satisfies these assumptions but is clearly not a model for continuous motion. We will include as our final assumption for our model this continuity: the function X_t is a continuous function of t .

It turns out that the above assumptions uniquely describe the process at least up to a scaling constant. Suppose the process X_t satisfies these assumptions. What is the distribution of the random variable X_t ? For ease, we will discuss the case $t = 1$. For any n , we can then write

$$X_1 = [X_{1/n} - X_0] + [X_{2/n} - X_{1/n}] + \dots + [X_{n/n} - X_{(n-1)/n}].$$

In other words, X_1 can be written as the sum of n independent, identically distributed random variables. Moreover, if n is large, each of the random variables is small. To be more precise, if we let

$$M_n = \max\{|X_{1/n} - X_0|, |X_{2/n} - X_{1/n}|, \dots, |X_{n/n} - X_{(n-1)/n}|\},$$

then as $n \rightarrow \infty$, $M_n \rightarrow 0$. This is a consequence of the assumption that X_t is a continuous function of t (if M_n did not go to 0 then there would be a “jump” in the path of X_t). It is a theorem of probability theory that the only distribution that can be written as the sum of n independent, identically distributed random variables such that the maximum of the variables goes to 0 is a normal distribution. We can thus conclude that the distribution of X_t is a normal distribution. We now formalize this definition.

Definition. A *Brownian motion* or a *Wiener process with variance parameter σ^2* is a stochastic process X_t taking values in the real numbers satisfying

- (i) $X_0 = 0$;
- (ii) For any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, the random variables $X_{t_1} - X_{s_1}, \dots, X_{t_n} - X_{s_n}$ are independent;
- (iii) For any $s < t$, the random variable $X_t - X_s$ has a normal distribution with mean 0 and variance $(t - s)\sigma^2$;
- (iv) The paths are continuous, i.e., the function $t \mapsto X_t$ is a continuous function of t .

While it is standard to include the fact that the increments are normally distributed in the definition, it is worth remembering that this fact can actually be deduced from the physical assumptions. *Standard Brownian motion* is a Brownian motion with $\sigma^2 = 1$. We can also speak of a Brownian motion starting at x ; this is a process satisfying conditions (ii) through (iv) and the initial condition $X_0 = x$. If X_t is a Brownian motion (starting at 0), then $Y_t = X_t + x$ is a Brownian motion starting at x .

Brownian motion can be constructed as a limit of random walks. Suppose S_n is an unbiased random walk on the integers. We can write

$$S_n = Y_1 + \dots + Y_n,$$

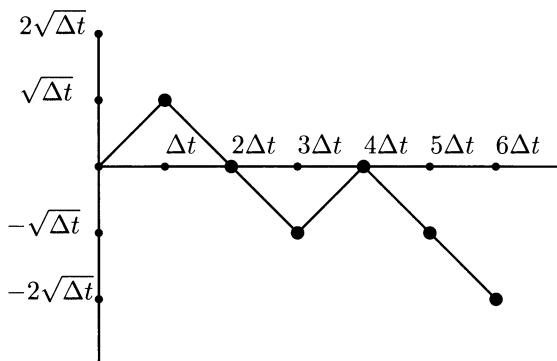
where the random variables Y_i are independent,

$$\mathbb{P}\{Y_i = 1\} = \mathbb{P}\{Y_i = -1\} = \frac{1}{2}.$$

Now instead of having time increments of size 1 we will have increments of size $\Delta t = 1/N$ where N is an integer. We will set

$$W_{k\Delta t}^{(N)} = a_N S_k,$$

where we choose a normalizing constant a_N so that W_1 has variance 1. Since $\text{Var}(S_N) = N$, it is clear that we must choose $a_N = N^{-1/2}$. Hence in this discrete approximation, the size of the jump in time $\Delta t = 1/N$ is $1/\sqrt{N} = (\Delta t)^{1/2}$. We can consider the discrete approximation as a process for all values of t by linear interpolation (see the figure below).



As $N \rightarrow \infty$, this discrete approximation approaches a continuous-time, continuous-space process. By the central limit theorem the distribution of

$$W_1^{(N)} = \frac{S_N}{\sqrt{N}}$$

approaches a normal distribution with mean 0 and variance 1. Similarly, the distribution of $W_t^{(N)}$ approaches a normal distribution with mean 0 and variance t . The limiting process can be shown to be a standard Brownian motion. (It requires some sophisticated mathematics to state explicitly what kind of limit is being taken here. We will not worry about this detail.)

The path of a Brownian motion is very rough. Consider the increment $X_{t+\Delta t} - X_t$ for small Δt . The distribution of this increment has mean 0, variance Δt so

$$\mathbb{E}(|X_{t+\Delta t} - X_t|^2) = \Delta t.$$

In other words the typical size of an increment, $|X_{t+\Delta t} - X_t|$, is about $\sqrt{\Delta t}$. As $\Delta t \rightarrow 0$, $\sqrt{\Delta t} \rightarrow 0$, which is consistent with the continuity of the paths. What about differentiability? Does it make sense to talk about dX_t/dt ? Recall the definition of the derivative from calculus,

$$\frac{dX_t}{dt} = \lim_{\Delta t \rightarrow 0} \frac{X_{t+\Delta t} - X_t}{\Delta t}.$$

When Δt is small, the absolute value of the numerator is on the order of $\sqrt{\Delta t}$ which is much larger than Δt . Hence, this limit does not exist. By a sharpening of this argument one can prove the following.

Fact. *The path of a Brownian motion X_t is nowhere differentiable.*

Care is needed in proving statements such as the one above. The intuitive argument can be used fairly easily to prove the statement “for each t , the

probability that X_t is not differentiable at t is 1.” This is not as strong as the fact above which states “the probability that X_t is not differentiable at all values of t is 1.” This distinction is a little tricky to understand. As a possibly easier example consider the following two statements: “For each t , the probability that $X_t \neq 1$ is 1” and “The probability that $X_t \neq 1$ for all values of t is 1.” These statements are not the same, and, in fact, the first is true and the second is false. For any given t , X_t has a normal distribution; hence the probability of taking on any particular value is 0 (this is true for any continuous distribution). However, the probability that $X_1 > 1$ is certainly greater than 0. If $X_0 = 0$ and $X_1 > 1$, then the continuity of X_t implies that $X_t = 1$ for some $0 < t < 1$. Hence the probability that $X_t = 1$ for some $0 < t < 1$ is greater than 0. The difficulty here comes with the fact that the real numbers are uncountable. We can write

$$\{X_t = 1 \text{ for some } 0 \leq t \leq 1\} = \bigcup_{0 \leq t \leq 1} \{X_t = 1\}.$$

The right-hand side is a union of sets each with probability 0. However, it is an *uncountable* union of such sets. The axioms of probability imply that the countable union of sets of probability 0 has probability 0 but does not say the same for an uncountable union. This phenomenon arises whenever one deals in continuous probability. For example, if Y is any continuous random variable then

$$\{-\infty < Y < \infty\} = \bigcup_{-\infty < y < \infty} \{Y = y\}.$$

The right-hand side is a union of events with probability 0, but the left-hand side has probability 1.

In stochastic processes with continuous time and space, many difficult technical problems can arise in trying to deal with uncountable unions of sets. We will ignore most of these issues here. Most of these problems are relatively easily overcome for Brownian motion.

8.2 Markov Property

Let X_t be a standard Brownian motion. We will let \mathcal{F}_t represent the information contained in $X_s, s \leq t$, in other words all the information that can be obtained from watching the Brownian motion up through time t . Suppose $s < t$ and consider the conditional expectation $E(X_t | \mathcal{F}_s)$. Note that

$$E(X_t | \mathcal{F}_s) = E(X_s | \mathcal{F}_s) + E(X_t - X_s | \mathcal{F}_s).$$

Since X_s is \mathcal{F}_s measurable, the first term on the right-hand side equals X_s . Since $X_t - X_s$ is independent of \mathcal{F}_s , the second term equals $\mathbb{E}(X_t - X_s) = 0$.

Hence

$$E(X_t \mid \mathcal{F}_s) = X_s = E(X_t \mid X_s).$$

The equality of the left-hand and right-hand sides above illustrates the Markov property of Brownian motion, i.e., in order to predict X_t given all the information up through time s , it suffices to consider only the value of the Brownian motion at time s . More generally, the Markov property implies that for functions f ,

$$E[f(X_t) \mid \mathcal{F}_s] = E[f(X_t) \mid X_s].$$

Brownian motion satisfies this property. This follows from an even stronger property of Brownian motion: if $Y_t = X_{s+t} - X_s$, then Y_t is a Brownian motion independent of \mathcal{F}_s . In other words $Z_t = X_{s+t}$ is a Brownian motion starting at the (random) starting point X_s .

Let $p_t(x, y)$ denote the transition densities, i.e., the density of X_t for Brownian motion starting at x . Since $X_t - X_0$ is normal, mean 0, variance t ,

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/2t}, \quad -\infty < y < \infty.$$

The transition densities satisfy the *Chapman–Kolmogorov equation*

$$p_{s+t}(x, y) = \int_{-\infty}^{\infty} p_s(x, z) p_t(z, y) dz.$$

This can be verified directly for this transition function, but one can also see this by appealing to the Markov property. Since $Z_t = X_{s+t}$ is a Brownian motion starting at X_s , the Chapman–Kolmogorov equation averages the density $p_t(z, y)$ over all possible starting points z .

In order to do many useful computations about Brownian motions, a more general Markov property is needed. This is generally referred to as the *strong Markov property*. We first need the notion of a real-valued stopping time. The definition is a generalization of the definition of a stopping time given for discrete-time processes. We say that a random variable T taking values in $[0, \infty]$ is a *stopping time* for Brownian motion if for each t the (indicator function of the) event $\{T \leq t\}$ is measurable with respect to \mathcal{F}_t . In other words, to know whether or not the process has stopped before time t , one only needs to look at the Brownian motion up through time t . The most important examples will be stopping times of the form

$$T_x = \inf\{t : X_t = x\}.$$

If T is a stopping time, we write \mathcal{F}_T for the information contained in the Brownian motion up through the stopping time T (one gets to view the path

up through time T but not beyond). We will let Y_t denote the process beyond time T ,

$$Y_t = X_{t+T} - X_T.$$

Strong Markov Property. Y_t is a Brownian motion independent of \mathcal{F}_T .

It is easier to see what this means by considering an example of how the property is used. Suppose the Brownian motion starts at 0 and we want to calculate the probability that there exists some t with $0 \leq t \leq 1$ and $X_t \geq 1$. Let $T = T_1$ be the first time that the Brownian motion equals 1. Then, by continuity, the event $\{X_t \geq 1 \text{ for some } 0 \leq t \leq 1\}$ is the same as the event $\{T \leq 1\}$. Since

$$\mathbb{P}\{T = 1\} \leq \mathbb{P}\{X_1 = 1\} = 0,$$

we can see that

$$\mathbb{P}\{T \leq 1\} = \mathbb{P}\{T < 1\}.$$

Now consider the event $\{X_1 \geq 1\}$. Since X_1 is normal, mean 0, variance 1,

$$\mathbb{P}\{X_1 \geq 1\} = \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Also,

$$\mathbb{P}\{X_1 \geq 1\} = \mathbb{P}\{T \leq 1\} \mathbb{P}\{X_1 \geq 1 \mid T \leq 1\}.$$

Now we use the strong Markov property. Suppose $T \leq 1$. We may assume in fact that $T < 1$ (since $T = 1$ has probability 0 of occurring). Then, given T , $X_1 - X_T = X_1 - 1$ is a normal random variable, mean 0, variance $1 - T$. Regardless of the variance, we know by the symmetry of the normal distribution that the probability that this normal random variable is greater than or equal to 0 is 1/2. Hence, we conclude

$$\mathbb{P}\{X_1 - 1 \geq 0 \mid T \leq 1\} = 1/2.$$

Therefore

$$\mathbb{P}\{T \leq 1\} = 2 \mathbb{P}\{X_1 \geq 1\} = 2 \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

This result is a particular case of the reflection principle. We now state the general result which is proved in the same way.

Reflection Principle. Suppose X_t is a Brownian motion with variance parameter σ^2 starting at a and $a < b$. Then for any $t > 0$,

$$\begin{aligned} \mathbb{P}\{X_s \geq b \text{ for some } 0 \leq s \leq t\} &= 2 \mathbb{P}\{X_t \geq b \mid X_0 = a\} \\ &= 2 \int_b^\infty \frac{1}{\sqrt{2\pi t \sigma^2}} e^{-(x-a)^2/2\sigma^2 t} dx. \end{aligned}$$

Example 1. Let $t > 1$ and let us compute the probability that a standard Brownian motion crosses the x -axis sometime between times 1 and t , i.e.,

$$\mathbb{P}\{X_s = 0 \text{ for some } 1 \leq s \leq t\}.$$

We first condition on what happens at time $t = 1$. Suppose $X_1 = b > 0$. Then the probability that $X_s = 0$ for some $1 \leq s \leq t$ is the same as the probability that $X_s \leq -b$ for some $0 \leq s \leq t - 1$. This is the same (by symmetry) as the probability that $X_s \geq b$ for some $0 \leq s \leq t - 1$. This probability is given by the reflection principle, so

$$\mathbb{P}\{X_s = 0 \text{ for some } 1 \leq s \leq t \mid X_1 = b\} = 2 \int_b^\infty \frac{1}{\sqrt{2\pi(t-1)}} e^{-x^2/2(t-1)} dx.$$

By symmetry, again, the probability is the same if $X_1 = -b$. Hence, by averaging over all possible values of b we get

$$\begin{aligned} &\mathbb{P}\{X_s = 0 \text{ for some } 1 \leq s \leq t\} \\ &= \int_{-\infty}^\infty p_1(0, b) \mathbb{P}\{X_s = 0 \text{ for some } 1 \leq s \leq t \mid X_1 = b\} db \\ &= 2 \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-b^2/2} \left[2 \int_b^\infty \frac{1}{\sqrt{2\pi(t-1)}} e^{-x^2/2(t-1)} dx \right] db. \end{aligned}$$

The substitution $y = x/\sqrt{t-1}$ in the inside integral reduces this integral to

$$4 \int_0^\infty \int_{b(t-1)^{-1/2}}^\infty \frac{1}{2\pi} e^{-(b^2+y^2)/2} dy db.$$

This integral can be computed using polar coordinates. Note that the region $\{0 < b < \infty, b(t-1)^{-1/2} < y < \infty\}$ corresponds to the polar region $\{0 < r < \infty, \arctan(\sqrt{t-1})^{-1} < \theta < \pi/2\}$. Hence the probability equals

$$\begin{aligned} &4 \int_0^\infty \int_{\arctan((\sqrt{t-1})^{-1})}^{\pi/2} \frac{1}{2\pi} e^{-r^2/2} r d\theta dr \\ &= 4 \left(\frac{\pi}{2} - \arctan \frac{1}{\sqrt{t-1}} \right) \frac{1}{2\pi} \int_0^\infty r e^{-r^2/2} dr \\ &= 1 - \frac{2}{\pi} \arctan \frac{1}{\sqrt{t-1}}. \end{aligned}$$

Example 2. We will show that (with probability one)

$$\lim_{t \rightarrow \infty} \frac{X_t}{t} = 0.$$

First, we consider the limit taken over only integer times. Note that for n an integer,

$$X_n = (X_1 - X_0) + \cdots + (X_n - X_{n-1}),$$

is a sum of independent, identically distributed random variables. It follows from the (strong) law of large numbers that

$$\lim_{n \rightarrow \infty} \frac{X_n}{n} = 0.$$

For each n , let

$$M_n = \sup\{|X_t - X_n| : n \leq t \leq n+1\}.$$

If we can show that

$$\lim_{n \rightarrow \infty} \frac{M_n}{n} = 0,$$

we will be finished since for any t , if $n \leq t < n+1$,

$$\frac{|X_t|}{t} \leq \frac{|X_t|}{n} \leq \frac{|X_n| + |M_n|}{n}.$$

For any $a > 0$, symmetry and the reflection principle state that

$$\begin{aligned} \mathbb{P}\{|M_n| \geq a\} &\leq 2 \mathbb{P}\{M_n \geq a\} = 4 \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &\leq 4 \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-xa/2} dx \\ &= \frac{8}{a\sqrt{2\pi}} e^{-a^2/2}. \end{aligned}$$

If we plug in $a = 2(\ln n)^{1/2}$, we get

$$\mathbb{P}\{|M_n| \geq 2\sqrt{\ln n}\} \leq \frac{8}{2\sqrt{2\pi \ln n} n^2}.$$

In particular, for all n sufficiently large, the probability is less than n^{-2} . If we let I_n denote the indicator function of the event $\{|M_n| \geq 2\sqrt{\ln n}\}$ and

$$I = \sum_{n=0}^{\infty} I_n,$$

we find that $\mathbb{E}(I) < \infty$. This states that the expected number of times that $|M_n| \geq 2\sqrt{\ln n}$ is finite and hence that, with probability one, $|M_n| \geq 2\sqrt{\ln n}$ only finitely often. In particular, this implies that $n^{-1}M_n \rightarrow 0$.

8.3 Zero Set of Brownian Motion

In this section we will investigate the (random) set

$$Z = \{t : X_t = 0\}.$$

It turns out that this set is an interesting “fractal” subset of the real line.

In analyzing this set we will use two important scaling results about Brownian motion which will be proved in the exercises (see Exercises 8.7 and 8.8).

Scaling Properties. *Suppose X_t is a standard Brownian motion. Then,*

(1) *If $a > 0$, and $Y_t = a^{-1/2}X_{at}$, then Y_t is a standard Brownian motion.*

(2) *If X_t is a standard Brownian motion and $Y_t = tX_{1/t}$, then Y_t is a standard Brownian motion.*

In an example in the previous section, we proved that

$$\mathbb{P}\{Z \cap [1, t] \neq \emptyset\} = 1 - \frac{2}{\pi} \arctan \frac{1}{\sqrt{t-1}}.$$

As $t \rightarrow \infty$ the quantity on the right-hand side tends to 1. This tells us that with probability 1 the Brownian motion eventually returns to the origin, and hence (with the help of the strong Markov property) that it returns infinitely often. This means that the Brownian motion for large t has both positive and negative values.

What happens near $t = 0$? Let $Y_t = tX_{1/t}$. Then Y_t is also a standard Brownian motion. As time goes to infinity in the process X , time goes to 0 in Y . Hence, since X_t has both positive and negative values for arbitrarily large values of t , Y_t has positive and negative values for arbitrarily small values of t . This states that in any interval about 0 the Brownian motion takes on both positive and negative values (and hence by continuity also the value 0)!

One topological property that Z satisfies is the fact that Z is a closed set. This means that if a sequence of points $t_i \in Z$ and $t_i \rightarrow t$, then $t \in Z$. This follows from the continuity of the function X_t . For any continuous function, if $t_i \rightarrow t$, then $X_{t_i} \rightarrow X_t$. We have seen that 0 is not an isolated point of Z , i.e., there are positive numbers $t_i \in Z$ such that $t_i \rightarrow 0$. It can be shown that none of the points of Z are isolated points. From a topological perspective Z looks like the Cantor set (see the example below for a definition).

How “big” is the set Z ? To discuss this we need to discuss the notion of a dimension of a set. There are two similar notions of dimension, Hausdorff dimension and box dimension, which can give fractional dimensions to sets. (There is a phrase “fractal dimension” which is used a lot in scientific literature. As a rule, the people who use this phrase are not distinguishing between

Hausdorff and box dimension and could mean either one.) The notion of dimension we will discuss here will be that of box dimension, but all the sets we will discuss have Hausdorff dimension equal to their box dimension. Suppose we have a bounded set A in d -dimensional space \mathbb{R}^d . Suppose we cover A with d -dimensional balls of diameter ϵ . How many such balls are needed? If A is a line segment of length 1 (one-dimensional set), then ϵ^{-1} such balls are needed. If A is a two-dimensional square, however, on the order of ϵ^{-2} such balls are needed. One can see that for a standard k -dimensional set, we need ϵ^{-k} such balls. This leads us to define the (box) dimension of the set A to be the number D such that for small ϵ the number of balls of diameter ϵ needed to cover A is on the order of ϵ^{-D} .

Example. Consider the fractal subset of $[0, 1]$, the Cantor set. The Cantor set A can be defined as a limit of approximate Cantor sets A_n . We start with $A_0 = [0, 1]$. The next set A_1 is obtained by removing the open middle interval $(1/3, 2/3)$, so that

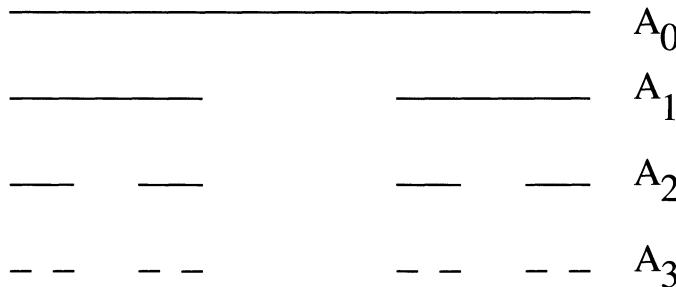
$$A_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

The second set A_2 is obtained by removing the middle thirds of the two intervals in A_1 , hence

$$A_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

In general A_{n+1} is obtained from A_n by removing the “middle third” of each interval. The Cantor set A is then the limit of these sets A_n ,

$$A = \bigcap_{n=1}^{\infty} A_n.$$



Note that A_n consists of 2^n intervals each of length 3^{-n} . Suppose we try

to cover A by intervals of length 3^{-n} ,

$$\left[\frac{k-1}{3^n}, \frac{k}{3^n} \right].$$

We need 2^n such intervals. Hence the dimension D of the Cantor set is the number such that $2^n = (3^{-n})^{-D}$, i.e.,

$$D = \frac{\ln 2}{\ln 3} \approx .631.$$

Now consider the set Z and consider $Z_1 = Z \cap [0, 1]$. We will try to cover Z_1 by one-dimensional balls (i.e., intervals) of diameter (length) $\epsilon = 1/n$. For ease we will consider the n intervals

$$\left[\frac{k-1}{n}, \frac{k}{n} \right], \quad k = 1, 2, \dots, n.$$

How many of these intervals are needed to cover Z_1 ? Such an interval is needed if $Z_1 \cap [(k-1)/n, k/n] \neq \emptyset$. What is

$$P(k, n) = \mathbb{P} \left\{ Z_1 \cap \left[\frac{k-1}{n}, \frac{k}{n} \right] \neq \emptyset \right\}?$$

Assume $k \geq 1$ (if $k = 0$, the probability is 1 since $0 \in Z$). By the scaling property of Brownian motion, $Y_t = ((k-1)/n)^{-1/2} X_{nt/(k-1)}$ is a standard Brownian motion. Hence

$$P(k, n) = \mathbb{P} \left\{ Y_t = 0 \text{ for some } 1 \leq t \leq \frac{k}{k-1} \right\}.$$

This probability was calculated in the previous section,

$$P(k, n) = 1 - \frac{2}{\pi} \arctan \sqrt{k-1}.$$

Therefore, the expected number of the intervals needed to cover Z_1 looks like

$$\sum_{k=1}^n P(k, n) = \sum_{k=1}^n \left[1 - \frac{2}{\pi} \arctan \sqrt{k-1} \right].$$

To estimate the sum, we need to consider the Taylor series for $\arctan(1/t)$ at $t = 0$ (which requires remembering the derivative of \arctan),

$$\arctan \frac{1}{t} = \frac{\pi}{2} - t + O(t^2).$$

In other words, for x large,

$$\arctan x \approx \frac{\pi}{2} - \frac{1}{x}.$$

Hence

$$\sum_{k=1}^n P(k, n) \approx 1 + \sum_{k=2}^n \frac{2}{\pi \sqrt{k-1}} \approx \frac{2}{\pi} \int_1^n (x-1)^{-1/2} dx \approx \frac{4}{\pi} \sqrt{n}.$$

Hence it takes on the order of \sqrt{n} intervals of length $1/n$ to cover Z_1 , or, in other words,

Fact. *The fractal dimension of the zero set Z is $1/2$.*

8.4 Brownian Motion in Several Dimensions

Suppose X_t^1, \dots, X_t^d are independent (one-dimensional) standard Brownian motions. We will call the vector-valued stochastic process

$$X_t = (X_t^1, \dots, X_t^d)$$

a standard d -dimensional Brownian motion. In other words, a d -dimensional Brownian motion is a process in which each component performs a Brownian motion, and the component Brownian motions are independent.

It is not difficult to show that X_t defined as above satisfies the following:

- (i) $X_0 = 0$;
- (ii) for any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, the (vector-valued) random variables $X_{t_1} - X_{s_1}, \dots, X_{t_n} - X_{s_{n-1}}$ are independent;
- (iii) the random variable $X_t - X_s$ has a joint normal distribution with mean 0 and covariance matrix $(t-s)\mathbf{I}$, i.e., has density $f(x_1, \dots, x_d)$ equal to

$$\left(\frac{1}{\sqrt{2\pi r}} e^{-x_1^2/2r} \right) \dots \left(\frac{1}{\sqrt{2\pi r}} e^{-x_d^2/2r} \right) = \frac{1}{(2\pi r)^{d/2}} e^{-|x|^2/2r},$$

where $r = t - s$;

- (iv) X_t is a continuous function of t .

We could use (i) through (iv) as the definition of X_t , but we would quickly discover that we could construct X_t by taking d independent one-dimensional Brownian motions. As in the one-dimensional case we let $p_t(x, y), x, y \in \mathbb{R}^d$ denote the probability density of X_t assuming $X_0 = x$ (it is clear how to define a Brownian motion starting at any point in \mathbb{R}^d),

$$p_t(x, y) = \frac{1}{(2\pi t)^{d/2}} e^{-|y-x|^2/2t}.$$

Again, this satisfies the Chapman–Kolmogorov equation

$$p_{s+t}(x, y) = \int_{\mathbb{R}^d} p_s(x, z) p_t(z, y) dz_1 \dots dz_d.$$

Brownian motion is closely related to the theory of diffusion. Suppose that a large number of particles are distributed in \mathbb{R}^d according to a density $f(y)$. Let $f(t, y)$ denote the density of the particles at time t (so that $f(0, y) = f(y)$). If we assume that the particles perform standard Brownian motions, independently, then we can write the density of particles at time t . If a particle starts at position x , then the probability density for its position at time t is $p_t(x, y)$. By integrating, we get

$$f(t, y) = \int_{\mathbb{R}^d} f(x) p_t(x, y) dx_1 \cdots dx_d.$$

The symmetry of Brownian motion tells us that $p_t(x, y) = p_t(y, x)$. Hence we can write the right-hand side as

$$\int_{\mathbb{R}^d} f(x) p_t(y, x) dx_1 \cdots dx_d.$$

The right-hand side represents the expected value of $f(X_t)$ assuming $X_0 = y$. We can then write this,

$$f(t, y) = \mathbb{E}^y[f(X_t)].$$

The notation \mathbb{E}^y is used to denote expectations of X_t assuming $X_0 = y$.

We will now derive a differential equation that $f(t, x)$ satisfies. Consider $\partial f / \partial t$; for ease we will take $t = 0$, $d = 1$. If f is sufficiently nice, we can write the Taylor series for f about x ,

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2} f''(x)(y - x)^2 + o((y - x)^2),$$

where $o(\cdot)$ denotes an error term such that $o((y - x)^2)/(y - x)^2 \rightarrow 0$ as $y \rightarrow x$. Therefore,

$$\begin{aligned} \left. \frac{\partial f}{\partial t} \right|_{t=0} &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}^x [f(X_t) - f(X_0)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} [f'(x) \mathbb{E}^x [X_t - x] \\ &\quad + \frac{1}{2} f''(x) \mathbb{E}^x [(X_t - x)^2] + o((X_t - x)^2)]. \end{aligned}$$

We know that $\mathbb{E}^x[X_t - x] = 0$ and $\mathbb{E}^x[(X_t - x)^2] = \text{Var}(X_t) = t$. Also since $(X_t - x)^2$ is of order t , the term $t^{-1}o(\cdot)$ tends to 0. Hence we get

$$\left. \frac{\partial f}{\partial t} \right|_{t=0} = \frac{1}{2} f''(x).$$

The same argument holds for all t giving

$$\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2 f}{\partial x^2}.$$

Similarly, we can extend this argument to d dimensions and show that f satisfies the equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \Delta f,$$

where Δ denotes the Laplacian,

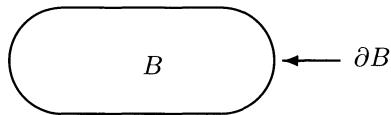
$$\Delta f(t, x_1, \dots, x_d) = \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}.$$

This equation is often called the *heat equation*. One can find a similar solution to the heat equation with diffusion constant D ,

$$\frac{\partial f}{\partial t} = \frac{D}{2} \Delta f,$$

by considering Brownian motions with variance parameter $\sigma^2 = D$.

Sometimes it is useful to consider the heat equation in a bounded domain. Let B be a bounded region of \mathbb{R}^d with boundary ∂B .



Imagine an initial heat distribution on B , $f(x), x \in B$ is given. Suppose also that the temperature is fixed at the boundary, i.e., there is a function $g(y), y \in \partial B$ representing the fixed temperature at point y . If $u(t, x)$ denotes the temperature at x at time t , then $u(t, x)$ satisfies

$$(i) \quad \frac{\partial u}{\partial t} = \frac{D}{2} \Delta u, \quad x \in B,$$

$$(ii) \quad u(t, x) = g(x), \quad x \in \partial B,$$

$$(iii) \quad u(0, x) = f(x), \quad x \in B.$$

The solution of (i) through (iii) can be written in terms of Brownian motion. Let X_t be a d -dimensional Brownian motion with variance parameter $\sigma^2 = D$. Let $\tau = \tau_{\partial B}$ be the first time that the Brownian motion hits the boundary ∂B ,

$$\tau = \inf\{t : X_t \in \partial B\}.$$

Then the solution can be written as

$$u(t, x) = \mathbb{E}^x [f(X_t)I\{\tau > t\} + g(X_\tau)I\{\tau \leq t\}].$$

In other words, at time t , take the average value of the following: $f(X_t)$ for the paths that have not hit ∂B and $g(X_\tau)$ for those paths that have hit ∂B . As $t \rightarrow \infty$, the temperature approaches a steady-state distribution $v(x)$ with boundary value $g(x)$. The steady-state solution satisfies

$$(i) \quad \Delta v(x) = 0, \quad x \in B,$$

$$(ii) \quad v(x) = g(x), \quad x \in \partial B.$$

The solution is given by

$$v(x) = \lim_{t \rightarrow \infty} u(t, x) = \mathbb{E}^x [g(X_\tau)].$$

Example 1. Let $d = 1$ and suppose that $B = (a, b)$ with $0 \leq a < b < \infty$. Then $\partial B = \{a, b\}$. Take $a < x < b$ and consider

$$\tau = \inf\{t : X_t = a \text{ or } b\},$$

where X_t is a standard Brownian motion. Let g be the function on ∂B , $g(a) = 0, g(b) = 1$. Then

$$v(x) = \mathbb{E}^x [g(X_\tau)] = \mathbb{P}^x \{X_\tau = b\}$$

(here we have used \mathbb{P}^x to denote a probability assuming $X_0 = x$). We know by above that $v(x)$ satisfies

$$\frac{d^2v}{dx^2} = 0, \quad a < x < b,$$

$$v(a) = 0, \quad v(b) = 1.$$

We can solve this differential equation easily and we get

$$v(x) = \frac{x - a}{b - a}.$$

This is the Brownian motion analogue of the gambler's ruin estimate.

Example 2. Let $d = 1$ and suppose that $B = (0, \pi)$ and that $X_0 = y \in (0, \pi)$. Let $u(t, x)$ be the solution of the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial u}{\partial x}$$

with boundary conditions $u(t, 0) = u(t, \pi) = 0$ and such that as t goes to 0, $u(t, x)$ approaches the “delta function” at y . Then $u(t, x), 0 < x < \pi$ also denotes the density of the Brownian motion restricted to those paths that have not left $(0, \pi)$. The function u can be found explicitly using the technique of separation of variables. First, it is easy to check that for all integers n , the function $e^{-tn^2/2} \sin(nx)$ satisfies the heat equation and equals zero on the boundary. Therefore, for any choice of constants C_n , the function

$$u(t, x) = \sum_{n=1}^{\infty} C_n e^{-tn^2/2} \sin(nx),$$

satisfies the heat equation and the boundary condition. If we want $u(0, x) = f(x)$, then we need to choose the constants so that

$$f(x) = \sum_{n=1}^{\infty} C_n \sin(nx).$$

Since

$$\int_0^\pi \sin(nx) \sin(mx) dx = 0 \quad \text{if } n \neq m,$$

we can see that C_n must satisfy

$$\int_0^\pi f(x) \sin(nx) dx = C_n \int_0^\pi \sin^2(nx) dx = \frac{\pi}{2} C_n.$$

In the case where f is the delta function at y , we choose

$$C_n = \frac{2}{\pi} \int_0^\pi f(x) \sin(nx) dx = \frac{2}{\pi} \sin(ny).$$

Hence,

$$u(t, x) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-tn^2/2} \sin(ny) \sin(nx).$$

As $t \rightarrow \infty$,

$$u(t, x) \sim \frac{2}{\pi} e^{-t/2} \sin y \sin x.$$

Example 3. If $d \geq 1$, $D = 1$, $g \equiv 0$, then one can try to write the solution of the heat equation in the form

$$u(t, x) = \sum_{n=1}^{\infty} C_n e^{-\lambda_n t/2} \phi_n(x),$$

where the functions ϕ_n are eigenfunctions of Δ with eigenvalue $-\lambda_n$ and Dirichlet boundary conditions, i.e.,

$$\Delta \phi_n(x) = -\lambda_n \phi_n(x), \quad x \in B \quad \phi(x) = 0, \quad x \in \partial B.$$

In order to do this, we need to find a collection of such eigenfunctions that are orthogonal,

$$\int_B \phi_n(x) \phi_m(x) dx_1 \cdots dx_d = 0, \quad n \neq m,$$

and are complete, i.e., each f can be written as

$$f(x) = \sum_{n=1}^{\infty} C_n \phi_n(x).$$

For a number of regions, such as balls in \mathbb{R}^d , the eigenfunctions and eigenvalues are known. For a much wider class of regions, one can prove the existence of such a collection of functions. See a book on partial differential equations for more information. If B is a bounded, connected region, the eigenfunction ϕ_1 associated to the largest eigenvalue $-\lambda_1$ (the eigenvalue of smallest absolute value) can be chosen so that if $X_0 = y$, the density $u(t, x)$ satisfies

$$u(t, x) \sim e^{-\lambda_1 t/2} \phi_1(y) \phi_1(x), \quad t \rightarrow \infty.$$

In the previous example, $\lambda_1 = 1$ and $\phi_1(x) = \sqrt{2/\pi} \sin x$.

8.5 Recurrence and Transience

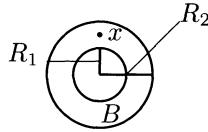
In this section we ask whether the Brownian motion keeps returning to the origin. We have already answered this question for one-dimensional Brownian motion; if X_t is a standard (one-dimensional) Brownian motion, then X_t is recurrent, i.e., there are arbitrarily large times t with $X_t = 0$.

Now suppose X_t is a standard d -dimensional Brownian motion. Let $0 < R_1 < R_2 < \infty$ and let $B = B(R_1, R_2)$ be the annulus

$$B = \{x \in \mathbb{R}^d : R_1 < |x| < R_2\},$$

with boundary

$$\partial B = \{x \in \mathbb{R}^d : |x| = R_1 \text{ or } |x| = R_2\}.$$



Suppose $x \in B$. Let $f(x) = f(x, R_1, R_2)$ be the probability that a standard Brownian motion starting at x hits the sphere $\{y : |y| = R_2\}$ before it hits the sphere $\{y : |y| = R_1\}$. If we let

$$\tau = \tau_{\partial B} = \inf\{t : X_t \in \partial B\},$$

then we can write

$$f(x) = \mathbb{E}^x[g(X_\tau)],$$

where $g(y) = 1$ for $|y| = R_2$ and $g(y) = 0$ for $y = R_1$. We saw in the last section that f is the function satisfying

$$(i) \quad \Delta f(x) = 0, \quad x \in B,$$

$$(ii) \quad f(y) = 0, \quad |y| = R_1; \quad f(y) = 1, \quad |y| = R_2.$$

To find f , we first note that the symmetry of Brownian motion implies $f(x) = \phi(|x|)$ for some ϕ , i.e., the value of f depends only on the absolute value of x . We can write the equation (i) in spherical coordinates. The form of the Laplacian Δ in spherical coordinates is somewhat messy; however, it is not so bad for functions $\phi(r)$ that depend only on the radius. One can check that

$$\Delta \phi(r) = \frac{d^2 \phi}{dr^2} + \frac{d-1}{r} \frac{d\phi}{dr}.$$

The general solution to the equation

$$\phi''(r) + \frac{d-1}{r} \phi'(r) = 0$$

is given by

$$\phi(r) = \begin{cases} c_1 \ln r + c_2, & d = 2, \\ c_1 r^{2-d} + c_2, & d \geq 3. \end{cases}$$

[The second-order equation for $\phi(r)$ is a first-order equation for $\psi(r) = \phi'(r)$ which can be solved by separation of variables.] Putting in the boundary conditions $\phi(R_1) = 0$ and $\phi(R_2) = 1$, we see that

$$f(x) = \phi(|x|) = \frac{\ln |x| - \ln R_1}{\ln R_2 - \ln R_1}, \quad d = 2,$$

$$f(x) = \phi(|x|) = \frac{R_1^{2-d} - |x|^{2-d}}{R_1^{2-d} - R_2^{2-d}}, \quad d \geq 3.$$

Consider now the two-dimensional case. Let $x \in \mathbb{R}^2$ and suppose that a Brownian motion starts at x (or that the Brownian motion is at x at some time t). Take any $\epsilon > 0$, and ask the question: What is the probability that the Brownian motion never returns to the disc of radius ϵ about 0? The argument above gives us the probability of reaching the circle of radius R_2 before reaching the disc. The probability we are interested in is therefore

$$\lim_{R_2 \rightarrow \infty} \mathbb{P}^x\{|X_t| = R_2 \text{ before } |X_t| = \epsilon\} = \lim_{R_2 \rightarrow \infty} \frac{\ln|x| - \ln \epsilon}{\ln R_2 - \ln \epsilon} = 0.$$

Hence, with probability one the Brownian motion always returns to the disc of radius ϵ and hence it returns infinitely often and at arbitrarily large times. Does it ever return to the point 0, i.e., are there times t with $X_t = 0$? Again, start the walk at $x \neq 0$. If there is a positive probability of reaching 0, then there must be an R_2 such that the probability of reaching 0 before reaching the circle of radius R_2 is positive. But this latter probability can be written as

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}^x\{|X_t| = \epsilon \text{ before } |X_t| = R_2\} = \lim_{\epsilon \rightarrow 0} \left[1 - \frac{\ln|x| - \ln \epsilon}{\ln R_2 - \ln \epsilon} \right] = 0.$$

Hence the Brownian motion never actually returns to 0. To summarize, the Brownian motion in two dimensions returns arbitrarily close to 0 infinitely often, but never actually returns to 0. We say that the Brownian motion in two dimensions is *neighborhood recurrent* but not *point recurrent*.

Now consider $d \geq 3$. Again we take $\epsilon > 0$ and ask what is the probability that the Brownian motion starting at x never returns to the ball of radius ϵ . If $|x| > \epsilon$, this is given by

$$\lim_{R_2 \rightarrow \infty} \frac{\epsilon^{2-d} - |x|^{2-d}}{\epsilon^{2-d} - R_2^{2-d}} = 1 - \left(\frac{\epsilon}{|x|} \right)^{d-2} < 1.$$

Since the probability is less than 1, we can see that eventually the Brownian motion escapes from any ball around the origin and hence goes off to infinity. We say that in this case the Brownian motion is *transient*.

8.6 Fractal Nature of Brownian Motion

Let X_t be a standard d -dimensional Brownian motion and let A represent the (random) set of points visited by the path,

$$A = \{x \in \mathbb{R}^d : X_t = x \text{ for some } t\}.$$

In this section we will consider the dimension of the set A for $d \geq 2$.

In order to consider a bounded set, let $A_1 = A \cap \{x : |x| \leq 1\}$. Fix an ϵ and let us try to cover A_1 with balls of diameter ϵ . First consider the whole ball of radius 1, $\{x : |x| \leq 1\}$ and cover it by balls of diameter ϵ . The number of such balls needed is of the order of ϵ^{-d} (which is consistent with the fact that the ball is a dimension d set). How many of these balls are needed to cover A_1 ?

First, consider $d = 2$. By the argument given in the previous section, every open ball is visited by the Brownian motion. Hence A intersects every ball and all the balls are needed. Hence the dimension of A is two.

Now consider $d > 2$. Take a typical ball of diameter ϵ . What is the probability that it is needed in the covering, i.e., what is the probability that Brownian motion visits the ball? By the calculations done in the previous section, a ball of radius $\epsilon/2$ around a point x (with $|x| > \epsilon/2$) is visited with probability $(\epsilon/2|x|)^{d-2}$. Hence, if ϵ is small and $|x|$ is of order 1, the probability is about a constant times ϵ^{d-2} . Since each of the about ϵ^{-d} balls is chosen for the covering with probability about ϵ^{d-2} , the total number of balls needed is about $\epsilon^{d-2}\epsilon^{-d} = \epsilon^{-2}$. Hence the dimension of the set A is two. We have just sketched the idea behind this following fact:

Fact. *The path of a d -dimensional Brownian motion ($d \geq 2$) has fractal dimension two.*

8.7 Scaling Rules

The fractal nature of Brownian motion is closely related to the scaling rule: if X_t is a standard one-dimensional Brownian motion and $b > 0$, then $Y_t = b^{-1/2}X_{bt}$ is also a standard Brownian motion. A process satisfying the properties discussed on page 173 must satisfy this scaling rule. Suppose that we were willing to give up the condition that X_t is a continuous function of t . Could we get different scaling laws? Is there a process that is symmetric about zero satisfying the other conditions that has a different scaling exponent λ by which we mean that $Y_t = b^{-\lambda}X_{bt}$ has the same distribution as X_t ?

Let us suppose that such a process exist with scaling exponent λ . If we assume that X_t has a finite variance then λ must equal $1/2$. This follows from the simple calculation

$$\begin{aligned}\text{Var}(X_1) &= \text{Var}[X_{1/n} + (X_{2/n} - X_{1/n}) + \cdots + (X_{n/n} - X_{(n-1)/n})] \\ &= n \text{Var}(X_{1/n}) = n \text{Var}(b^{-\lambda} X_1) = n^{1-2\lambda} \text{Var}(X_1),\end{aligned}$$

which implies that $\lambda = 1/2$.

Let

$$M_n = \max \{ |X_{1/n}|, |X_{2/n} - X_{1/n}|, \dots, |X_{n/n} - X_{(n-1)/n}| \}.$$

If the paths have jumps, then we expect $\mathbb{P}\{M_n \geq \epsilon\}$ not to go to zero as $n \rightarrow \infty$ for some value of ϵ . However, assuming the paths are not too wild, we would expect that $\mathbb{P}\{M_n \geq K\}$ would be less than, say $1/2$, for some value of K . Note that

$$\begin{aligned} \mathbb{P}\{M_n \leq r\} &= \mathbb{P}\{|X_{j/n} - X_{(j-1)/n}| \leq r \text{ for } j = 1, \dots, n\} \\ &= \mathbb{P}\{|X_{1/n}| \leq r\}^n \\ &= \mathbb{P}\{n^{-\lambda} |X_1| \leq r\}^n = \mathbb{P}\{|X_1| \leq r n^\lambda\}^n. \end{aligned}$$

If we recall that $(1 - \frac{a}{n})^n \rightarrow e^{-a}$, we can see that a good candidate for the distribution of X_1 would be one satisfying $\mathbb{P}\{|X_1| > n^\lambda\} \sim cn^{-1}$, or

$$\mathbb{P}\{|X_1| \geq y\} \sim c y^{-1/\lambda}.$$

If $\lambda < 1/2$, then it is not difficult to check that such an X_1 would have a finite variance. But this implies that $\lambda = 1/2$. Hence there are no examples with $\lambda < 1/2$. For $\lambda > 1/2$, there are examples and these are called the *symmetric stable distribution* and the corresponding processes are called *symmetric stable processes*. The density of these processes cannot be given explicitly except in the case $\lambda = 1$ which is the *Cauchy distribution* with density

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

8.8 Brownian Motion with Drift

Consider a d -dimensional Brownian motion X_t with variance parameter σ^2 starting at $x \in \mathbb{R}^d$. Let $\mu \in \mathbb{R}^d$ and

$$Y_t = X_t + t\mu.$$

Then Y_t is called d -dimensional Brownian motion with drift μ and variance parameter σ^2 starting at x . One can check easily that Y_t satisfies

- (i) $Y_0 = x$;
- (ii) if $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, then $Y_{t_1} - Y_{s_1}, \dots, Y_{t_n} - Y_{s_n}$ are independent;
- (iii) $Y_t - Y_s$ has a normal distribution with mean $\mu(t-s)$ and covariance matrix $\sigma^2(t-s)\mathbf{I}$;
- (iv) Y_t is a continuous function of t .

The motion Y_t consists of a “straight line” motion in the direction μ with random fluctuations. Note that $\mathbb{E}(Y_t) = t\mu$.

The density of Y_t given $Y_0 = x$, $p_t(x, y)$ is easily seen to be

$$p_t(x, y) = \frac{1}{(2\pi\sigma^2 t)^{d/2}} e^{-|y-x-t\mu|^2/2t\sigma^2}.$$

This satisfies the Chapman–Kolmogorov equation,

$$p_{s+t}(x, y) = \int_{\mathbb{R}^d} p_s(x, z) p_t(z, y) dz_1 \cdots dz_d.$$

Suppose we start with a density on \mathbb{R}^d , $f(x)$. Consider the function

$$f(t, x) = \mathbb{E}^x[f(Y_t)].$$

For ease we will consider the case $d = 1, t = 0$. We again write f in a Taylor series about x ,

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2}f''(x)(y - x)^2 + o((y - x)^2).$$

Hence,

$$\begin{aligned} \mathbb{E}^x[f(Y_t)] &= f(x) + f'(x)\mathbb{E}^x[Y_t - x] \\ &\quad + \frac{1}{2}f''(x)\mathbb{E}^x[(Y_t - x)^2] + o(\mathbb{E}^x(Y_t - x)^2). \end{aligned}$$

A Brownian motion with drift μ and variance parameter σ^2 starting at x can be obtained by letting $Y_t = X_t + t\mu + x$, where X_t is a (zero drift) Brownian motion with variance parameter σ^2 starting at 0. Hence,

$$\mathbb{E}^x[Y_t - x] = \mathbb{E}[X_t + t\mu] = t\mu,$$

$$\begin{aligned} \mathbb{E}^x[(Y_t - x)^2] &= \mathbb{E}[(X_t + t\mu)^2] = [\mathbb{E}(X_t + t\mu)]^2 + \text{Var}(X_t + t\mu) \\ &= (t\mu)^2 + \sigma^2 t. \end{aligned}$$

Also, since $(Y_t - x)^2$ is order t , $o((Y_t - x)^2)$ is $o(t)$. Therefore,

$$\begin{aligned} \frac{\partial f}{\partial t} \Big|_{t=0} &= \lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(Y_t)] - \mathbb{E}^x[f(Y_0)]}{t} \\ &= \mu f'(x) + \frac{\sigma^2}{2} f''(x). \end{aligned}$$

We see that the inclusion of a drift has added a first derivative with respect to x .

In d dimensions, if the drift $\mu = (\mu_1, \dots, \mu_d)$, we would get

$$\frac{\partial f}{\partial t} = \sum_{i=1}^d \mu_i \frac{\partial f}{\partial x_i} + \frac{\sigma^2}{2} \Delta f.$$

8.9 Exercises

8.1 Let X be a normal random variable, mean 0 variance 1. Show that if $a > 0$

$$\mathbb{P}\{X \geq a\} \leq \frac{2}{a\sqrt{2\pi}} e^{-a^2/2}.$$

(Hint:

$$\int_a^\infty e^{-x^2/2} dx \leq \int_a^\infty e^{-ax/2} dx.)$$

8.2 Let X_{n1}, \dots, X_{nn} be independent normal random variables with mean 0 and variance $1/n$. Then

$$X = X_{n1} + \dots + X_{nn},$$

is a normal random variable with mean 0, variance 1. Let

$$M_n = \max \{|X_{n1}|, \dots, |X_{nn}|\}.$$

Show that for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{M_n > \epsilon\} = 0.$$

(Hint: it will be useful to use the estimate from Problem 8.1. It may also be useful to remember that if Y is normal mean 0, variance σ^2 , then $\sigma^{-1}Y$ is normal mean 0, variance 1.)

8.3 Let X_{n1}, \dots, X_{nn} be independent Poisson random variables with mean $1/n$. Then

$$X = X_{n1} + \dots + X_{nn},$$

is a Poisson random variable with mean 1. Let

$$M_n = \max\{X_{n1}, \dots, X_{nn}\}.$$

Find

$$\lim_{n \rightarrow \infty} \mathbb{P}\{M_n > 1/2\}.$$

8.4 Let X_t denote a standard (one-dimensional) Brownian motion. Find the following probabilities. Give your answers as rational numbers or decimals to at least three places.

- (a) $X_2 > 2$
- (b) $X_2 > X_1$
- (c) $X_2 > X_1 > X_3$
- (d) $X_t = 0$ for some t with $2 \leq t \leq 3$
- (e) $X_t < 4$ for all t with $0 < t < 3$
- (f) $X_t > 0$ for all $t > 10$.

8.5 Random variables Y_1, \dots, Y_n have a joint normal distribution with mean 0 if there exist independent random variables X_1, \dots, X_n , each normal mean 0, variance 1, and constants a_{ij} such that

$$Y_i = a_{i1}X_1 + \dots + a_{in}X_n.$$

Let X_t be a standard Brownian motion. Let $s_1 \leq s_2 \leq \dots \leq s_n$. Explain why it follows from the definition of a Brownian motion that X_{s_1}, \dots, X_{s_n} have a joint normal distribution.

8.6 If Y_1, \dots, Y_n have a joint normal distribution with mean 0, then the covariance matrix is the matrix Γ whose (i, j) entry is $\mathbb{E}(Y_i Y_j)$. Let X_t and s_1, \dots, s_n be as in Exercise 8.5.

- (a) Find the covariance matrix Γ for X_{s_1}, \dots, X_{s_n} .
- (b) The moment generating function (mgf) for Y_1, \dots, Y_n is the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$f(t_1, \dots, t_n) = \mathbb{E}[e^{t_1 Y_1 + \dots + t_n Y_n}].$$

Find the mgf for Y_1, \dots, Y_n in terms of its covariance matrix Γ .

(c) If two distributions have the same mgf, then the two distributions are the same. Use this fact to prove the following: if Y_1, \dots, Y_n have a mean 0 joint normal distribution, and $\mathbb{E}[Y_i Y_j] = 0$ for all $i \neq j$, then Y_1, \dots, Y_n are independent.

8.7 Suppose X_t is a standard Brownian motion and $Y_t = a^{-1/2}X_{at}$ with $a > 0$. Show that Y_t is a standard Brownian motion.

8.8 Suppose X_t is a standard Brownian motion and $Y_t = tX_{1/t}$. Show that Y_t is a standard Brownian motion. (Hint: it may be useful to use Exercise 8.6 (c).)

8.9 Let X_t be a standard Brownian motion. Compute the following conditional probability:

$$\mathbb{P}\{X_2 > 0 \mid X_1 > 0\}.$$

Are the events $\{X_1 > 0\}$ and $\{X_2 > 0\}$ independent?

8.10 Let X_t and Y_t be independent standard (one-dimensional) Brownian motions.

(a) Show that $Z_t = X_t - Y_t$ is a Brownian motion. What is the variance parameter for Z_t ?

(b) True or False: With probability 1, $X_t = Y_t$ for infinitely many values of t .

8.11 Let X_t be a standard (one-dimensional) Brownian motion starting at 0 and let

$$M = \max\{X_t : 0 \leq t \leq 1\}.$$

Find the density for M and compute its expectation and variance.

8.12 Let X_t be a standard (one-dimensional) Brownian motion starting at 0 and let

$$T = \min\{t : |X_t| = 1\}, \quad \tilde{T} = \min\{t : X_t = 1\}.$$

(a) Show that there exists positive constants c, β such that for all $t > 0$,

$$\mathbb{P}\{T > t\} \leq c e^{-\beta t}.$$

Conclude that $\mathbb{E}[T] < \infty$.

(b) Use the reflection principle to find the density of \tilde{T} , and show that $\mathbb{E}[\tilde{T}] = \infty$.

8.13 Let X_t, T be as in Exercise 8.12 and let

$$T^* = \min\{t : X_t = 1 \text{ or } X_t = -3\}.$$

(a) Explain why X_{T^*} and T^* are independent random variables.

(b) Show that T^* and X_{T^*} are *not* independent.

8.14 Let X_t be a standard (one-dimensional) Brownian motion started at a point y chosen uniformly on the interval $(0, 1)$. Suppose the motion is stopped whenever it reaches 0 or 1, and let $u(t, x), 0 < x < 1$ denote the density of the position X_t restricted to those paths that have not left $(0, 1)$. Find $u(t, x)$ explicitly in terms of an infinite series and use the series to find the function h and the constant β such that as $t \rightarrow \infty$,

$$u(t, x) \sim e^{-\beta t} h(x).$$

8.15 Let the Cantor-like set A be defined as follows. Let $A_0 = [0, 1]$,

$$A_1 = \left[0, \frac{2}{5}\right] \cup \left[\frac{3}{5}, 1\right],$$

and A_n is obtained from A_{n-1} by removing the “middle fifth” from each interval in A_{n-1} . Let

$$A = \bigcap_{n=0}^{\infty} A_n.$$

What is the fractal dimension of A ?

8.16 Suppose that X has a Cauchy distribution, i.e., has density

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

- (a) If $a > 0$, let $Y = a^{-1}X$. What is the density of Y ?
- (b) Suppose that Y, Z are independent random variables each with a Cauchy distribution. Show that the average $(Y+Z)/2$ also has a Cauchy distribution.
- (c) For which $r > 0$ is $\mathbb{E}[|X|^r] < \infty$?

8.17 Let $X_t = (X_t^1, X_t^2)$ denote a standard two-dimensional Brownian motion. Let

$$\sigma_t = \min\{s : X_s^2 = t\}, \quad Y_t = X_{\sigma_t}^1.$$

- (a) Which of the following properties does the process Y_t satisfy?
 - (i) $Y_0 = 0$;
 - (ii) For $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n$, the random variables $Y_{t_1} - Y_{s_1}, \dots, Y_{t_n} - Y_{s_n}$ are independent;
 - (iii) If $0 \leq s < t$, then the distribution of $Y_t - Y_s$ is the same as that of Y_{t-s} ;
 - (iv) Y_t is a continuous function of t .
- (b) For which $\lambda > 0$ does the process $Z_t = a^{-\lambda} Y_{at}$ have the same distribution as Y_{at} ?

Chapter 9

Stochastic Integration

9.1 Integration with Respect to Random Walk

The goal of this chapter is to introduce the idea of integration with respect to Brownian motion. To give the reader a sense for the integral, we will start by discussing integration with respect to simple random walk. Let X_1, X_2, \dots be independent random variables, $\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = 1/2$ and let S_n denote the corresponding simple random walk

$$S_n = X_1 + \cdots + X_n.$$

As in Section 5.2, Example 3, we think of X_n as being the result of a game at time n and we can consider possible betting strategies on the games.

Let \mathcal{F}_n denote the information contained in X_1, \dots, X_n . Let B_n be the “bet” on the n th game. B_n can be either positive or negative, a negative value being the same as betting that X_n will turn up -1 . The important assumption that we make is that the bettor must make the bet using only the information available up to, but not including, the n th game, i.e., we assume that B_n is measurable with respect to \mathcal{F}_{n-1} . The winnings up to time n , Z_n , can be written as

$$Z_n = \sum_{i=1}^n B_i X_i = \sum_{i=1}^n B_i [S_i - S_{i-1}] = \sum_{i=1}^n B_i \Delta S_i,$$

where we write $\Delta S_i = S_i - S_{i-1}$. We call Z_n the integral of B_n with respect to S_n .

There are two important properties that this integral satisfies. The first was shown in Section 5.2, Example 3: the process Z_n is a martingale with respect to \mathcal{F}_n , i.e., if $m < n$,

$$E(Z_n | \mathcal{F}_m) = Z_m.$$

In particular, $\mathbb{E}(Z_n) = 0$. The second property deals with the second moment of Z_n . Assume that the bets B_n have finite second moments, $\mathbb{E}(B_n^2) < \infty$. Then

$$\text{Var}(Z_n) = \mathbb{E}(Z_n^2) = \sum_{i=1}^n \mathbb{E}(B_i^2).$$

To see this, we expand the square to write

$$Z_n^2 = \sum_{i=1}^n B_i^2 X_i^2 + 2 \sum_{1 \leq i < j \leq n} B_i B_j X_i X_j.$$

Note that $X_i^2 = 1$ and hence

$$\mathbb{E} \left(\sum_{i=1}^n B_i^2 X_i^2 \right) = \sum_{i=1}^n \mathbb{E}(B_i^2).$$

Suppose $i < j$. Then B_i, X_i, B_j are all measurable with respect to \mathcal{F}_{j-1} while X_j is independent of \mathcal{F}_{j-1} . Using (5.3), we see that

$$\mathbb{E}(B_i B_j X_i X_j | \mathcal{F}_{j-1}) = B_i B_j X_i \mathbb{E}(X_j | \mathcal{F}_{j-1}) = B_i B_j X_i \mathbb{E}(X_j) = 0,$$

and hence

$$\mathbb{E}(B_i B_j X_i X_j) = \mathbb{E} [\mathbb{E}(B_i B_j X_i X_j | \mathcal{F}_{j-1})] = 0.$$

9.2 Integration with Respect to Brownian Motion

Here we describe a continuous analogue of the discrete integral given in the last section. Instead of a simple random walk, we will take a standard (one-dimensional) Brownian motion, which we will write W_t . We can think of this as a continuous fair game such that if one bets one unit for the entire period $[s, t]$ then one's winnings in this time period would be $W_t - W_s$.

Let Y_t denote the amount that is bet at time t . What we would like to do is define

$$Z_t = \int_0^t Y_s dW_s.$$

The process Z_t should denote the amount won in this game up to time t if the amount bet at time s is Y_s . It is a nontrivial mathematical problem to define this integral. The roughness of the paths of the Brownian motion prevent one from defining the integral as a “Riemann–Stieljes” integral.

We will make two assumptions about our betting strategy Y_s . The first assumption is that $\mathbb{E}(Y_t^2) < \infty$ for all t and for each t ,

$$\int_0^t \mathbb{E}(Y_s^2) ds < \infty.$$

This condition will certainly be satisfied if we restrict ourselves to bounded betting strategies. The second assumption is critical and corresponds to our

assumption in the discrete case that the bettor cannot look into the future to determine the bet. Let \mathcal{F}_t denote the information contained in the Brownian motion up through time t . We assume that Y_t is \mathcal{F}_t -measurable. In other words, the bettor can see the entire Brownian motion up through time t before choosing the bet, but cannot see anything after time t .

It is not too difficult to define the integral if we make the restrictive assumption that the bettor can change the bet only at a certain finite set of times, say $t_1 < t_2 < \dots < t_n$. The bets then take the form

$$Y_t = \begin{cases} Y_0, & 0 \leq t < t_1, \\ Y_1, & t_1 \leq t < t_2, \\ \vdots \\ Y_n, & t_n \leq t < \infty. \end{cases}$$

Here Y_0, \dots, Y_n are random variables with $\mathbb{E}(Y_i^2) < \infty$, and Y_i must be measurable with respect to \mathcal{F}_{t_i} (where $t_0 = 0$). We will call a betting strategy that can change at only a finite number of times a *simple* strategy. For a simple strategy, we define the stochastic integral for $t_j \leq t < t_{j+1}$ by

$$Z_t = \int_0^t Y_s dW_s = \sum_{i=1}^j Y_{i-1}[W_{t_i} - W_{t_{i-1}}] + Y_j[W_t - W_{t_j}].$$

There are three important properties that the stochastic integral of a simple strategy satisfies. The first is linearity: if X_s and Y_s are two simple strategies and a, b are real numbers, then $aX_s + bY_s$ is a simple strategy and

$$\int_0^t (aX_s + bY_s) dW_s = a \int_0^t X_s dW_s + b \int_0^t Y_s dW_s.$$

This can be easily checked.

The other two properties are direct analogues of the properties of the discrete stochastic integral of the previous section. We say a continuous-time process Z_t is a martingale with respect to \mathcal{F}_t if each Z_t is \mathcal{F}_t -measurable; $\mathbb{E}(|Z_t|) < \infty$ for each t ; and if $s < t$,

$$E(Z_t | \mathcal{F}_s) = Z_s. \quad (9.1)$$

The second property is that the stochastic integral Z_t as defined above is a martingale with respect to the information \mathcal{F}_t derived from the Brownian motion. It is easy to see that Z_t is \mathcal{F}_t -measurable and the condition $\mathbb{E}(|Z_t|) < \infty$ follows from the fact that the second moments of the Y_i exist. We will now verify (9.1). First assume $t_j \leq s < t \leq t_{j+1}$ for some j . Then we can write

$$Z_t = Z_s + Y_j[W_t - W_s].$$

Since Y_j and Z_s are \mathcal{F}_s -measurable and $W_t - W_s$ is independent of \mathcal{F}_s ,

$$E(Z_t | \mathcal{F}_s) = Z_s + Y_j E(W_t - W_s | \mathcal{F}_s) = Z_s + Y_j \mathbb{E}(W_t - W_s) = Z_s.$$

In particular, if $t_j \leq t \leq t_{j+1}$,

$$E(Z_{t_{j+1}} | \mathcal{F}_t) = Z_t, \quad E(Z_t | \mathcal{F}_{t_j}) = Z_{t_j}.$$

Note that $E(Z_t | \mathcal{F}_{t_{j-1}}) = E(E(Z_t | \mathcal{F}_{t_j}) | \mathcal{F}_{t_{j-1}}) = E(Z_{t_j} | \mathcal{F}_{t_{j-1}}) = Z_{t_{j-1}}$, and by iteration we can see that for all $i \leq j$, $E(Z_t | \mathcal{F}_{t_i}) = Z_{t_i}$. Finally, if $t_i \leq s < t_{i+1}, t_j \leq t < t_{j+1}$ for some $i < j$, then

$$E(Z_t | \mathcal{F}_s) = E(E(Z_t | \mathcal{F}_{t_{i+1}}) | \mathcal{F}_s) = E(Z_{t_{i+1}} | \mathcal{F}_s) = Z_s.$$

This gives (9.1).

The third property gives a way to calculate the second moment,

$$\mathbb{E}(Z_t^2) = \int_0^t \mathbb{E}(Y_s^2) ds. \quad (9.2)$$

The right-hand side is a standard calculus “ ds ” integral. To prove this, assume that $t_j \leq t < t_{j+1}$. Note that $\mathbb{E}(Y_s^2)$ is a step function in s so

$$\int_0^t \mathbb{E}(Y_s^2) ds = \sum_{i=0}^{j-1} \mathbb{E}(Y_i^2)(t_{i+1} - t_i) + \mathbb{E}(Y_j^2)(t - t_j).$$

If we expand the square, we see that

$$Z_t^2 = \sum_{i=1}^j Y_{i-1}^2 [W_{t_i} - W_{t_{i-1}}]^2 + Y_j^2 [W_t - W_{t_j}]^2 + (\text{cross terms}),$$

where “cross terms” represents a sum of terms of the form

$$Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] [W_{t_k} - W_{t_{k-1}}], \quad i < k,$$

or

$$Y_{i-1} Y_j [W_{t_i} - W_{t_{i-1}}] [W_t - W_j].$$

If $i < k$,

$$\begin{aligned} & E(Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] [W_{t_k} - W_{t_{k-1}}] | \mathcal{F}_{t_{k-1}}) \\ &= Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] E(W_{t_k} - W_{t_{k-1}} | \mathcal{F}_{t_{k-1}}) \\ &= Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] \mathbb{E}(W_{t_k} - W_{t_{k-1}}) = 0, \end{aligned}$$

and hence

$$\mathbb{E}(Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] [W_{t_k} - W_{t_{k-1}}]) =$$

$$\mathbb{E}[E(Y_{i-1} Y_{k-1} [W_{t_i} - W_{t_{i-1}}] [W_{t_k} - W_{t_{k-1}}] | \mathcal{F}_{t_{k-1}})] = 0.$$

Similarly,

$$\mathbb{E}(Y_{i-1}Y_j[W_{t_i} - W_{t_{i-1}}][W_t - W_j]) = 0.$$

Therefore

$$\mathbb{E}(Z_t^2) = \sum_{i=1}^j \mathbb{E}(Y_{i-1}^2[W_{t_i} - W_{t_{i-1}}]^2) + \mathbb{E}(Y_j^2[W_t - W_{t_j}]^2).$$

Note that

$$\begin{aligned} \mathbb{E}[Y_{i-1}^2[W_{t_i} - W_{t_{i-1}}]^2 \mid \mathcal{F}_{t_{i-1}}] &= Y_{i-1}^2 \mathbb{E}[(W_{t_i} - W_{t_{i-1}})^2 \mid \mathcal{F}_{t_{i-1}}] \\ &= Y_{i-1}^2 \mathbb{E}[(W_{t_i} - W_{t_{i-1}})^2] \\ &= Y_{i-1}^2(t_i - t_{i-1}). \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[Y_{i-1}^2[W_{t_i} - W_{t_{i-1}}]^2] &= \mathbb{E}(\mathbb{E}[Y_{i-1}^2[W_{t_i} - W_{t_{i-1}}]^2 \mid \mathcal{F}_{t_{i-1}}]) \\ &= \mathbb{E}(Y_{i-1}^2)(t_i - t_{i-1}). \end{aligned}$$

Similarly,

$$\mathbb{E}[Y_j^2[W_t - W_{t_j}]^2] = \mathbb{E}(Y_j^2)(t - t_j).$$

This proves (9.2).

To define the stochastic integral for betting rules Y_s that are not simple, we do the standard mathematical procedure for defining continuous objects—approximate by discrete and take a limit. Let Y_s be measurable with respect to \mathcal{F}_s , satisfying the second moment conditions listed above. A little more must be assumed about the Y_s to be mathematically precise: the paths of Y_s (i.e., Y_s considered as a function of s) should be right continuous and have left limits; we will not worry about this in our informal treatment. For each $n \geq 0$, define the approximate strategy $Y_s^{(n)}$ by

$$Y_s^{(n)} = n \int_{(k-1)/n}^{k/n} Y_r dr, \quad \frac{k}{n} < s \leq \frac{k+1}{n},$$

where we set $Y_s^{(n)} = 0$ for $s \leq 1/n$. We have arranged the approximation so that for each t , $Y_s^{(n)}$, $0 \leq s \leq t$ is a simple strategy that is \mathcal{F}_s -measurable. The key estimate that can be proved (we will not do it) is that

$$Y_s^{(n)} \rightarrow Y_s$$

in the sense that for each t

$$\lim_{n \rightarrow \infty} \int_0^t \mathbb{E}([Y_s - Y_s^{(n)}]^2) ds = 0.$$

This allows us to define the stochastic integral

$$Z_t = \int_0^t Y_s dW_s,$$

by saying that Z_t is the mean-square limit of the random variables

$$Z_t^{(n)} = \int_0^t Y_s^{(n)} dW_s.$$

The first and third properties of the stochastic integral allow this definition to work since as $n, m \rightarrow \infty$,

$$\mathbb{E}([Z_t^{(n)} - Z_t^{(m)}]^2) = \int_0^t \mathbb{E}([Y_s^{(n)} - Y_s^{(m)}]^2) ds \rightarrow 0.$$

In the process of showing the limit exists, one also shows that the three properties of the integral still hold.

Linearity:

$$\int_0^t [aX_s + bY_s] dW_s = a \int_0^t X_s dW_s + b \int_0^t Y_s dW_s.$$

Martingale Property: $Z_t = \int_0^t Y_s dW_s$ is a martingale with respect to \mathcal{F}_t . In particular, $\mathbb{E}(Z_t) = 0$ for all t .

Second Moment Calculation:

$$\text{Var}\left(\int_0^t Y_s dW_s\right) = \mathbb{E}\left[\left(\int_0^t Y_s dW_s\right)^2\right] = \int_0^t \mathbb{E}[Y_s^2] ds.$$

The relationship

$$Z_t = \int_0^t Y_s dW_s$$

is often written in the differential form

$$dZ_t = Y_t dW_t.$$

The process Z_t can be thought of as a process that at time t looks like a Brownian motion with variance parameter Y_t^2 (recall that if W_t is a standard Brownian motion, then σW_t is a Brownian motion with variance parameter σ^2 .) Sometimes one has a process

$$Z_t = \int_0^t X_s ds + \int_0^t Y_s dW_s,$$

where the “ ds ” integral is a standard calculus integral. In differential form this is written

$$dZ_t = X_t \, dt + Y_t \, dW_t.$$

This represents a process that at time t looks like a Brownian motion with drift X_t and variance parameter Y_t^2 .

9.3 Itô’s Formula

How does one calculate stochastic integrals? As an example, consider the integral

$$Z_t = \int_0^t W_s \, dW_s.$$

W_s is \mathcal{F}_s -measurable and this integral is well defined. One might hope that standard calculus rules would work for stochastic integrals in which case we would have

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} W_0^2 = \frac{1}{2} W_t^2.$$

However, a quick examination of this equation shows that it cannot be true: the left-hand side is a random variable with expectation 0 but the right-hand side has expectation $t/2$. In this section, we derive a formula that will allow us to calculate this integral exactly. This formula is usually called Itô’s formula and it is the fundamental theorem of stochastic calculus.

Let us start by reviewing the ordinary fundamental theorem of calculus. Suppose we have a continuously differentiable function $f(t)$. Around each t_0 we can expand $f(t)$,

$$f(t) = f(t_0) + f'(t_0)(t - t_0) + o(t - t_0).$$

We can write $f(t)$ as a telescoping sum

$$f(t) = f(0) + \sum_{j=0}^{n-1} \left[f\left(\frac{(j+1)t}{n}\right) - f\left(\frac{jt}{n}\right) \right].$$

We now use the Taylor’s series about jt/n to write

$$f\left(\frac{(j+1)t}{n}\right) = f\left(\frac{jt}{n}\right) + f'\left(\frac{jt}{n}\right) \frac{t}{n} + t o\left(\frac{1}{n}\right),$$

and

$$f(t) - f(0) = \sum_{j=0}^{n-1} f' \left(\frac{jt}{n} \right) \frac{t}{n} + \sum_{j=0}^{n-1} t o \left(\frac{1}{n} \right).$$

As $n \rightarrow \infty$ the second term on the right tends to 0 and the first term tends to the integral of f' . We therefore get

$$f(t) - f(0) = \int_0^t f'(s) ds,$$

which we all know very well.

Now let W_t be a Brownian motion, and f a function with at least two continuous derivatives. At each x_0 we can expand $f(x)$,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o((x - x_0)^2).$$

Write $f(W_t)$ as a telescoping sum,

$$f(W_t) = f(W_0) + \sum_{j=0}^{n-1} [f(W_{\frac{j+1}{n}t}) - f(W_{\frac{j}{n}t})].$$

By using the Taylor series expansion about $W_{\frac{j}{n}t}$ we can write

$$\begin{aligned} f(W_{\frac{j+1}{n}t}) &= f(W_{\frac{j}{n}t}) + f'(W_{\frac{j}{n}t})[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}] \\ &\quad + \frac{1}{2} f''(W_{\frac{j}{n}t})[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2 + t o \left(\frac{1}{n} \right). \end{aligned}$$

The $o(\cdot)$ is smaller than order n^{-1} since $[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2$ is of order (t/n) . We then get

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_{j=0}^{n-1} f'(W_{\frac{j}{n}t})[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}] \\ &\quad + \frac{1}{2} \sum_{j=0}^{n-1} f''(W_{\frac{j}{n}t})[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2 + \sum_{j=0}^{n-1} t o \left(\frac{1}{n} \right). \end{aligned} \quad (9.3)$$

As $n \rightarrow \infty$, the third term on the right goes to 0. Since f' is continuous, the first term will approach

$$\int_0^t f'(W_s) dW_s.$$

To see what the second term converges to, let us consider the general question of the limit of

$$\sum_{j=0}^{n-1} g(W_{\frac{j}{n}t})[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2,$$

where g is a continuous function. First consider the case where g is identically 1. Let

$$Q_t^{(n)} = \sum_{j=0}^{n-1} [W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2.$$

The limit

$$Q_t = \lim_{n \rightarrow \infty} Q_t^{(n)}$$

is often called the *quadratic variation* of W_t . $[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2$ has the same distribution as $(t/n)U^2$, where U is normal mean 0, variance 1. Note that

$$\mathbb{E}(U^2) = 1 \quad \text{Var}(U^2) = \mathbb{E}(U^4) - [\mathbb{E}(U^2)]^2 = 2.$$

Hence, since the increments of W are independent,

$$\mathbb{E}(Q_t^{(n)}) = \sum_{j=0}^{n-1} \mathbb{E}([W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2) = t,$$

$$\text{Var}(Q_t^{(n)}) = \sum_{j=0}^{n-1} \text{Var}([W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2) = n \text{Var}((t/n)U^2) = \frac{2t^2}{n}.$$

As $n \rightarrow \infty$, the expectation of $Q_t^{(n)}$ stays constant but the variance goes to 0. In other words, the limiting random variable Q_t is just a constant, and the quadratic variation of Brownian motion up to time t is the constant random variable equal to t .

For any g let

$$Q_t^{(n)}(g) = \sum_{j=0}^{n-1} g(t)[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}]^2,$$

and

$$Q_t(g) = \lim_{n \rightarrow \infty} Q_t^{(n)}(g).$$

If g is a step function of the form

$$g(s) = u(W_{\frac{j}{m}t}), \quad \frac{j}{m}t \leq s < \frac{j+1}{m}t,$$

then

$$\begin{aligned}
 Q_t(g) &= \lim_{n \rightarrow \infty} Q_t^{(n)}(g) \\
 &= \lim_{k \rightarrow \infty} Q_t^{(km)}(g) \\
 &= \lim_{k \rightarrow \infty} \sum_{j=0}^{m-1} u(W_{\frac{j}{m}t}) \sum_{i=0}^{k-1} [W_{\frac{kj+i+1}{km}t} - W_{\frac{kj+i}{km}t}]^2 \\
 &= \sum_{j=0}^{m-1} u(W_{\frac{j}{m}t}) \lim_{k \rightarrow \infty} \sum_{i=0}^{k-1} [W_{\frac{kj+i+1}{km}t} - W_{\frac{kj+i}{km}t}]^2.
 \end{aligned}$$

The result about quadratic variation tells us that

$$\lim_{k \rightarrow \infty} \sum_{i=0}^{k-1} [W_{\frac{kj+i+1}{km}t} - W_{\frac{kj+i}{km}t}]^2 = \frac{t}{m}.$$

Hence

$$Q_t(g) = \sum_{j=0}^{m-1} u(W_{\frac{j}{m}t}) \frac{t}{m}.$$

Now assume g is continuous. For each n , let g_n be the step function

$$g_n(s) = g(t), \quad \frac{j}{n} t \leq s < \frac{j+1}{n} t.$$

Note that

$$|Q_t(g) - Q_t(g_n)| \leq \|g - g_n\| Q_t = t \|g - g_n\|,$$

where

$$\|g - g_n\| = \sup_{0 \leq s \leq t} |g(s) - g_n(s)|.$$

The continuity of g implies that $\|g - g_n\| \rightarrow 0$ as $n \rightarrow \infty$. Hence

$$Q_t(g) = \lim_{n \rightarrow \infty} Q_t(g_n) = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} g\left(\frac{j}{n} t\right) \frac{t}{n}.$$

The last expression is the usual representation of the integral of g as a limit of Riemann sums. Therefore, if g is continuous,

$$Q_t(g) = \int_0^t g(s) ds.$$

Note that if h is continuous then since W_t is continuous, the function $g(t) = h(W_t)$ is continuous. If we plug this result into (9.3) we can conclude the following.

Itô's Formula. *If f is a function with two continuous derivatives, and W_t is a standard Brownian motion,*

$$f(W_t) - f(W_0) = \int_0^t f'(W_s) dW_s + \frac{1}{2} \int_0^t f''(W_s) ds.$$

This formula is sometimes written in the differential form,

$$df(W_t) = f'(W_t) dW_t + \frac{1}{2} f''(W_t) dt.$$

Example 1. Let $f(t) = t^2$. Then $f'(t) = 2t$, $f''(t) = 2$, and

$$W_t^2 = \int_0^t 2W_s dW_s + \frac{1}{2} \int_0^t 2 ds,$$

or

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t.$$

This turns out to be a particularly nice example; in general, one cannot use Itô's formula to calculate integrals exactly.

Example 2. Consider the process

$$X_t = e^{W_t}.$$

This process is called *geometric Brownian motion* and is often used to model stock prices. Itô's formula with $f(t) = e^t$ says that

$$X_t - 1 = \int_0^t e^{W_s} dW_s + \frac{1}{2} \int_0^t e^{W_s} ds.$$

In other words X_t satisfies the stochastic differential equation

$$dX_t = X_t dW_t + \frac{1}{2} X_t dt.$$

9.4 Extensions of Itô's Formula

Suppose W_t is a standard Brownian motion and Z_t satisfies

$$dZ_t = X_t dt + Y_t dW_t, \tag{9.4}$$

where X_t, Y_t are \mathcal{F}_t -measurable and have continuous paths. In other words,

$$Z_t = Z_0 + \int_0^t X_s \, ds + \int_0^t Y_s \, dW_s.$$

If R_t is \mathcal{F}_t -measurable we define $\int_0^t R_s \, dZ_s$ by

$$\begin{aligned} \int_0^t R_s \, dZ_s &= \int_0^t R_s (X_s \, ds + Y_s \, dW_s) \\ &= \int_0^t R_s X_s \, ds + \int_0^t R_s Y_s \, dW_s. \end{aligned}$$

Suppose f has two continuous derivatives. As in the previous section we can write

$$\begin{aligned} f(Z_t) - f(Z_0) &= \sum_{j=0}^{n-1} f'(Z_{\frac{j}{n}t}) [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}] \\ &\quad + \frac{1}{2} \sum_{j=0}^{n-1} f''(Z_{\frac{j}{n}t}) [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}]^2 + \sum_{j=0}^{n-1} t o\left(\frac{1}{n}\right). \end{aligned} \quad (9.5)$$

As $n \rightarrow \infty$, the last summation goes to zero. Since Z_t satisfies (9.4),

$$Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t} \approx X_{\frac{j}{n}t} \frac{t}{n} + Y_{\frac{j}{n}t} [W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}].$$

In the limit, we get

$$\begin{aligned} &\lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} f'(Z_{\frac{j}{n}t}) [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}] \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} f'(Z_{\frac{j}{n}t}) X_{\frac{j}{n}t} \frac{t}{n} + \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} f'(Z_{\frac{j}{n}t}) Y_{\frac{j}{n}t} [W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}] \\ &= \int_0^t f'(Z_s) X_s \, ds + \int_0^t f'(Z_s) Y_s \, dW_s \\ &= \int_0^t f'(Z_s) \, dZ_s. \end{aligned}$$

Similarly to the last section, we can see that

$$\lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} f''(Z_{\frac{j}{n}t}) [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}]^2 = \int_0^t f''(Z_s) d\langle Z \rangle_s,$$

where $\langle Z \rangle_t$ denotes the quadratic variation of Z_t ,

$$\langle Z \rangle_t = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}]^2.$$

If we consider only the quadratic variation of the stochastic integral part of Z_t , we get

$$\begin{aligned} \langle \int_0^t Y_s dW_s \rangle_t &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} \left[\int_{jt/n}^{(j+1)t/n} Y_s dW_s \right]^2 \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} ([Y_{\frac{j}{n}t} + o(1)] [W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}])^2 \\ &= \int_0^t Y_s^2 ds. \end{aligned}$$

We have left out a number of details here but the basic idea is the same as in the previous section. Also,

$$Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t} = O\left(\frac{1}{n}\right) + \int_{jt/n}^{(j+1)t/n} Y_s dW_s,$$

and hence

$$[Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}]^2 = O\left(\frac{1}{n^{3/2}}\right) + \left[\int_{jt/n}^{(j+1)t/n} Y_s dW_s \right]^2.$$

We therefore get

$$\begin{aligned} \langle Z \rangle_t &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} [Z_{\frac{j+1}{n}t} - Z_{\frac{j}{n}t}]^2 \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} \left[\int_{jt/n}^{(j+1)t/n} Y_s dW_s \right]^2 \\ &= \langle \int_0^t Y_s dW_s \rangle_t = \int_0^t Y_s^2 ds. \end{aligned}$$

In other words, the quadratic variation of Z is the same as the quadratic variation of its “stochastic integral” part. Combining all of this we get the following.

Itô's Formula II. *If f has two continuous derivatives and Z_t satisfies (9.4), then*

$$\begin{aligned} f(Z_t) - f(Z_0) &= \int_0^t f'(Z_s) dZ_s + \frac{1}{2} \int_0^t f''(Z_s) d\langle Z \rangle_s \\ &= \int_0^t f'(Z_s) Y_s dW_s + \int_0^t [f'(Z_s) X_s + \frac{1}{2} f''(Z_s) Y_s^2] ds. \end{aligned}$$

We now generalize a little more and assume that $f(t, x)$ is a function of both time t and space x . We will need to assume that f has two continuous derivatives in x , and one continuous derivative in t . We will write $f'(t, x)$, $f''(t, x)$ for the partials with respect to x and $\dot{f}(t, x)$ for the partial with respect to t . We can expand $f(t, Z_t) - f(0, Z_0)$ into two telescoping sums

$$\begin{aligned} f(t, Z_t) - f(0, Z_0) &= \sum_{j=0}^{n-1} \left[f\left(\frac{j+1}{n}t, Z_{\frac{j+1}{n}t}\right) - f\left(\frac{j}{n}t, Z_{\frac{j+1}{n}t}\right) \right] \\ &\quad + \sum_{j=0}^{n-1} \left[f\left(\frac{j}{n}t, Z_{\frac{j+1}{n}t}\right) - f\left(\frac{j}{n}t, Z_{\frac{j}{n}t}\right) \right]. \end{aligned}$$

Using the approximation

$$f\left(\frac{j+1}{n}t, Z_{\frac{j+1}{n}t}\right) - f\left(\frac{j}{n}t, Z_{\frac{j+1}{n}t}\right) \approx \frac{t}{n} \dot{f}\left(\frac{j}{n}t, Z_{\frac{j+1}{n}t}\right),$$

it can be shown that

$$\lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} \left[f\left(\frac{j+1}{n}t, Z_{\frac{j+1}{n}t}\right) - f\left(\frac{j}{n}t, Z_{\frac{j+1}{n}t}\right) \right] = \int_0^t \dot{f}(s, Z_s) \, ds.$$

The limit of the second telescoping sum can be handled as before. This gives the following.

Itô's Formula III. *If $f(t, x)$ has two continuous derivatives in x and one continuous derivative in t , and Z_t satisfies (9.4), then*

$$\begin{aligned} f(t, Z_t) - f(0, Z_0) &= \int_0^t \dot{f}(s, Z_s) \, ds + \int_0^t f'(s, Z_s) \, dZ_s + \frac{1}{2} \int_0^t f''(s, Z_s) \, d\langle Z \rangle_s \\ &= \int_0^t f'(s, Z_s) Y_s \, dW_s \\ &\quad + \int_0^t [\dot{f}(s, Z_s) + f'(s, Z_s) X_s + \frac{1}{2} f''(s, Z_s) Y_s^2] \, ds. \end{aligned}$$

A particular case of this formula occurs when $Z_t = W_t$. Then,

$$\begin{aligned} f(t, W_t) - f(0, W_0) &= \\ \int_0^t f'(s, W_s) \, dW_s + \int_0^t &[\dot{f}(s, W_s) + \frac{1}{2} f''(s, W_s)] \, ds. \end{aligned}$$

Example 1. Let $f(t, x) = e^{at+bx}$ where a, b are real numbers. Then $Z_t = e^{at+bW_t}$ satisfies the stochastic differential equation

$$dZ_t = b Z_t dW_t + \left(a + \frac{b^2}{2}\right) Z_t dt.$$

Equivalently, the solution to the equation

$$dZ_t = r Z_t dt + b Z_t dW_t$$

is

$$Z_t = \exp \left\{ b W_t + \left(r - \frac{b^2}{2}\right) t\right\}. \quad (9.6)$$

Another generalization comes from considering Brownian motion in more than one dimension. Suppose $W_t = (W_t^1, \dots, W_t^d)$ is a standard d -dimensional Brownian motion and $f(x^1, \dots, x^d)$ is a function from \mathbb{R}^d to \mathbb{R} that has continuous second derivatives. If we expand f in a Taylor series about $x = (x^1, \dots, x^d)$, we get

$$\begin{aligned} f(y) &= f(x) + \sum_{i=1}^d f_i(x) (y^i - x^i) \\ &\quad + \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d f_{jk}(x) (y^j - x^j) (y^k - x^k) + o(|y - x|^2). \end{aligned}$$

Here we use subscripts to denote partial derivatives. As before, we can write $f(W_t)$ as a telescoping sum to show that

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_{i=1}^d \left[\lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} f_i(W_{\frac{l}{n}t}) (W_{\frac{l+1}{n}t}^i - W_{\frac{l}{n}t}^i) \right] \\ &\quad + \sum_{j=1}^d \left[\frac{1}{2} \lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} f_{jj}(W_{\frac{l}{n}t}) (W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j)^2 \right] \\ &\quad + \sum_{j \neq k} \left[\frac{1}{2} \lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} f_{jk}(W_{\frac{l}{n}t}) (W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j) (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k) \right]. \end{aligned}$$

The first two terms are of the type we have seen. To find the limit of the last one, we show that the “covariation”

$$\langle W^j, W^k \rangle_t := \lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} (W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j) (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k) = 0.$$

This is done similarly to the quadratic variation in the previous section. In this case,

$$\begin{aligned} & \mathbb{E} \left[\sum_{l=0}^{n-1} (W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j) (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k) \right] \\ &= \sum_{l=0}^{n-1} \mathbb{E} \left[(W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j) (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k) \right] \\ &= \sum_{l=0}^{n-1} \mathbb{E} \left(W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j \right) \mathbb{E} \left(W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k \right) = 0, \end{aligned}$$

and the variance of the random variable is

$$\begin{aligned} & \sum_{l=0}^{n-1} \text{Var}[(W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j) (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k)] \\ &= \sum_{l=0}^{n-1} \mathbb{E} [(W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j)^2 (W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k)^2] \\ &= \sum_{l=0}^{n-1} \mathbb{E} [(W_{\frac{l+1}{n}t}^j - W_{\frac{l}{n}t}^j)^2] \mathbb{E} [(W_{\frac{l+1}{n}t}^k - W_{\frac{l}{n}t}^k)^2] \\ &= \sum_{l=0}^{n-1} \frac{t}{n} \frac{t}{n} \longrightarrow 0. \end{aligned}$$

Therefore, the last term in the telescoping sum for $f(W_t) - f(W_0)$ vanishes in the limit. If f also has a t -dependence, it can be handled as above. We now summarize. Recall that the Laplacian of f is defined by

$$\Delta f(x) = \sum_{j=1}^d f_{jj}(x).$$

Itô's Formula IV. Suppose $f(t, x^1, \dots, x^d)$ is a function with one continuous derivative in t and two continuous derivatives in $x = (x^1, \dots, x^d)$. Suppose $W_t = (W_t^1, \dots, W_t^d)$ is a standard d -dimensional Brownian motion. Then,

$$f(t, W_t) - f(0, W_0) = \sum_{i=1}^d \int_0^t f_i(W_s) dW_s^i + \int_0^t [\dot{f}(W_s) + \frac{1}{2} \Delta f(W_s)] ds.$$

Stochastic calculus is similar to usual calculus with an additional rule added. Let us consider calculus from a differential perspective. If $h(t)$ is a function, and $\Delta h(t) = h(t + \Delta t) - h(t)$, then $h'(t)$ is defined by the rule

$$\Delta h(t) = h'(t) \Delta t + o(\Delta t), \quad \Delta t \rightarrow 0.$$

To calculate $h'(t)$, we calculate $\Delta h(t)$ and then throw away all the terms that are $o(\Delta t)$. For example, suppose $h(t) = f(t)g(t)$ where f and g are differentiable. Then

$$\begin{aligned}\Delta h(t) &= f(t + \Delta t)g(t + \Delta t) - f(t)g(t) \\ &= f(t + \Delta t)[g(t + \Delta t) - g(t)] + g(t)[f(t + \Delta t) - f(t)] \\ &= [f(t) + f'(t)\Delta t + o(\Delta t)][g'(t)\Delta t + o(\Delta t)] \\ &\quad + g(t)[f'(t)\Delta t + o(\Delta t)] \\ &= [f(t)g'(t) + f'(t)g(t)]\Delta t + o(\Delta t).\end{aligned}$$

This gives the product rule $(fg)' = fg' + f'g$.

If W_t^1, \dots, W_t^d are independent Brownian motions, then ΔW_t^j , the increment of the Brownian motion, is of order $\sqrt{\Delta t}$. Hence, if we multiply two of them together, we get something of order Δt , which cannot be thrown away. If we multiply three of them together, or if we multiply one of them times a term of order Δt , then the product is of order $(\Delta t)^{3/2}$ and can be thrown away. So, in order to do stochastic calculus one needs only add to usual calculus the rule for handling products of two Brownian increments. Ito's formula tells us what to do. In differential notation, we have

$$(\Delta W_t^j)^2 = \Delta \langle W^j \rangle_t = \Delta t,$$

$$(\Delta W_t^j)(\Delta W_t^k) = \Delta \langle W^j, W^k \rangle_t = 0, \quad j \neq k.$$

More generally, if

$$dZ_t^1 = X_t^1 dt + \sum_{j=1}^d Y_t^{j,1} dW_t^j, \tag{9.7}$$

$$dZ_t^2 = X_t^2 dt + \sum_{j=1}^d Y_t^{j,2} dW_t^j, \tag{9.8}$$

then the covariation term is

$$\langle Z^1, Z^2 \rangle_t = \sum_{j=1}^d \int_0^t (Y_t^{j,1} Y_t^{j,2}) dt,$$

i.e.,

$$d\langle Z^1, Z^2 \rangle_t = \sum_{j=1}^d (Y_t^{j,1} Y_t^{j,2}) dt.$$

This allows us to derive the stochastic calculus product rule. Note that we can write

$$\Delta(Z_t^1 Z_t^2) = Z_{t+\Delta t}^1 \Delta Z_t^2 + Z_t^2 \Delta Z_t^1 = Z_t^1 \Delta Z_t^2 + Z_t^2 \Delta Z_t^1 + \Delta Z_t^1 \Delta Z_t^2.$$

Product Rule. If Z_t^1, Z_t^2 satisfy (9.7) and (9.8), then

$$d(Z_t^1 Z_t^2) = Z_t^1 dZ_t^2 + Z_t^2 dZ_t^1 + d\langle Z^1, Z^2 \rangle_t. \quad (9.9)$$

Example 2. Exponential Martingale. Suppose $dZ_t = Y_t dW_t$, so that Z_t is a martingale. Itô's formula shows that

$$d[e^{Z_t}] = e^{Z_t} Y_t dW_t + \frac{1}{2} e^{Z_t} Y_t^2 dt = e^{Z_t} dZ_t + \frac{1}{2} e^{Z_t} d\langle Z \rangle_t.$$

Assume sufficient boundedness so that $\mathbb{E}[e^{Z_t}] < \infty$; boundedness of Y_t is sufficient. One can see from the differential equation that e^{Z_t} is a submartingale (i.e., $E(e^{Z_t} | \mathcal{F}_s) \geq e^{Z_s}$) but not a martingale (if Y is nonzero). One way to obtain a martingale is to subtract the “ dt ” term. Another way is to multiply e^{Z_t} by an appropriate process. Let $M_t = e^{Z_t} R_t$ where

$$R_t = \exp \left\{ -\frac{1}{2} \int_0^t Y_s^2 ds \right\}.$$

Note that R_t is random but differentiable; in fact, $\dot{R}_t = -(Y_t^2/2) R_t$. Since R_t is differentiable, $\langle e^Z, R \rangle_t = 0$ (since $\Delta(e^{Z_t} R_t)$ is of order $(\Delta t)^{3/2}$). Therefore by the product rule we get,

$$dM_t = R_t d(e^{Z_t}) + e^{Z_t} dR_t = M_t Y_t dW_t = M_t dZ_t.$$

Hence, M_t is a martingale. This is sometimes called the *exponential martingale* since it satisfies an stochastic differential equation analogous to the exponential differential equation $f'(t) = a f(t)$.

9.5 Continuous Martingales

If W_t is a standard Brownian motion; Y_t is measurable with respect to \mathcal{F}_t , the information in $W_s, 0 \leq s \leq t$; and

$$\int_0^t \mathbb{E}[Y_s^2] ds < \infty,$$

then

$$M_t = \int_0^t Y_s dW_s, \quad (9.10)$$

is a square-integrable martingale, i.e., a martingale with respect to $\{\mathcal{F}_t\}$ satisfying $\mathbb{E}[M_t^2] < \infty$. It is also a *continuous martingale* which means that with probability one the function $t \mapsto M_t$ is continuous. Many of the results from Chapter 5 have analogues for continuous martingales which can be proved with little extra effort. Note that if $\delta > 0$, then $\hat{M}_n := M_{\delta n}$ is a (discrete time) martingale with respect to $\hat{\mathcal{F}}_n = \mathcal{F}_{\delta n}$. If T is a stopping time with respect to \mathcal{F}_t we define the stopping time $T^{(\delta)}$ as the smallest integer n such that $\delta n \geq T$. To determine whether or not $T^{(\delta)} = n$ it suffices to see the Brownian motion W_t up through time δn ; therefore $T^{(\delta)}$ is a stopping time for the discrete time martingale. By letting $\delta \rightarrow 0$, the following extensions of results from Chapter 5 can be established.

Optional Sampling Theorem I. *If M_t is a continuous martingale and T is a bounded stopping time with respect to $\{\mathcal{F}_t\}$, then*

$$\mathbb{E}[M_T] = M_0.$$

Optional Sampling Theorem II. *If M_t is a continuous martingale and T is a stopping time with respect to $\{\mathcal{F}_t\}$ satisfying $\mathbb{P}\{T < \infty\} = 1$,*

$$\mathbb{E}[|M_T|] < \infty,$$

and

$$\lim_{t \rightarrow \infty} \mathbb{E}[|M_t| \mathbf{1}\{|T| > t\}] = 0,$$

then

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Maximal Inequality. *If M_t is a continuous square-integrable martingale of the form (9.10), then for every $a > 0$,*

$$\mathbb{P}\left\{\max_{0 \leq s \leq t} |M_s| \geq a\right\} \leq \frac{\mathbb{E}[M_t^2]}{a^2} = \frac{1}{a^2} \int_0^t \mathbb{E}[Y_s^2] ds.$$

If M_t is a continuous martingale with respect to \mathcal{F}_t and T is a stopping time then $J_t = M_{t \wedge T}$ is a continuous martingale. Here $t \wedge T = \min\{t, T\}$. Suppose U is an open subset of \mathbb{R}^d and $f(t, x^1, \dots, x^d)$ is a continuous function that

has one continuous derivative in t and two continuous derivatives in the spatial variable provided $x = (x^1, \dots, x^d) \in U$. Suppose Z_t satisfies (9.4) and let T denote the first time t such that Z_t is not in U . Then Itô's formula describes the evolution of $f(t \wedge T, Z_{t \wedge T})$ for $t < T$. As an example, suppose W_t is a standard d -dimensional Brownian motion and U is a bounded open set in \mathbb{R}^d . Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuous function such that $\Delta f(x) = 0$ for x in U . Then, Itô's formula shows that $M_t = f(W_{t \wedge T})$ is a continuous martingale. If $W_0 \in \bar{U}$, then M_t is a bounded martingale (since f is a bounded function on the compact set \bar{U}), and therefore if $x \in \bar{U}$,

$$f(x) = \mathbb{E}[M_0 \mid W_0 = x] = \mathbb{E}[M_T \mid W_0 = x] = \mathbb{E}[f(W_T) \mid W_0 = x].$$

9.6 Girsanov Transformation

Suppose that we play a simple game. A coin is flipped. If it comes up heads we win \$1; otherwise, we lose a \$1. However, suppose the coin is unfair so that it has probability $3/4$ of coming up tails each time. Then this is an unfair game. There are two natural ways to try to make this a fair game.

- Change the payoff so that we win \$1.50 if it comes up heads and lose only .50 if it comes up tails. In this case the expected winning is zero.
- Change (or replace) the coin so that the probability of a heads is $1/2$.

In this section, we will discuss a way of changing a continuous process with drift to a process without drift that is analogous to the second option above.

Suppose Z_t satisfies

$$dZ_t = X_t dt + Y_t dW_t, \quad (9.11)$$

where W_t is a standard Brownian motion. We let \mathcal{F}_t denote the information in $\{W_s : s \leq t\}$, and we assume that X_t, Y_t are \mathcal{F}_t -measurable. If X_t is nonzero, then Z_t is not a martingale. One way to get a martingale from Z_t is to subtract the “ dt ” term. This is analogous to the first option in the previous paragraph. We will describe another way to obtain a martingale, analogous to the second option, which is called the *Girsanov* or *Cameron-Martin transformation*.

Instead of subtracting the drift, we will change the weight on paths. By giving greater weight to those paths that are moving in the direction opposite the drift, we will balance things so that the average drift is zero. To illustrate the idea, we will start with a discrete example. Suppose J_1, J_2, \dots are independent random variables with

$$\mathbb{P}\{J_j = 1\} = 1 - \mathbb{P}\{J_j = -1\} = p,$$

where $0 < p < 1$. Let $S_0 = 0, S_n = J_1 + \dots + J_n$, and let \mathcal{F}_n denote the information contained in J_1, \dots, J_n . If $p \neq 1/2$, then S_n is not a martingale with respect to \mathcal{F}_n . While $S_n - n(2p - 1)$ is a martingale, we will consider a different martingale obtained by keeping the same paths but changing the measure. Our process S_n can be considered as a measure \mathbb{P} on random walk paths of length n that gives measure $p^{(n+S_n)/2} (1-p)^{(n-S_n)/2}$ to each particular path (note that the number of first n steps that are “+1” is $(n + S_n)/2$ and the number of steps that are “-1” is $(n - S_n)/2$). We can write

$$p^{(n+S_n)/2} (1-p)^{(n-S_n)/2} = [4p(1-p)]^{n/2} \left(\frac{p}{1-p}\right)^{S_n/2} 2^{-n}.$$

Let

$$M_n = [4p(1-p)]^{-n/2} \left(\frac{1-p}{p}\right)^{S_n/2}.$$

We define a measure on paths $\tilde{\mathbb{P}}$ by $\tilde{\mathbb{P}} = M_n \mathbb{P}$. To be more precise, if A is \mathcal{F}_n -measurable, then

$$\tilde{\mathbb{P}}(A) = \mathbb{E}[I_A M_n],$$

where I_A denotes the indicator function. Note that $\tilde{\mathbb{P}}$ gives measure 2^{-n} to each path. In particular, the process S_n , under the measure $\tilde{\mathbb{P}}$, is a martingale.

To generalize this idea, we will give a characterization of the weighting function M_n . What makes this work is the fact that both M_n and $M_n S_n$ are martingales (under the measure \mathbb{P}), see Exercise 5.10. We need M_n to be a martingale in order for the measure to be well defined as we now demonstrate. Suppose A is measurable with respect to \mathcal{F}_m and $m < n$. Then A is also \mathcal{F}_n -measurable, so the two formulas for $\tilde{\mathbb{P}}(A)$ should give the same answer. But, since M_n is a martingale,

$$\mathbb{E}[M_n I_A] = \mathbb{E}[E(M_n I_A | \mathcal{F}_m)] = \mathbb{E}[I_A E(M_n | \mathcal{F}_m)] = \mathbb{E}[M_m I_A].$$

In order for S_n to be a martingale under the measure $\tilde{\mathbb{P}}$, we need to show if $m < n$, then

$$E_{\tilde{\mathbb{P}}}(S_n | \mathcal{F}_m) = S_m.$$

Here, $E_{\tilde{\mathbb{P}}}(S_n | \mathcal{F}_m)$ denotes the conditional expectation using the measure $\tilde{\mathbb{P}}$. Using the definition of conditional expectation, we see that this equality is equivalent to showing for all events A that are \mathcal{F}_m -measurable,

$$\mathbb{E}[1_A S_m M_m] = \mathbb{E}[1_A S_n M_n].$$

But, this is just another way of saying that $E[M_n S_n | \mathcal{F}_m] = M_m S_m$, i.e., that $M_n S_n$ is a martingale.

We return to continuous case and assume that Z_t satisfies (9.11). The new weight will be given in terms of a nonnegative martingale M_t (with respect to \mathcal{F}_t) with $M_0 = 1$. We will define a new measure $\tilde{\mathbb{P}}$ by the relation “ $d\tilde{\mathbb{P}} = M_t d\mathbb{P}$ ”. To be more precise, if A is an \mathcal{F}_t -measurable event, then

$$\tilde{\mathbb{P}}(A) = \mathbb{E}[I_A M_t].$$

If $s < t$, and A is \mathcal{F}_s -measurable, then it is also \mathcal{F}_t -measurable, so it may look like $\tilde{\mathbb{P}}$ is not well defined. However, since M_t is a martingale,

$$\mathbb{E}[I_A M_t] = \mathbb{E}[E(I_A M_t | \mathcal{F}_s)] = \mathbb{E}[I_A E(M_t | \mathcal{F}_s)] = \mathbb{E}[I_A M_s].$$

This shows that $\tilde{\mathbb{P}}(A)$ is well defined. We say that M_t is the *Radon-Nikodym derivative* of $\tilde{\mathbb{P}}$ with respect to \mathbb{P} .

We want to choose M_t so that Z_t is a $\tilde{\mathbb{P}}$ -martingale, i.e., a martingale if we use the measure $\tilde{\mathbb{P}}$. This will be true if M_t and $M_t Z_t$ are both martingales (with respect to \mathbb{P}). This can be seen using an argument as in the discrete case above.

Suppose M_t is a martingale of the form $dM_t = R_t dW_t$. Then the product rule tells us that

$$\begin{aligned} d[M_t Z_t] &= M_t dZ_t + Z_t dM_t + d\langle M, Z \rangle_t \\ &= [M_t X_t + R_t Y_t] dt + [M_t Y_t + Z_t R_t] dW_t \end{aligned}$$

If $R_t = -M_t X_t / Y_t$ and certain boundedness conditions hold, then this will be a martingale.

Girsanov transformation. If Z_t satisfies (9.11) and M_t is a martingale satisfying

$$dM_t = -\frac{X_t}{Y_t} M_t dW_t,$$

then Z_t is a martingale with respect to the measure $\tilde{\mathbb{P}}$ where

$$d\tilde{\mathbb{P}} = M_t d\mathbb{P}.$$

Example 1. Suppose Z_t is Brownian motion with drift, i.e.,

$$dZ_t = \mu dt + dW_t.$$

We want $R_t = -\mu M_t$, so we need M_t to satisfy the equation

$$dM_t = -\mu M_t dW_t.$$

The solution to this is

$$M_t = e^{-\mu W_t - (\mu^2/2)t} = \frac{e^{-\mu W_t}}{\mathbb{E}[e^{-\mu W_t}]}.$$

Hence if we weight Brownian motion with drift by M_t we get standard Brownian motion. Note that if $\mu > 0$, then M_t is larger for paths with W_t (and hence Z_t) smaller.

Example 2. Suppose Z_t satisfies

$$dZ_t = r Z_t dt + b Z_t dW_t,$$

see (9.6). Then we would like to find M_t satisfying $dM_t = -(r/b) M_t dW_t$. We have seen that

$$M_t = \exp\{-(r/b)W_t - (r/b)^2 t/2\}$$

satisfies this. Hence Z_t is a $\tilde{\mathbb{P}}$ -martingale where $d\tilde{\mathbb{P}} = M_t d\mathbb{P}$.

9.7 Feynman-Kac Formula

Suppose Z_t satisfies the stochastic differential equation

$$dZ_t = a(Z_t) dt + b(Z_t) dW_t, \quad (9.12)$$

where $a(x), b(x)$ are fixed functions. Such a Z_t is often called a (*time homogeneous*) *diffusion*. Note that Z_t is Markovian, i.e., the dependence of the future $\{Z_s : s \geq t\}$ on the past \mathcal{F}_t lies entirely on the value Z_t . There is a close relationship between diffusions and certain second order partial differential equations.

Suppose $f(x), v(x)$ are two functions and let

$$J_t = \exp \left\{ \int_0^t v(Z_s) ds \right\},$$

$$V(t, x) = \mathbb{E}^x[f(Z_t) J_t].$$

Here $\mathbb{E}^x[Y]$ denotes $\mathbb{E}[Y | Z_0 = x]$. We assume that this expectation exists for all t, x . If $s < t$, then

$$\begin{aligned} E[f(Z_t) J_t | \mathcal{F}_s] &= J_s E \left[f(Z_t) \exp \left\{ \int_s^t v(Z_r) dr \right\} | \mathcal{F}_s \right] \\ &= J_s V(t-s, Z_s). \end{aligned}$$

The left-hand side is a martingale since if $r < s$, then

$$E[E[f(Z_t) J_t | \mathcal{F}_s] | \mathcal{F}_r] = E[f(Z_t) J_t | \mathcal{F}_r].$$

Hence, we know that if $M_s = J_s V(t-s, Z_s)$, then M_s is a martingale for $0 \leq s \leq t$. Assuming sufficient differentiability, we can use Itô's formula and the product rule (9.9) to write

$$\begin{aligned} dM_s &= J_s dV(t-s, Z_s) + V(t-s, Z_s) \dot{J}_s ds \\ &= J_s [v(Z_s) V(t-s, Z_s) - \dot{V}(t-s, Z_s) + V'(t-s, Z_s) a(Z_s) \\ &\quad + \frac{1}{2} V''(t-s, Z_s) b^2(Z_s)] ds + J_s V'(t-s, Z_s) b(Z_s) dW_s. \end{aligned}$$

Since M_s is a martingale, the dt term must always be zero, and V satisfies

$$\dot{V}(t, x) = \frac{1}{2} b^2(x) V''(t, x) + a(x) V'(t, x) + v(x) V(t, x).$$

Feynman-Kac Formula. *The solution to the partial differential equation*

$$\dot{V}(t, x) = \frac{1}{2} b^2(x) V''(t, x) + a(x) V'(t, x) + v(x) V(t, x)$$

with initial condition $V(0, x) = f(x)$ is

$$V(t, x) = \mathbb{E}^x \left[f(Z_t) \exp \left\{ \int_0^t v(Z_s) ds \right\} \right],$$

where Z_t satisfies (9.12).

By setting $v = 0$, we see that $V(t, x) = \mathbb{E}^x[f(Z_t)]$ satisfies

$$\dot{V}(t, x) = \frac{1}{2} b^2(x) V''(t, x) + a(x) V'(t, x).$$

We can write

$$\mathbb{E}^x[f(Z_t)] = \int_{-\infty}^{\infty} f(y) p(t, x, y) dy,$$

where $p(t, x, \cdot)$ denotes the density of the random variable Z_t assuming $Z_0 = x$. If we fix x , then $p(t, y) = p(t, x, y)$ is the solution to the equation with initial condition “delta function” at x . In particular, p satisfies

$$\dot{p}(t, y) = \frac{1}{2} b^2(y) p''(t, y) + a(y) p'(t, y).$$

In the next section we will need a Feynman-Kac formula for a time inhomogeneous diffusion

$$dZ_t = a(t, Z_t) dt + b(t, Z_t) dW_t. \tag{9.13}$$

Let $v(t, x), f(x)$ be given functions. We fix a t_0 and consider only $0 \leq t \leq t_0$. Let

$$J_t = \exp \left\{ \int_0^t v(s, Z_s) ds \right\},$$

and let

$$V(t, x) = \mathbb{E} \left[f(Z_{t_0}) \exp \left\{ \int_t^{t_0} v(s, Z_s) ds \right\} \mid Z_t = z \right].$$

Then,

$$E[f(Z_{t_0}) J_{t_0} \mid \mathcal{F}_t] = J_t V(t, Z_t).$$

Since the left-hand side is a martingale, so is the right-hand side. Using the product rule and Itô's formula we see that

$$-\dot{V}(t, x) = \frac{1}{2} b^2(t, Z_t) V''(t, x) + a(t, Z_t) V'(t, z) + v(t, Z_t) V(t, x) = 0. \quad (9.14)$$

Note that $V(t_0, x) = f(x)$.

Feynman-Kac Formula II. *The solution to (9.14) for $0 \leq t < t_0$ with $V(t_0, x) = f(x)$ is*

$$V(t, x) = \mathbb{E}^x \left[f(Z_{t_0}) \exp \left\{ \int_t^{t_0} v(s, Z_s) ds \right\} \right], \quad (9.15)$$

where Z_t satisfies (9.13).

9.8 Black-Scholes Formula

The Black-Scholes formula is a way to calculate the current value of an option that is based on the price of a stock following a stochastic differential equation. Suppose S_t denotes the price of a stock, and S_t satisfies

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

By (9.6), the solution of this is

$$S_t = S_0 \exp \left\{ \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right\}.$$

Assume also that one can buy or sell a bond with guaranteed interest rate r . If we let Y_t be the amount of money invested in bonds, then if we do not buy or sell any bonds the amount grows according to the equation

$$dY_t = r Y_t dt.$$

A *European call option* (with strike price K at time T) is an opportunity to buy one share of the stock at time T for price K . If $S_T \leq K$ such an

option is useless, but if $S_T > K$, then it has a value of $S_T - K$, which is the profit obtained by buying the stock and then selling it immediately. We can write the value as $(S_T - K)_+$ where $x_+ = \max\{x, 0\}$. The *Black-Scholes formula* determines the value of this option at a time $t \leq T$ under the assumption that there are no *arbitrage* opportunities. Let V_t denote this value. Clearly $V_T = (S_T - K)_+$, and V_t should be measurable with respect to \mathcal{F}_t , the information at time t . It is reasonable to assume that $V_t = V(t, S_t)$; we will determine this function. Note that $V(T, x) = (x - K)_+$.

We can think of the option as an asset with value V_t at time $t \leq T$. Suppose we sell such an option at time $t < T$ and invest the money in a portfolio consisting of a combination of the stock and the bond, say X_t shares of the stock and Y_t invested in the bond. We assume we have a buying and selling strategy between bonds and stocks based on the stock price at a certain time. Here Y_t is determined by the X_t and the relationship that stocks are bought only with money obtained from selling bonds and vice versa.

The value of the total portfolio (one option sold plus the total of assets in bonds and stocks) at time s is

$$U_t = -V(t, S_t) + O_t,$$

where

$$O_t = X_t S_t + Y_t. \quad (9.16)$$

For ease, let us assume that $U_0 = 0$, i.e., at time $t = 0$ we sold one option and invested that money in some combination of bond and stock.

Suppose we monitor this investment up to time T (switching between shares of the stock and the bond based on the price of the stock) using a strategy that *guarantees* that $U_T \geq 0$. If it is also true that with positive probability $U_T > 0$, then we have found a way to gain money (with positive probability) without any risk. This is called an *arbitrage*. Similarly, if there is a strategy to guarantee $U_T \leq 0$ with a chance that $U_T < 0$, then there are arbitrage possibilities by buying an option. The main assumption in the Black-Scholes formula is: there are no arbitrage opportunities with “self-financing” strategies.

The self-financing assumption is that the change in the total value of the bond/stock portfolio is given by

$$dO_t = X_t dS_t + r Y_t dt. \quad (9.17)$$

In other words, the change in the value is the number of shares of stock times the change in stock price plus the number of units of the bond times the change in bond price. Assuming (9.17), we can use Itô’s formula to write

$$\begin{aligned} dU_t &= -dV(t, S_t) + dO_t \\ &= -\dot{V}(t, S_t) dt - V'(t, S_t) dS_t - \frac{1}{2} V''(t, S_t) d\langle S \rangle_t \\ &\quad + X_t dS_t + r Y_t dt \end{aligned}$$

Now, to remove the randomness from the value of the portfolio we choose $X_t = V'(t, S_t)$. This makes the coefficient of dW_t zero and

$$dU_t = \left[-\dot{V}(t, S_t) - \frac{1}{2}V''(t, S_t)\sigma^2 S_t^2 + r Y_t \right] dt. \quad (9.18)$$

The assumption of no arbitrage tells us that this must equal zero.

Using the product rule (9.9) on (9.16), we see that

$$dO_t = X_t dS_t + dY_t + S_t dX_t + d\langle X, S \rangle_t.$$

Hence, the self-financing condition (9.17) can be written as

$$dY_t = r Y_t dt - S_t dX_t - d\langle X, S \rangle_t.$$

Since $X_t = V'(t, S_t)$, Itô's formula gives

$$\begin{aligned} dX_t &= [\dot{V}'(t, S_t) + V''(t, S_t)\mu S_t + \frac{1}{2}V'''(t, S_t)\sigma^2 S_t^2] dt \\ &\quad + V''(t, S_t)\sigma S_t dW_t. \end{aligned}$$

Hence Y_t must satisfy

$$\begin{aligned} dY_t &= r Y_t dt - \left[\dot{V}'(t, S_t) S_t + V''(t, S_t)(\mu + \sigma^2) S_t^2 \right. \\ &\quad \left. + \frac{1}{2}V'''(t, S_t)\sigma^2 S_t^3 \right] dt - V''(t, S_t)\sigma S_t^2 dW_t. \end{aligned} \quad (9.19)$$

Let

$$Y_t = V(t, S_t) - S_t X_t = V(t, S_t) - S_t V'(t, S_t),$$

and assume that $V(t, x)$ has been chosen so the quantity in (9.18) vanishes, i.e.,

$$\dot{V}(t, x) + \frac{1}{2}x^2\sigma^2 V''(t, x) + r x V'(t, x) - r V(t, x) = 0. \quad (9.20)$$

Then an Itô's formula calculation shows that (9.19) holds.

One can get lost in the calculation, so it is worth understanding why it works. If there are no arbitrage opportunities and the option is priced properly, then any strategy that produces no randomness must also produce no gain or loss. Hence the current value of the portfolio, O_t , must also be the price of the option at that time, i.e., $V(t, S_t) = O_t$. Since we know that we must have $V'(t, S_t)$ shares of the stock to hedge the option, the assets in bonds must be

$$Y_t = O_t - X_t S_t = V(t, S_t) - V'(t, S_t) S_t.$$

Plugging into (9.18) we get the *Black-Scholes equation* (9.20).

Note that the Black-Scholes equation has r and σ^2 as parameters but μ does not appear! The value of the option depends only on the bond rate and the variance parameter (sometimes called the *volatility*) σ^2 . We need to find the solution of this equation with boundary condition $V(T, x) = (x - K)_+$. The dependence on r can be removed by a simple change of variables: if \tilde{V} satisfies (9.20) with $r = 0$,

$$\dot{\tilde{V}}(t, x) + \frac{1}{2} x^2 \sigma^2 \tilde{V}''(t, x) = 0, \quad (9.21)$$

and $\tilde{V}(t, x) = e^{r(t-T)} V(t, e^{r(T-t)} x)$, then $\tilde{V}(t, x)$ satisfies (9.20) and $\tilde{V}(T, x) = V(T, x)$. This can be checked by differentiation (Exercise 9.7); however, there is a simple reason why this is true. If money grows at rate r , then x dollars at time T is the equivalent of $e^{r(t-T)} x$ dollars at time t . Hence, it suffices to solve the equation when $r = 0$.

A probabilistic form for the solution of (9.21) is given by the Feynman-Kac formula (9.15); in fact, this form can be used for options with different payoffs $V(T, x) = g(x)$. Assume $r = 0$. Remembering that $V(t, S_t) = O_t$, we get

$$dV(t, S_t) = V'(t, S_t) dS_t.$$

If $V(t, S_t)$ were a martingale, we would know that

$$\mathbb{E}[V(t, S_t)] = \mathbb{E}[V(T, S_T)] = \mathbb{E}[g(S_T)].$$

Recall that S_t satisfies

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

This is a martingale only if $\mu = 0$. However, we have seen that the value of the option does not depend on the value of μ , so we can set $\mu = 0$. If $\mu = 0$ the solution to the stochastic differential equation is

$$S_t = \exp \left\{ \sigma W_t - \frac{\sigma^2}{2} t \right\}.$$

Then we have

$$\begin{aligned} V(T-t, x) &= \mathbb{E}[g(S_t) \mid S_t = x] \\ &= \mathbb{E} \left[g \left(\exp \left\{ \sigma W_t - \frac{\sigma^2 t}{2} \right\} \right) \mid W_t = \frac{\log x}{\sigma} \right] \\ &= \mathbb{E} \left[g \left(x e^{-\sigma^2 t/2} e^{\sigma \sqrt{t} N} \right) \right], \end{aligned}$$

where N is a standard unit normal.

Suppose $g(y) = (y - K)_+$. Then,

$$V(T-t, e^{\sigma^2 t/2} x) = \mathbb{E}[(x e^{\sigma \sqrt{t} N} - K)_+].$$

A straightforward, although tedious, calculation (see Exercise 9.4) shows that the right-hand side is

$$x e^{\sigma^2 t/2} \Phi\left(\frac{\log(x/K) + \sigma^2 t}{\sigma \sqrt{t}}\right) - K \Phi\left(\frac{\log(x/K)}{\sigma \sqrt{t}}\right),$$

where Φ denote the standard normal distribution function. Hence $V(T-t, x)$ is given by

$$x \Phi\left(\frac{\log(x/K) + (1/2) \sigma^2 t}{\sigma \sqrt{t}}\right) - K \Phi\left(\frac{\log(x/K) - (1/2) \sigma^2 t}{\sigma \sqrt{t}}\right),$$

This is the solution for $r = 0$, and we can easily convert it to the solution for general r .

Black-Scholes Formula. Suppose $V(t, x)$ is the solution to (9.19) satisfying $V(T, x) = (x - K)_+$. Then $V(T-t, x)$ equals

$$x \Phi\left(\frac{\log(x/K) + (r + \frac{\sigma^2}{2})t}{\sigma \sqrt{t}}\right) - K e^{-rt} \Phi\left(\frac{\log(x/K) + (r - \frac{\sigma^2}{2})t}{\sigma \sqrt{t}}\right),$$

where Φ is the standard normal distribution function.

Let us generalize and assume that S_t satisfies

$$dS_t = \mu(t, S_t) S_t dt + \sigma(t, S_t) S_t dW_t,$$

where $\mu(t, x), \sigma(t, x)$ are given functions. We cannot give an explicit solution to this stochastic differential equation. However, we can still give an expression for the value of a European call option. We assume that we have a self-financing portfolio with value $O_t = X_t S_t + Y_t$ that “hedges” the option. If $V(t, x)$ denotes the value of the option, then we choose $X_t = -V'(t, S_t)$ in order to remove the randomness. Assuming no arbitrage, the value of the portfolio using the hedging strategy is exactly the same as the value of the option at that time. Therefore $Y_t = O_t - X_t S_t = V(t, S_t) - V'(t, S_t) S_t$. Hence, we again obtain the Black-Scholes equation (9.20) where σ^2 is replaced with $\sigma^2(t, x)$. We need to find the solution to

$$\dot{V}(t, x) + \frac{1}{2} \sigma^2(t, x) V''(t, x) + r x V'(t, x) - r V(t, x) = 0,$$

with $V(T, x) = g(x)$. Note again that $\mu(t, x)$ does not appear in the equation. In most cases, there is no closed form for this solution. However, the Feynman-Kac formula (9.15) gives the value in terms of an expectation that can be estimated by simulation.

9.9 Simulation

Consider a stochastic differential equation

$$dX_t = a(X_t) dt + b(X_t) dW_t,$$

where a and b are relatively nice functions of x and W_t denotes a standard Brownian motion. The solution is a process X_t that at any particular time looks like a Brownian motion with drift parameter $a(X_t)$ and variance parameter $b(X_t)$. While it is often difficult to give an explicit solution to the equation, it is easy to simulate the process on a computer using a random walk.

Choose some small number Δt . We can approximate the Brownian motion by a simple random walk with time increments Δt and space increments $\sqrt{\Delta t}$. To do this let Y_1, Y_2, \dots be independent random variables with

$$\mathbb{P}\{Y_i = 1\} = \mathbb{P}\{Y_i = -1\} = \frac{1}{2}.$$

We set $X_0 = 0$ and for $n > 0$,

$$X_{n\Delta t} = X_{(n-1)\Delta t} + a(X_{(n-1)\Delta t})\Delta t + b(X_{(n-1)\Delta t})\sqrt{\Delta t} Y_n.$$

In practice, it is often just as easy to make the increments normal. If Z_1, Z_2, \dots are independent standard unit normals, we can set $X_0 = 0$ and for $n > 0$,

$$X_{n\Delta t} = X_{(n-1)\Delta t} + a(X_{(n-1)\Delta t})\Delta t + b(X_{(n-1)\Delta t})\sqrt{\Delta t} Z_n.$$

9.10 Exercises

9.1 Let W_t be a standard one-dimensional Brownian motion with $W_0 = 1$ and let r be a real number. Let T be the first time that $W_t = 0$. Let $R_t = W_t^r$.

(a) Write the stochastic differential equation for R_t (valid for $t < T$), i.e., find f, g such that

$$dR_t = f(R_t) dt + g(R_t) dW_t.$$

(b) Find a function F such that $M_{t \wedge T}$ is a martingale where

$$M_t = R_t \exp \left\{ \int_0^t F(R_s) ds \right\}.$$

9.2 Let $d > 1$ and let W_t denote a standard d -dimensional Brownian motion starting at $x \neq 0$. Let $M_t = \log|W_t|$ if $d = 2$ and $M_t = |W_t|^{2-d}$ if $d > 2$. Show that M_t is a martingale.

9.3 Let W_t be a standard one-dimensional Brownian motion and let $a, b > 0$. Let $T_{a,-b}$ be the first time t such that $W_t = a$ or $W_t = -b$.

- (a) Use the martingale W_t to find $\mathbb{P}\{W_{T_{a,-b}} = a\}$.
- (b) Use the martingale $W_t^2 - t$ to find $\mathbb{E}[T_{a,-b}]$.
- (c) Explain why the random variables $T_{a,-a}$ and $W_{T_{a,-a}}$ are independent.
- (d) Are the random variables $T_{a,-b}$ and $W_{T_{a,-b}}$ independent for all a, b ?
- (e) Use the martingale $e^{\lambda W_t - (\lambda/2)t}$ to compute the moment generating function for $T_{a,-a}$.

9.4 Suppose N is a standard unit normal and $X = ae^{bN}$ where $a, b > 0$. Show that the density of X is

$$f(x) = \frac{1}{xb} \phi\left(\frac{\log(x/a)}{b}\right), \quad 0 < x < \infty,$$

where $\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the density for N . If $K > 0$, show that

$$\begin{aligned} \int_0^\infty (x - K)_+ f(x) dx &= \\ a e^{b^2/2} \Phi\left(\frac{\log(a/K) + b^2}{b}\right) - K \Phi\left(\frac{\log(a/K)}{b}\right), \end{aligned}$$

where Φ denotes the distribution function for N .

9.5 Let X_1, X_2, \dots be independent $N(0, 1)$ random variables and let f be a bounded continuous function. Let $Z_0 = 0$ and for $n > 0$,

$$Z_n = Z_{n-1} + f(Z_{n-1}) + X_n.$$

We will do the Girsanov transformation for Z_n to make Z_n a martingale (with respect to \mathcal{F}_n , where \mathcal{F}_n is the information in X_1, \dots, X_n).

(a) If a is a real number, compute $\mathbb{E}[X_1 e^{aX_1}]$. (One can do it directly, or one can differentiate the moment generating function $\mathbb{E}[e^{aX_1}]$ with respect to a .)

(b) Let $M_0 = 1$ and for $n > 0$,

$$M_n = \exp \left\{ - \sum_{j=1}^n f(Z_{j-1}) X_j - \sum_{j=1}^n \frac{f(Z_{j-1})^2}{2} \right\}.$$

Show that M_n is a martingale with respect to \mathcal{F}_n .

(c) Show that $M_n Z_n$ is a martingale with respect to \mathcal{F}_n .

(d) Show that Z_n is a $\tilde{\mathbb{P}}$ -martingale where $d\tilde{\mathbb{P}} = M_n d\mathbb{P}$.

9.6 Suppose W_t is a standard one-dimensional Brownian motion. Suppose $Z_0 = 1$ and Z_t satisfies the Bessel equation

$$dZ_t = \frac{a}{Z_t} dt + dW_t.$$

Here a is a real number and we only consider $t < T = \min\{s : Z_s = 0\}$.

- (a) Find a nonconstant differentiable function ϕ such that $M_t = \phi(Z_{t \wedge T})$ is a martingale. (Hint: use Itô's formula to find a differential equation that ϕ should satisfy and then solve the equation.)
- (b) If $0 < \epsilon < 1 < \alpha$ and $S = S(\epsilon, \alpha)$ denotes the first time t such that $Z_t = \epsilon$ or $Z_t = \alpha$, find $\mathbb{P}\{Z_S = \epsilon\}$.
- (c) Find the probability that there exists some time t with $Z_t = \epsilon$. For which values of a is this probability equal to one?
- (d) For which values of a does the process reach the origin in finite time?

9.7 Show that if $V(t, x)$ satisfies (9.21), then $\tilde{V}(t, x) := e^{r(t-T)} V(t, e^{r(T-t)}x)$ satisfies (9.20).

9.8 COMPUTER SIMULATION. Assume X_t is a process satisfying the stochastic differential equation

$$dX_t = a(X_t) dt + b(X_t) dW_t,$$

where

$$a(x) = 0,$$

$$b(x) = \begin{cases} 2, & x > 0 \\ 1, & x < 0. \end{cases}$$

Using $\Delta t = 1/100$ run many simulations of X_t . Estimate the following

- (a) $\mathbb{E}(X_1)$,
- (b) $\mathbb{P}\{X_1 > 0\}$ You may wish to use both ± 1 and normal increments and compare the results.

9.9 Do Exercise 9.8 with

$$a(x) = x,$$

$$b(x) = |x|^{3/4}.$$

Suggestions for Further Reading

There are many possibilities for additional reading. We make a few suggestions here, but this is not intended to be a complete list.

Background in probability at an undergraduate level:

G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press.

J. Pitman, *Probability*, Springer-Verlag.

Stochastic processes at the level of this book:

G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press.

S. Karlin and H. Taylor, *A First Course in Stochastic Processes* and *A Second Course in Stochastic Processes*, Academic Press.

S. Resnik, *Adventures in Stochastic Processes*, Birkhäuser.

To pursue stochastic processes at a higher level, it is necessary to have a background in advanced calculus (undergraduate real analysis) and measure theory. One possibility for each of these is:

R. Strichartz, *The Way of Analysis*, Jones and Bartlett Mathematics.

R. Bartle, *The Elements of Integration and Lebesgue Measure*, Wiley.

The next step is to learn probability at a measure-theoretic level. These books contain some of the measure theory as well:

P. Billingsley, *Probability and Measure*, Wiley.

R. Durrett, *Probability: Theory and Examples*, Thomson Brooks/Cole.

J. Jacod & P. Protter, *Probability Essentials*, Springer-Verlag.

D. Williams, *Probability with Martingales*, Cambridge University Press.

For treatments of Brownian motion and stochastic calculus using measure-theoretic probability theory:

K. Chung & R. Williams, *An Introduction to Stochastic Integration*, Birkhäuser.

R. Durrett, *Stochastic Calculus: A Practical Introduction*, CRC Press.

I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag.

B. Øksendal, *Stochastic Differential Equations*, Springer-Verlag.

Index

- aperiodic 22
arbitrage 224
- backward equation 82
birth-and-death processes 74-81
Black-Scholes formula 227
branching process 53-57, 116, 127
Brownian motion 173-198
 geometric 209
 one-dimensional 174
 several dimensions 184
 standard 174
 with drift 193
 zero set 181
- Cantor set 182-183
Chapman–Kolmogorov equation 44, 177, 184, 194
communication classes 20
conditional expectation 101-109
convolution 137
- difference equations 3-6
differential equations 1-3
Dirichlet problem 92, 186
discounting 96
Doob maximal inequality 123, 124, 217
- equilibrium distribution, see invariant distribution
explosion 81
exponential alarm clocks 69, 72
exponential distribution 68
exponential martingale 216
extinction probability 55-57
- Feynman-Kac formula 221-223
- forward equation 82
fractal dimension 181-182
- gambler's ruin 30-31, 112-113, 187
generating function 55
Gibbs samplers 165
Girsanov transformation 220
graph 12
- harmonic function 119, 168, 218
heat equation 186-189
- infinitesimal generator 70
invariant distribution 15, 51-52, 73, 78
- irreducibility 20, 76
Ising model 164
Ito's formula 205-216
- Jensen's inequality 125
- Markov chains
 countable 43
 finite 9
 finite, continuous-time 68-74
 Monte Carlo 162-166
 reversible 155
- Markov property 1, 9, 176-177
martingale 106, 217
martingale betting strategy 107-108
martingale convergence theorem 117
measurable 103
Metropolis algorithms 165
- null recurrence 51-53, 78
- optimal stopping 87-97
optional sampling theorem 112, 115, 217

- periodic 21-24
- Perron–Frobenius theorem 17, 40-41
- Poisson process 65-68, 131
- Polya's urn 109, 116-117, 119, 122
- positive recurrence 51-53, 78
- quadratic variation 51-51, 207, 211
- queues 10, 44-45
 - $G/M/1$ 150-151
 - $M/G/1$ 133, 148-149
 - $M/M/k$ 75
- random harmonic series 115-116, 119
- random walk
 - absorbing boundary 12, 18, 30-31
 - biased 12
 - graph 12, 21, 31
 - partially reflecting 44, 52-53
 - reflecting boundary 11, 18, 29
 - simple 44, 46-49
 - symmetric 12
- recurrence 50-53, 77, 119-120, 189-191
- recurrent class 20, 29-30
- reflection principle 122, 178
- renewal equation 138
- renewal process 131
 - age 133, 138-141, 145
 - central limit theorem 135
 - law of large numbers 134
 - lifetime 141-142, 147
 - residual life 141-142, 146
- renewal theorems 136-137
- return times 25, 51, 131-132
- self-financing 224
- simple strategy 201
- state space 1
- stationary distribution, see invariant distribution
- steady-state distribution, see invariant distribution
- Stirling's formula 47-63
- stochastic integral 199-228
- stochastic matrix 10
- stochastic process 1
- stopping time 88, 110, 177
- strong Markov property 147
- substochastic matrix 27
- submartingale 109, 123
- superharmonic function 62, 89
- supermartingale 109
- transience 50-53, 77, 119-120, 189-191
- transient class 20, 26-30
- transition matrix 10
- uniform integrability 114-116
- value 89
- waiting times 67-68, 69
- Wald's equation 129, 149
- Wiener process 174
- Yule process 76, 79-80

Introduction to Stochastic Processes

Focusing on mathematical ideas rather than proofs, **Introduction to Stochastic Processes, Second Edition** provides quick access to important foundations of probability theory applicable to problems in many fields. Approaching all problems and theorems without any measure theory, the book provides a concise and informal introduction to stochastic processes evolving with time.

Here's what's new in the Second Edition:

- Expanded chapter on stochastic integration that introduces modern mathematical finance
- Expanded discussion of Itô's formula including Girsanov theory, the Feynman-Kac formula, and the Black-Scholes formula in stochastic integration
- New topics such as Doob's maximal inequality and a discussion on self similarity in the chapter on Brownian motion

This concise, informal introduction is designed to meet the needs of students and professionals not only in mathematics and statistics, but in the many fields in which the concepts presented are also important, including computer science, economics, business, biological sciences, psychology, and engineering. It acquaints readers with the possibilities of applying stochastic processes in their work.

