Graduate Texts in Mathematics

GTM

Rabi Bhattacharya
Edward C. Waymire

Random Walk, Brownian Motion, and Martingales

Springer

Graduate Texts in Mathematics 292

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

More information about this series at http://www.springer.com/series/136

Rabi Bhattacharya • Edward C. Waymire

Random Walk, Brownian Motion, and Martingales

Springer

Rabi Bhattacharya
Department of Mathematics
The University of Arizona
Tucson, AZ, USA

Edward C. Waymire
Department of Mathematics
Oregon State Univeristy
Corvallis, OR, USA

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-030-78937-4 ISBN 978-3-030-78939-8 (eBook)
https://doi.org/10.1007/978-3-030-78939-8

Mathematics Subject Classification: 60G05, 60G07, 60G17, 60G18, 60G22, 60G42, 60G50, 60G60, 60J65, 60J80

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To

Gouri (In Loving Memory)

and

Linda

Preface

The theory of stochastic processes has numerous and longstanding significant interactions with pure and applied mathematics as well as most branches of physical and biological sciences. Our aim is to present a graduate-level course built on four pillars of the subject—random walk, branching processes, Brownian motion, and martingales. Much of the theory is developed by building on simple examples. This approach helps to develop intuition and it provides a specific context in which to check more subtle aspects of the proofs.

The prerequisite is a one-semester/quarter of graduate level probability. In detail, it should include the following: (1) Caratheodory's theorem for construction of measures, integration of functions on a measure space, dominated and monotone convergence theorems, and the Radon–Nikodym theorem; (2) Kolmogorov's existence theorem for construction of probability measures on infinite product spaces, and independence; (3) characteristic functions; (4) the strong law of large numbers and the classical central limit theorem; and (5) conditional expectation. There are many excellent texts and online resources for this purpose. In addition, throughout this text the authors' footnotes to the second edition of their text *A Basic Course in Probability Theory*, denoted BCPT, provide specific page or chapter references to the prerequisite material in analysis and probability as needed.

Chapter 1 begins with technical definitions of stochastic processes in discrete and continuous time, and of random fields, illustrated with examples. The text moves on to the first major topic, namely random walks. Chapters 2, 3, and 7 provide a comprehensive account of the simple random walk in one dimension, made up of successive sums of i.i.d. random variables, each taking a value $+1$ with probability p, and a value -1 with probability $q = 1 - p, 0 < p < 1$. The one-dimensional version of the Einstein/Smoluchowski theory of Brownian motion is derived from it later under the topic of Brownian motion. Chapter 4 is devoted to simple d-dimensional symmetric random walk whose steps assign equal probability (namely $1/2d$) to each of the $2d$ neighboring points of the origin in the d-dimensional integer lattice. Fascinating contrasts in the dynamics occur for different values of d.

Chapter 5 on Poisson and compound Poisson processes provides well-known continuous time analogues of random walks. That is, they are processes with independent increments and are therefore Markov. The study of general continuous time processes, as well as random fields, involves some technical issues such as sample path continuity. These are covered in Chapter 6, which provides broad criteria for regularity due to Kolmogorov and Chentsov. In particular, it is shown that Brownian motion has continuous sample paths.

Chapter 7 shows that processes with independent increments, such as random walks, Poisson and compound processes, as well as Brownian motion, have the strong Markov property, a stronger and highly useful version of the Markov property. Chapter 8 deals with a different kind of asymptotic behavior of general lattice-valued random walks using an important technique referred to as coupling. The specific limit theorems concern renewal theory and the problem of estimating the speed of convergence to equilibrium for a class of Markov chains. Chapter 9 rounds out the study of discrete time processes with branching processes, another important class of Markov chains with important applications in a number of fields including biology and physics.

In addition to the construction and sample path regularity of Brownian motion in Chapter 6, and its strong Markov property in Chapter 7, the functional central limit theorem, proved in Chapter 17, is a cornerstone theorem linking Brownian motion as a universal space-time scaling limit of a random walk having finite second moments. It makes precise the idea that Brownian motion is approximately a random walk under appropriate scaling of time and space. Chapters 16 through 20 study further properties of Brownian motion and related processes, often from this point of view.

Martingale theory, the final major topic, is presented in Chapters 10 through 13. The theory, mostly due to Joseph L. Doob, has virtually revolutionized probability theory. This theory is now indispensable to the modeling and analysis of various phenomena in stochastic processes. Chapter 15 on the martingale characterization of the Poisson process and Chapter 14 on super-critical Binaymé–Galton–Watson branching process provide further illustrations of some of the power of martingale theory.

The last part of the book, Chapters 21 through 28, presents a number of important contemporary applications of stochastic processes. These are somewhat more specialized, or more technical, than the theory presented in the main text. Nonetheless, they are topics of much recent interest. For example, the rather deep general renewal theorem of Blackwell is essential for modern theory of ruin in insurance; the binomial tree model of mathematical finance is of much interest in financial mathematics; two different and widely discussed views of Hurst phenomena are still a matter of debate. Essential features of the modern theory of branching excursions, branching random walks, and multiplicative cascades appear in Chapters 21 and 22, while the final chapter is on a distinct but related probabilistic treatment of the global solvability of the $3d$ incompressible Navier–Stokes equation, one of the most important unsolved problems in mathematics. The instructor may choose one or two of these topics depending on time available and the interest of the class.

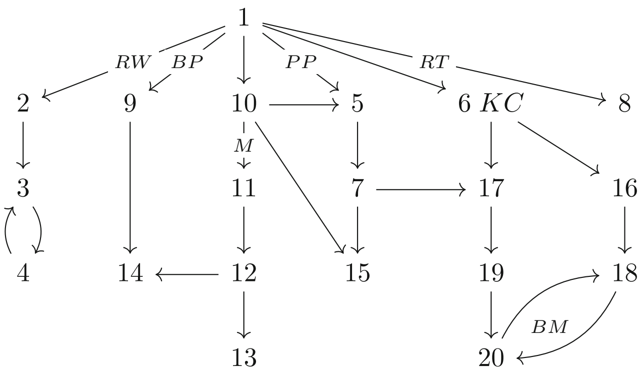
In general, journal and book attributions are cited as Author (year) in the text, or as footnotes, as they occur. Complete references to these are provided under *References* in the end pages. The list *Related Books and Monographs* is a compilation of some related textbooks and research monographs, including those cited in the text.

The authors are grateful to Springer Editor Loretta Bartolini for patience and guidance. We would like to thank our colleagues William Faris and Enrique Thomann for their continued advice and encouragement. We are grateful to Josie O'Harrow for technical assistance in providing TIKZ code for the figures that appear in the text. The authors gratefully acknowledge partial support (DMS1811317, DMS-1408947) of the National Science Foundation during the preparation of this book.

Tucson, AZ, USA — Rabi Bhattacharya
Corvallis, OR, USA — Edward C. Waymire

Ten Week Course Suggestions
(A) Random Walk and Brownian Motion: 1–4, 6–8,10–11,16–20
(B) General Stochastics: 1–4, 5, 7, 9–12,14,16–20

RW = Random Walk
BP = Branching Process
PP = Poisson Process
BM = Brownian Motion
KC = KolmogorovŰChentsov
RT = Renewal Theory
M = Martingale



Chapter Dependency Diagram

Double arrow indicates related material but not required in respective proofs.

Contents

Symbol Definition List xv

1 What Is a Stochastic Process? 1
Exercises 15

2 The Simple Random Walk I: Associated Boundary Value Distributions, Transience, and Recurrence 17
Exercises 22

3 The Simple Random Walk II: First Passage Times 27
Exercises 37

4 Multidimensional Random Walk 41
Exercises 44

5 The Poisson Process, Compound Poisson Process, and Poisson Random Field 47
Exercises 56

6 The Kolmogorov–Chentsov Theorem and Sample Path Regularity 61
Exercises 68

7 Random Walk, Brownian Motion, and the Strong Markov Property 71
Exercises 93

8 Coupling Methods for Markov Chains and the Renewal Theorem for Lattice Distributions 99
Exercises 109

9 Bienaymé–Galton–Watson Simple Branching Process and Extinction 113
Exercises 121

10 Martingales: Definitions and Examples 123
Exercises 132

11 Optional Stopping of (Sub)Martingales 135
Exercises 147

12 The Upcrossings Inequality and (Sub)Martingale Convergence 151
Exercises 161

13 Continuous Parameter Martingales 163
Exercises 171

14 Growth of Supercritical Bienaymé–Galton–Watson Simple Branching Processes 173
Exercises 183

15 Stochastic Calculus for Point Processes and a Martingale Characterization of the Poisson Process 185
Exercises 190

16 First Passage Time Distributions for Brownian Motion with Drift and a Local Limit Theorem 191
Exercises 197

17 The Functional Central Limit Theorem (FCLT) 199
Exercises 211

18 ArcSine Law Asymptotics 215
Exercises 224

19 Brownian Motion on the Half-Line: Absorption and Reflection 227
Exercises 232

20 The Brownian Bridge 235
Exercises 240

21 Special Topic: Branching Random Walk, Polymers, and Multiplicative Cascades 243
Exercises 260

22 Special Topic: Bienaymé–Galton–Watson Simple Branching Process and Excursions 263
Exercises 274

23 Special Topic: The Geometric Random Walk and the Binomial Tree Model of Mathematical Finance 279
Exercises 288

24 Special Topic: Optimal Stopping Rules 291
Exercises 301

25 Special Topic: A Comprehensive Renewal Theory for General Random Walks 305
Exercises 328

26 Special Topic: Ruin Problems in Insurance 329
Exercises 346

27 Special Topic: Fractional Brownian Motion and/or Trends: The Hurst Effect 347
Exercises 361

28 Special Topic: Incompressible Navier–Stokes Equations and the Le Jan–Sznitman Cascade 363
Exercises 376

References 379

Author Index 387

Subject Index 391

Symbol Definition List

Special Sets and Functions:

In the classic notation of G.H. Hardy, one writes $a(x) = O(b(x))$ to mean that there is a constant c (independent of x) such that $|a(x)| \leq c|b(x)|$ for all x. Also $a(x) = o(b(x))$ indicates that the ratio $a(x)/b(x) \to 0$ according to specified limit.

$\mathbb{Z}_+$, set of non-negative integers

$\mathbb{Z}_{++}$, set of positive integers

$\mathbb{R}_+$, set of non-negative real numbers

$\mathbb{R}_{++}$, set of positive real numbers

∂A, boundary of set A

A^o, interior of set A

A^-, closure of set A

A^c, complement of set A

$\mathbf{1}_B(x)$, indicator of the set B

$[X \in B]$, inverse image of the set B under X

$\#A$, $|A|$, cardinality for finite set A

δ_x Dirac delta (point mass)

$\otimes$, σ-field product

$\tau|n$, restriction of tree graph to first n generations

$v|n$, restriction of tree vertex to first n generations

$\overline{c, d}$

$\bigotimes$, product of $\sigma-$fields

$p(t; x, dy)$, homogeneous (stationary) transition probability

$p(s, t : x, dy)$, nonhomogeneous (nonstationary) transition probability

Function Spaces, Elements and Operations:

$C[0, 1]$, set of continuous, real-valued functions defined on [0, 1]

$\mathbb{R}^\infty$, infinite sequence space

$C([0, \infty) : \mathbb{R}^k)$, set of continuous functions on $[0, \infty)$ with values in $\mathbb{R}^k$

$C_b(S)$, set of continuous bounded, real-valued functions on a metric (or topological) space S

$B(S)$, set of bounded, measurable real-valued functions on a measurable space $(S, \mathcal{S})$

$C_b^0(S)$, continuous functions on a metric or topological space vanishing at infinity

$C(S : \mathbb{C})$, set of complex-valued functions on S

$\otimes_{e_\xi}$, Navier–Stokes projected convolution

e_i $\mathbf{e}$ i-th coordinate of unit vector

$i.o.$ infinitely often

$f * g$, convolution of functions

$Q_1 * Q_2$, convolution of probabilities

Cov, covariance

Var, variance

$\Rightarrow$, weak convergence

A^t, v^t matrix transpose

Chapter 1
What Is a Stochastic Process?

This chapter provides the mathematical framework and example illustrations of stochastic processes as families of random variables with values in some measurable (state) space S, such as the integers or the real line or higher dimensional Euclidean space, and indexed by some set Λ. Examples include i.i.d. sequences, random walks, Brownian motion, Poisson processes, branching processes, queue processes, Markov processes, and various martingale processes. Special emphasis is given to existence and constructions of these important classes of stochastic processes.

As remarked in the ***Preface****, throughout the text, the author's footnote references to the second edition of Bhattacharya and Waymire (2016) A Basic Course in Probability Theory, denoted BCPT, are used as an Appendix for prerequisite material in analysis and probability as needed. However, there are many excellent texts and on-line resources that can be used for this purpose.*

A stochastic process $X = \{X_t : t \in \Lambda\}$ is a family of random variables defined on a probability space that are generally linked by evolutionary rules indexing time, or interactions indexed by space. The temperature records X_t at the tth (integer) unit of time may be viewed as a family of random variables $\{X_0, X_1, \ldots\}$ indexed by the *discrete-time parameter* $t \in \mathbb{Z}_+$. The number X_t of page clicks on a given website during the (continuous) time interval $[0, t]$ gives rise to a collection of random variables $\{X_t : t \geq 0\}$ indexed by the *continuous-time parameter* t. The velocity X_t at a point t in a turbulent wind field provides a family of random variables $\{X_t : t \in \mathbb{R}^3\}$ indexed by a *multidimensional spatial parameter* t.

Given an index set Λ, a *stochastic process* indexed by Λ is a collection of random variables $\{X_t : t \in \Lambda\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a set S, equipped with a σ-field $\mathcal{S}$. The set S, or more precisely the (pair) measurable space $(S, \mathcal{S})$, is called the *state space* of the process. Typically S is a metric space

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_1

and $\mathcal{S}$ is its Borel σ-field generated by the open subsets of S. In particular, in the above, one may take, respectively: (i) $\Lambda = \mathbb{Z}_+$, $S = \mathbb{R}_+$; (ii) $\Lambda = [0, \infty)$, $S = \mathbb{Z}_+$; (iii) $\Lambda = \mathbb{R}^3$, $S = \mathbb{R}^3$, and accordingly with their respective Borel σ-fields under an implicitly assumed standard discrete or Euclidean metric. For the most part we shall study stochastic processes indexed by a one-dimensional set of real numbers (e.g., generally thought of as time). Here the natural ordering of numbers coincides with the sense of evolution of the process. The selection of discrete or continuous units to index a stochastic process is a model choice, but *linear order* is lost for stochastic processes indexed by a multidimensional parameter; such processes are usually referred to as *random fields*. More precisely one can make the following definition.

Definition 1.1. Let $(S, \mathcal{S})$ be a measurable space and Λ a non-empty set. A family $\mathbf{X} := \{X_t : t \in \Lambda\}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ taking values in S is referred to as a *stochastic process* with *state space* $(S, \mathcal{S})$ and *index set* Λ.

Remark 1.1. Sometimes one stipulates that $X_t \in S_t$ for $t \in \Lambda$, where each $(S_t, \mathcal{S}_t)$ is a measurable space. Typically each S_t is a Borel subset of a complete metric space S, with Borel σ-field $\mathcal{S}$, topologically referred to as a *standard space*. Also S_t may be referred to as the state space of X_t, and S as the state space for the process.

A somewhat generic picture is to view X_t as the state of a randomly moving particle in S at time t. The *distribution* of X_t for fixed index point $t \in \Lambda$ is a probability Q_t on $(S, \mathcal{S})$ defined by $Q_t(B) := P(X_t \in B)$, $B \in \mathcal{S}$. That is, the probability P defined on $(\Omega, \mathcal{F})$ induces a probability Q_t on $(S, \mathcal{S})$ via the map $X_t : \Omega \to S$ through inverse images defined and denoted by $[X_t \in B] \equiv X_t^{-1}(B) := \{\omega \in \Omega : X_t(\omega) \in B\}$, $B \in \mathcal{S}$; *the square bracket notation will continue to be used to denote inverse images of random variables throughout this text. That said, writing "$P(X \in B)$" in place of "$P([X \in B])$" is a minor abuse of notation that is also employed.* A little more generally, given a finite set of distinct indices $\{t_1, \ldots, t_m\} \subset \Lambda$ the corresponding *finite-dimensional distribution* of $(X_{t_1}, \ldots, X_{t_m})$ is defined by

$$Q_{t_1,\ldots,t_m}(B) = P((X_{t_1}, \ldots, X_{t_m}) \in B), \quad B \in \mathcal{S}^{\otimes m}. \tag{1.1}$$

Within the theory of stochastic processes it is useful to consider the distribution of states in terms of the more general distribution Q induced by the entire process $\mathbf{X} = \{X_t : t \in \Lambda\}$, viewed as a "random path" $t \to X_t$, within an appropriate *path space* $(\Gamma, \mathcal{G})$ where, for example, $\Gamma \subset S^\Lambda$ is a set of functions (sample paths) defined on Λ and taking values in S. Γ can typically also be viewed as a metric space and $\mathcal{G}$ its corresponding Borel σ-field. To illustrate, suppose that one has a stochastic process $\mathbf{X}$ on a probability space $(\Omega, \mathcal{F}, P)$, taking values in $S = \mathbb{R}$ whose *sample paths* $t \to X_t(\omega), 0 \le t \le 1$, $(\omega \in \Omega)$, are continuous. Then one might naturally consider $Q(G) = P(\mathbf{X} \in G) \equiv P([\mathbf{X} \in G])$ where G is a Borel subset of the metric space $\Gamma = C[0, 1]$, with supremum distance between functions (paths) in $C[0, 1]$. Path space distributions will be further illustrated in the forthcoming examples.

The following result of Kolmogorov asserts, conversely, the existence of a unique probability measure P on the product space $(\Omega = S^{\Lambda}, \mathcal{F} = \mathcal{S}^{\otimes\Lambda})$, such that, given a family of distributions $Q_{t_1,\ldots,t_m}$ on $(S^{\{t_1,\ldots,t_m\}}, \mathcal{S}^{\otimes\{t_1,\ldots,t_m\}})$ for all $m = 1, 2, \ldots,$ and all m-tuples $(t_1, \ldots, t_m)$ of distinct indices in Λ, the coordinate projections $X_t : \Omega \to S$ given by $X_t(\omega) = \omega(t), (\omega = \{\omega(t) : t \in \Lambda\})$, have the family of finite-dimensional distributions (1.1), provided some obvious simple consistency requirements and possible topological conditions are satisfied by the given probability distributions $Q_{t_1,\ldots,t_m}$.

The product space $S^{\{t_1,\ldots,t_m\}}$ may be replaced by the Cartesian product $S_{t_1} \times S_{t_2} \times \cdots \times S_{t_m}$, and its product σ-field denoted as $\mathcal{S}_{t_1} \otimes \mathcal{S}_{t_2} \otimes \cdots \otimes \mathcal{S}_{t_m}$ in the case X_t has state space S_t; see Remark 1.1.

For a precise general statement of Kolmogorov's existence (or extension) theorem, we allow the state space of X_t to be $(S_t, \mathcal{S}_t)$, not necessarily the same for all t. But for each t the state space is assumed to be *Polish*, i.e., a topological space metrizable as a complete separable metric space. The consistency conditions on the family of measures $Q_{t_1,\ldots,t_m}$ simply ensure that (a) the distribution of $(X_{t_{i_1}}, \ldots, X_{t_{i_m}})$ is the same as obtained from that of $(X_{t_1}, \ldots, X_{t_m})$ under the permutation $(i_1, \ldots, i_m)$ of $(1, \ldots, m)$, and (b) for any set of distinct indices $t_1, \ldots, t_m, t$, the distribution of $(X_{t_1}, \ldots, X_{t_m})$ is the same as obtained from that of $(X_{t_1}, \ldots, X_{t_m}, X_t)$, by letting X_t be free (i.e., simply requiring $X_t \in S_t$). Formally, the consistency conditions are the following:

(c1) For each $m = 1, 2, \ldots$ and every m-tuple of distinct indices $(t_1, \ldots, t_m)$, and for each permutation π of $(1, 2, \ldots, m)$,

$$Q_{t_{\pi(1)},\ldots,t_{\pi(m)}} = Q_{t_1,\ldots,t_m} \circ \pi^{-1} \tag{1.2}$$

on $(S_{t_{\pi(1)}} \times \cdots \times S_{t_{\pi(m)}}, \mathcal{S}_{t_\pi(1)} \otimes \cdots \otimes \mathcal{S}_{t_{\pi(m)}})$. Here $\pi : S_{t_1} \times \cdots \times S_{t_m} \to S_{t_{\pi(1)}} \times \cdots \times S_{t_{\pi(m)}}$ is given by $\pi(x_{t_1}, \ldots, x_{t_m}) = (x_{t_\pi(1)}, \ldots, x_{t_\pi(m)})$ and π^{-1} is its inverse transformation.

(c2) For each m and every $(m+1)$-tuple of distinct indices $t_1, \ldots, t_m, t$, one has for every $B \in \mathcal{S}_{t_1} \otimes \mathcal{S}_{t_2} \otimes \cdots \otimes \mathcal{S}_{t_m}$,

$$Q_{t_1,\ldots,t_m}(B) = Q_{t_1,\ldots,t_m,t}(B \times S_t). \tag{1.3}$$

Theorem 1.1 *(Kolmogorov's Existence Theorem*[1]*)***.** Assume that for every t the space S_t is Polish with $\mathcal{S}_t$ as its Borel σ-field. Then given a family of distributions $Q_{t_1,\ldots,t_m}$ on $(S_{t_1} \times S_{t_2} \times \cdots \times S_{t_m}, \mathcal{S}_{t_1} \otimes \mathcal{S}_{t_2} \otimes \cdots \otimes \mathcal{S}_{t_m})$ for all $m = 1, 2, \ldots,$ and all m-tuples $(t_1, \ldots, t_m)$ of distinct indices, satisfying the consistency conditions (c1), (c2)

---

[1]For a proof using Caratheodory's extension theorem see BCPT p. 168. Another elegant proof due to Edward Nelson using the Riesz Representation theorem may be found in Nelson (1959), or BCPT, p. 169.

above, there exists a unique probability P on $(\Omega = \prod_{t\in\Lambda} S_t, \mathcal{F} = \otimes_{t\in\Lambda}\mathcal{S}_t)$ such that the coordinate projections X_t $(t \in \Lambda)$ have the prescribed finite-dimensional distributions (1.1).

Remark 1.2. If the index set Λ is linearly ordered, e.g., if it is an interval, or the set $\mathbb{Z}$ of all integers, or the set $\mathbb{Z}_+$ of all non-negative integers, and one is provided with all finite-dimensional distributions $Q_{t_1,\dots,t_m}$ for $t_1 < t_2 < \cdots < t_m$, then one may just check (c2) with $t_m < t$. Condition (c1) is satisfied automatically by defining $Q_{t_{\pi(1)}}, \dots, Q_{t_{\pi(m)}}$ using (1.2).

Example 1 *(Product Probabilities and Independent Families of Random Variables)****.*** Let $(S_t, \mathcal{S}_t, Q_t), t \in \Lambda$, be a family of probability spaces, and $Q_{t_1,\dots,t_m}$ be the product probability measure $Q_{t_1} \times Q_{t_2} \times \cdots \times Q_{t_m}$ on $(S_{t_1} \times S_{t_2} \times \cdots \times S_{t_m}, \mathcal{S}_{t_1} \otimes \mathcal{S}_{t_2} \otimes \cdots \otimes \mathcal{S}_{t_m})$, such that

$$Q_{t_1,\dots,t_m}(B_{t_1} \times B_{t_2} \times \cdots \times B_{t_m}) = Q_{t_1}(B_{t_1})Q_{t_2}(B_{t_2}) \cdots Q_{t_m}(B_{t_m}), \tag{1.4}$$

for $B_{t_i} \in \mathcal{S}_{t_i}, i = 1, \dots, m$. The consistency conditions (c1), (c2) are then trivial to check. The probability measure P on $(\Omega = \prod_{t\in\Lambda} S_t, \mathcal{F} = \otimes_{t\in\Lambda}\mathcal{S}_t)$ with this family of finite-dimensional distributions is called the *product probability* of $(S_t, \mathcal{S}_t, Q_t), t \in \Lambda$. The coordinate projections $X_t, t \in \Lambda$, then form a family of independent random variables. It may be noted that in this case the statistical independence can be exploited to the effect that no topological assumption is needed on the spaces S_t for existence according to a theorem of Tulcea.[2] In the case $S_t = S$ are the same for all t, the family $\{X_t : t \in \Lambda\}$ is referred to as an i.i.d. (independent and identically distributed) S-valued family of random variables.

Often the stochastic process of interest is obtained as a functional of another process already known to satisfy Kolmogorov's, or Tulcea's, conditions, thus avoiding the more elaborate verification of these conditions for its construction. This point is illustrated in several of the following examples.

Example 2 *(General Random Walk and the Simple Random Walk)****.*** Let $S_t = \mathbb{R}^k$ for all t in the index set $\Lambda = \mathbb{Z}_{++} = \{1, 2, 3 \dots\}$. For an arbitrary common probability $Q_t = Q$ on $\mathbb{R}^k$ with Borel sigma-field $\mathcal{S}$, one may construct the product probability measure P on $(\Omega = \mathbb{R}^{\mathbb{Z}_{++}}, \mathcal{F} = \mathcal{S}^{\otimes Z_{++}})$ and the i.i.d. sequence $\{X_t : t \in \mathbb{Z}_{++}\}$ by coordinate projections, as in Example 1. The (general) random walk with step size distribution Q, starting at a point $x \in \mathbb{R}^k$, is then defined by a functional of the process of displacements via

$$S_0 = x, \quad S_n = S_0 + X_1 + \cdots + X_n \quad (n = 1, 2, \dots). \tag{1.5}$$

One may let $S_0 = X_0$, where X_0 is an $\mathbb{R}^k$-valued random variable, with some distribution Q_0, independent of the family of *increment* or *displacement* random

[2] For a general statement and proof of Tulcea's theorem, see Neveu (1965), pp. 162–167.

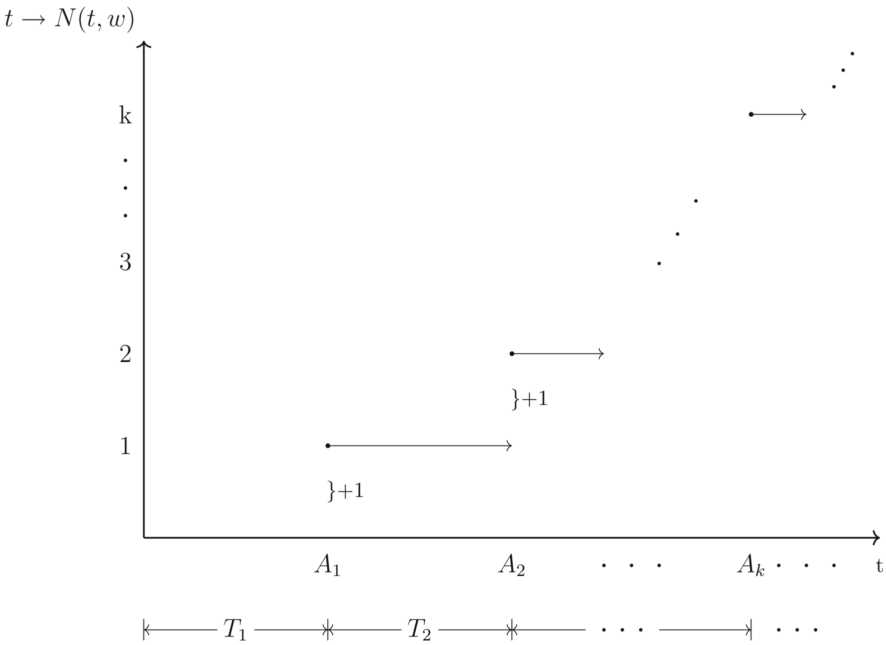


Fig. 1.1 Poisson Process Sample Path

variables $\{X_n : n \in \mathbb{Z}_{++}\}$. For this one may consider the product probability on the enlarged index set $\Lambda = \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ in Example 1, with Q_0 as the distribution of X_0, and $Q_n = Q$ for all $n \in \{1, 2, \ldots\}$. As an important special case one considers $k = 1$, $S = \{-1, +1\}$, $Q(\{+1\}) = p$, $Q(\{-1\}) = 1 - p = q$, say, $0 < p < 1$. The corresponding i.i.d. sequence of displacements (or increments) $\{X_n : n = 1, 2, \ldots\}$ is a sequence of Bernoulli random variables, and the one-dimensional simple random walk is then defined by (1.5), where the distribution Q_0 of X_0 on $\{-1, +1\}$ may be arbitrary, including a degenerate one (i.e., a one-point distribution $Q_0 = \delta_x$). When $p = q = 1/2$, this simple random walk is called a *simple symmetric random walk*. A generalization is the *k-dimensional simple (symmetric) random walk* on $S = \mathbb{Z}^k$ having step size distribution Q, concentrated on $\{\pm e_i : i = 1, 2, \ldots, k\}$, with $Q(\{e_i\}) = Q(\{-e_i\}) = 1/2k$, $(i = 1, 2, \ldots, k)$. Here e_i has $+1$ as its i-th coordinate and zeros as its remaining coordinates.

Another example of a construction of a process as a functional of an i.i.d. sequence is the following; see Figure 1.1.

Example 3 ***(****Poisson Process****).*** Let $T_1, T_2, \ldots$ be an i.i.d. sequence of exponential random variables with parameter $\theta > 0$ (or, mean $1/\theta$). That is, they have the common exponential density

$$\varphi(y) = \theta \exp\{-\theta y\}\mathbf{1}_{[0,\infty)}(y). \tag{1.6}$$

Define the Poisson process $\{N(t) : t \geq 0\}$ with parameter θ, as $N(t) = 0$ for $t < T_1$, and

$$N(t) = n, \quad t \in [T_1 + T_2 + \cdots + T_n, T_1 + T_2 + \cdots + T_{n+1}),\ n = 1, 2, \ldots. \tag{1.7}$$

One may check that the random variable $N(t)$ has the Poisson distribution with mean θt (Exercise 2). The *arrival times* $A_n = T_0 + T_1 + \cdots + T_{n-1}, n \geq 1$, may be viewed as a random walk on the positive half-line. A more in-depth analysis of this important stochastic process having right-continuous sample paths is provided in Chapter 5.

Example 4 ***(****Gaussian Processes/Random Fields****).*** Recall that a k-dimensional Gaussian, or normal, random (column) vector Y with mean (column) vector $m \in \mathbb{R}^k$ and a $k \times k$ nonsingular covariance (or dispersion) matrix Σ has the density

$$\varphi(y) = (2\pi)^{-\frac{k}{2}}(Det\,\Sigma)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y-m)^t \Sigma^{-1}(y-m)\} \quad (y \in \mathbb{R}^k), \tag{1.8}$$

where the superscript t is used to denote matrix transpose. The characteristic function of Y is $\varphi(\xi) \equiv \mathbb{E}\exp\{i\xi^t Y\} = \exp\{im^t\xi - \xi^t\Sigma\xi\}, (\xi \in \mathbb{R}^k)$. More generally, a Gaussian random variable with mean $m \in \mathbb{R}^k$ and a $k \times k$ symmetric non-negative definite covariance matrix Σ is defined as one having the characteristic function $\mathbb{E}\exp\{i\xi^t Y\} = \exp\{im^t\xi - \xi^t\Sigma\xi\}, (\xi \in \mathbb{R}^k)$. If Σ is singular, then Y does not have a density (with respect to Lebesgue measure on $\mathbb{R}^k$). A standard way to construct such a random variable is to first begin with a vector Z of k i.i.d. one-dimensional standard Gaussian random variables $Z_1, Z_2, \ldots, Z_k$, each with mean zero and variance one, $Z = (Z_1, Z_2, \ldots, Z_k)^t$, and define $Y = m + AZ$, where A is a $k \times k$ matrix such that $AA^t = \Sigma$. By the spectral theorem one may actually find a symmetric matrix A such that $AA = \Sigma$. It is simple to check that if $Y = (Y_1, \ldots, Y_k)^t$ is a k-dimensional Gaussian random variable with mean m and covariance matrix Σ, then, for $1 \leq d < k$ and any subset $(i_1, \ldots, i_d)$ of $(1, \ldots, k)$, $(Y_{i_1}, \ldots, Y_{i_d})^t$ is a d-dimensional Gaussian random variable with mean $(m_{i_1}, \ldots, m_{i_d})^t$ and covariance matrix $((\sigma_{ij}))_{i,j=i_1,\ldots,i_d}$, where σ_{ij} is the (i, j)-element of Σ. Consider now the construction of a sequence of Gaussian random variables $\{Y_n : n = 1, 2, \ldots\}$, i.e., a stochastic process indexed by $\Lambda = \mathbb{Z}_+ = \{1, 2, \ldots\}$ such that for any k-tuple of integers $(i_1, \ldots, i_k)$, $1 \leq i_1 < i_2 < \cdots < i_k$, $(Y_{i_1}, \ldots Y_{i_k})^t$ is a k-dimensional Gaussian random variable. Suppose then that $\{m_n : n = 1, 2, \ldots\}$ is a sequence of real numbers and, for each pair $(i.j)$, σ_{ij} are real numbers such that $((\sigma_{ij}))_{i,j=i_1,\ldots,i_k}$, is a $k \times k$ symmetric non-negative definite matrix. It follows from the preceding paragraph that Kolmogorov's consistency condition for such a construction of a sequence of Gaussian random variables $\{Y_n : n = 1, 2, \ldots\}$ with means $\mathbb{E}(Y_n) = m_n$, and covariances $\sigma_{ij} = \mathbb{E}(Y_i - m_i)(Y_j - m_j)$, $i, j \geq 1$, is satisfied if and only if the symmetric non-negative definiteness holds for all k-tuples $(i_1, \ldots, i_k)$, $1 \leq i_1 < i_2 < \cdots < i_k$, for all $k > 1$. One general method to construct such a Gaussian process is to consider an arbitrary (infinite dimensional) matrix $A = ((a_{ij}))_{i,j \in \mathbb{Z}_{++}}$, such that each row belongs to $\ell^2(\mathbb{Z}_{++})$, i.e., $\sum_{1 \leq j < \infty} a_{ij}^2 < \infty (\forall i = 1, 2, \ldots)$. Let Z be a sequence of independent standard Gaussian random variables, $Z = (Z_1, Z_2, \ldots, Z_n, \ldots)^t$, and $m = (m_1, \ldots, m_n, \ldots)^t$. Then

$$Y = m + AZ; \quad Y = (Y_1, Y_2, \ldots, Y_n, \ldots)^t, Y_n = m_n + \sum_{1 \leq j < \infty} a_{nj} Z_j, \tag{1.9}$$

is a Gaussian process (sequence) with mean m and covariance matrix $\Sigma = AA^t = ((\sigma_{ij}))_{1 \leq i,j < \infty}$.

An important special case occurs in which the means $m_n = m_1, n \geq 1$, are constant and $\sigma_{i,j}$ is a function of $i - j$, say $\sigma(i - j)$, $1 \leq j \leq i$. For in this case one may show (Exercise 3) that the distribution of the process $Y = (Y_1, Y_2, \ldots)$ is *translation-invariant* in the following sense.

Definition 1.2. A stochastic process $Y = (Y_1, Y_2, \ldots)$ is said to be a *stationary process* if its distribution is invariant under translations of the form $(Y_{1+h}, Y_{2+h}, \ldots)$ for any positive integer h.

Finally, to construct Gaussian processes with values in a multidimensional state space S one may simply enlarge the index set. For example, if $S = \mathbb{R}^3$ and $\Lambda = [0, \infty)$, then one may change the index t to $(t(1), t(2), t(3))$ to list the three coordinates of the state at time t and prescribe the joint distributions accordingly.

***Example 5** (Brownian Motion).* One-dimensional standard Brownian motion starting at $B_0 = 0 \in \mathbb{R}$ is a continuous parameter Gaussian process $B_t, t \geq 0$, defined on a probability space $(\Omega, \mathcal{F}, P)$, having independent mean zero Gaussian increments $B_t - B_s$, $0 \leq s < t$ over disjoint intervals, with variance $t - s$ and, most importantly, continuous sample paths $t \to B_t$ with probability one; see Figure 1.2. The process is referred to as a continuous parameter Gaussian process to convey that the finite-dimensional distribution of $(B_{t_1}, \ldots, B_{t_m})$, $0 \leq t_1 < t_2 < \cdots < t_m, m \geq 1$, is each a Gaussian distribution. This definition requires a construction and proof of existence, of which several will appear in this text. With that proviso, the process $X_t = x + \mu t + \sigma B_t, t \geq 0$, defines one-dimensional Brownian motion starting at $x \in \mathbb{R}$, drift parameter μ and diffusion coefficient σ^2. More generally, the k-dimensional standard Brownian motion starting at $0 \in \mathbb{R}^k$ is the random vector $(B_t^{(1)}, \ldots, B_t^{(k)})$, $t \geq 0$, for which the components are independent one-dimensional standard Brownian motions. The process $X_t = x + \mu t + \Sigma B_t, t \geq 0$, then defines the k-dimensional Brownian motion starting at $x \in \mathbb{R}^k$, with k-dimensional drift vector μ and (possibly singular) $k \times k$ diffusion matrix $\Sigma^t \Sigma$.

Remark 1.3. On uncountable index sets such as $[0, \infty)$ or $[0, \infty)^d$, $(d > 1)$, the Kolmogorov sigma-field, i.e., the product sigma-field $\mathcal{S}^{\otimes \Lambda}$, is too small to include sets such as the set of continuous paths or the set of right-continuous

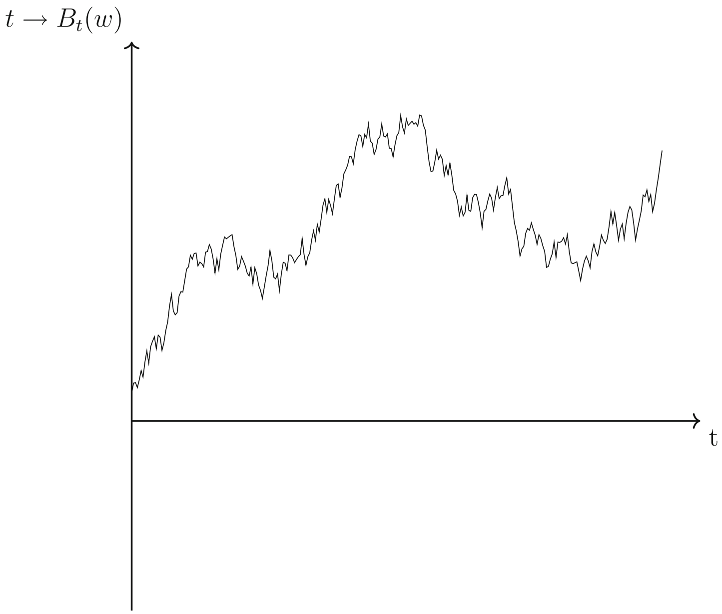


Fig. 1.2 Standard Brownian Motion

paths, and one may seek more regular versions of the process than the coordinate process on S. Indeed, the product sigma-field $\mathcal{S}^{\otimes \Lambda}$ only contains events which are (measurably) determined by countably many coordinates. So this would exclude measurability of the set of paths required to be continuous at a given point. The Kolmogorov–Chentsov Theorem (See Chapter 6) gives an important criterion based on smoothness of moments which yields stochastic processes with continuous sample paths. Similarly a criterion due to Dynkin[3] may be used to ensure right-continuity of sample paths. An alternative elegant route is via Doob's submartingale convergence theorem (See Chapter 13). One may view these arguments as essentially establishing that the relevant sets (of continuous sample paths, or right-continuous sample paths) have outer measure one.[4]

Remark 1.4 **(***Construction of Brownian Motion by Weak Convergence***).** There is another route that may sometimes be used to construct, or prove the existence of, a stochastic process by the theory of weak convergence on function spaces. An important example is the construction of Brownian motion $B_t, 0 \leq t < \infty$, as a limit (under weak convergence of probability measures on $C[0, \infty)$) of the simple symmetric random walk of Example 2 scaled as $B^{(n)}_{\frac{k}{n}} = S_k/\sqrt{n}(k = 0, 1, 2, \ldots)$,

[3]See Gikhman and Skorokhod (1969), pp. 159–169.

[4]See Caratheodory's extension theorem in BCPT p. 226.

and $B_t^{(n)}$ linearly interpolated for $t \in [k/n, (k+1)/n] (k = 0, 1, \dots)$; see Figure 17.1. This so-called functional central limit theorem, or FCLT, is discussed in detail in Chapter 17. The limiting process inherits the property of independent increments from that of the simple random walk, and the FCLT shows that the increments are Gaussian. One may, alternatively, construct such a process using Kolmogorov's existence theorem as in Example 2, together with the regularization due to Kolmogorov and Chentsov as mentioned earlier in Remark 1.3.

Kolmogorov's, or Tulcea's, theorem may be used to construct another very important class of stochastic processes known as *Markov processes*. We restrict here to the case of the discrete parameter, i.e., $\Lambda = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ in the following example.

***Example 6** (Discrete Parameter Markov Processes).* We begin this example with a definition.

Definition 1.3. On a probability space $(\Omega, \mathcal{F}, P)$, a stochastic process $\{X_n : n = 0, 1, 2, \dots\}$ having a state space $(S, \mathcal{S})$ is Markov if the conditional distribution of X_n, given $\sigma\{X_0, \dots, X_{n-1}\}$, is a function of X_{n-1}, alone.

Denote the conditional distributions on the event $[X_{n-1} = x]$ in Definition 1.3, referred to as (one-step) *transition probabilities*, by $p_n(x, dy)$. That is, for any $n \geq 1$, $B \in \mathcal{S}$, $x \in S$, $P(X_n \in B|\sigma\{X_0, \dots, X_{n-1}\}) = p_n(x, B)$ on the event $[X_{n-1} = x]$. Here one has (i) $x \to p_n(x, B)$ is measurable on $(S, \mathcal{S})$ into $(\mathbb{R}, \mathcal{B})$ for all $B \in \mathcal{S}$, and(ii) $B \to p_n(x, B)$ is a probability measure on $(S, \mathcal{S})$ for all $x \in \mathcal{S}$.

Unless otherwise stated we will assume that the transition probabilities $p_n(x, dy)$ are the same for all n, and simply denoted $p(x, dy)$. In this case the Markov process (or its transition probabilities) is said to be *time-homogeneous*. The transition probabilities are often said to be *stationary* in this case too, but this is not to be confused with a stationary process introduced in Definition 1.2, see Exercise 4.

Given the distribution of X_0, say $\mu(dx)$, and a transition probability $p(x, dy)$, the finite-dimensional probabilities of the process are given by the following successive steps:

STEP 1

$$
\begin{aligned}
Q_{0,1,\dots,n}(B_0 \times B_1 \times \cdots \times B_n) &\equiv P(X_0 \in B_0, X_1 \in B_1, \dots, X_n \in B_n) \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1) \cdots \mathbf{1}_{B_{n-1}}(X_{n-1})\mathbf{1}_{B_n}(X_n)|\sigma\{X_0, \dots, X_{n-1}\}]] \\
&= \mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1) \cdots \mathbf{1}_{B_{n-1}}(X_{n-1})g_1(X_{n-1})],
\end{aligned}
$$

where $g_1(X_{n-1}) = P(X_n \in B_n|\sigma\{X_0, \dots, X_{n-1}\}) = P(X_n \in B_n|\sigma\{X_{n-1}\})$. That is,

$$
g_1(x) = \int_{B_n} p(x, dy) = p(x, B_n). \tag{1.10}
$$

STEP 2

$$\begin{aligned}
&Q_{0,1,\ldots,n}(B_0 \times B_1 \times \cdots \times B_n)\\
&\quad = \mathbb{E}[\mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1)\cdots\mathbf{1}_{B_{n-1}}(X_{n-1})g_1(X_{n-1})|\sigma\{X_0,\ldots,X_{n-2}\}]]\\
&\quad = \mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1)\cdots\mathbf{1}_{B_{n-2}}(X_{n-2})\mathbb{E}[\mathbf{1}_{B_{n-1}}(X_{n-1})g_1(X_{n-1})|\sigma\{X_{n-2}\}]]\\
&\quad = \mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1)\cdots\mathbf{1}_{B_{n-2}}(X_{n-2})g_2(X_{n-2})],
\end{aligned}$$

where

$$g_2(x) = \int_{B_{n-1}} g_1(y)p(x,dy). \tag{1.11}$$

Proceeding in this manner, one obtains

STEP $n-1$

$$Q_{0,1,\ldots,n}(B_0 \times B_1 \times \cdots \times B_n) = \mathbb{E}[\mathbf{1}_{B_0}(X_0)\mathbf{1}_{B_1}(X_1)g_{n-1}(X_1)],$$

where

$$g_{n-1}(y) = \int_{B_2} g_{n-2}(y)p(x,dy)). \tag{1.12}$$

Finally,

STEP n

$$Q_{0,1,\ldots,n}(B_0 \times B_1 \times \cdots \times B_n) = \mathbb{E}[\mathbf{1}_{B_0}(X_0)g_n(X_0)] = \int_{B_0} g_n(y)\mu(dy), \tag{1.13}$$

where

$$g_n(x) = \int_{B_1} g_{n-1}(y)p(x,dy). \tag{1.14}$$

To summarize, one has the iterative computation $g_1(x) = p(x, B_n)$,

$$\begin{aligned}
&g_2(x) = \int_{B_{n-1}} g_1(y)p(x,dy), \ldots, g_{n-1}(x) = \int_{B_2} g_{n-2}(y)p(x,dy),\\
&g_n(x) = \int_{B_1} g_{n-1}(y)p(x,dy),
\end{aligned}$$

$$Q_{0,1,\ldots,n}(B_0 \times B_1 \times \cdots \times B_n) = \int_{B_0} g_n(y)\mu(dy), \quad (\forall B_0, B_1, \ldots, B_n \in \mathcal{S}), n \geq 1. \tag{1.15}$$

Conversely, to show that a Markov process exists by specification of an arbitrarily given transition probability $p(x, dy)$ and an arbitrary initial distribution $\mu(dx)$, one checks that the finite-dimensional distributions $Q_{0,1,\dots,n}(B_0, B_1, \dots, B_n)$, as computed in (1.15) entirely in terms of $p(x, dy)$ and $\mu(dx)$, satisfy the Kolmogorov consistency conditions (c1), (c2) above. For this it is enough to prove that

$$Q_{0,1,\dots,n}(B_0 \times B_1 \times \cdots \times B_n) = Q_{0,1,\dots,n,n+1}(B_0 \times B_1 \times \cdots \times B_n \times S), \qquad (1.16)$$

for all $B_0, B_1, \dots, B_n \in \mathcal{S}$. In this case following (1.15), the iterative computation of $Q_{0,1,\dots,n,n+1}(B_0 \times B_1 \times \cdots \times B_n \times S)$ yields, denoting the successive iterations as $h_1, h_2, \dots, h_n, h_{n+1}$,

$$h_1(x) = p(x, S) \equiv 1, h_2(x) = \int_{B_n} h_1(y)p(x, dy) = p(x, B_n) = g_1(x),$$

$$h_3(x) = \int_{B_{n-1}} h_2(y)p(x, dy) = \int_{B_{n-1}} g_1(y)p(x, dy) = g_2(x), \dots,$$

$$h_n(x) = \int_{B_2} h_{n-1}(y)p(x, dy) = \int_{B_2} g_{n-2}(y)p(x, dy) = g_{n-1}(x),$$

$$h_{n+1}(x) = \int_{B_1} h_n(y)p(x, dy) = \int_{B_1} g_{n-1}(y)p(x, dy) = g_n(x),$$

and finally,

$$\begin{aligned} &Q_{0,1,\dots,n+1}(B_0 \times B_1 \times \cdots \times B_n \times S) \\ &= \int_{B_0} h_{n+1}(y)\mu(dy) = \int_{B_0} g_n(y)\mu(dy) \\ &= Q_{0,1,\dots,n}(B_0 \times B_1 \times \cdots \times B_n) \end{aligned} \qquad (1.17)$$

establishes consistency (see Remark 1.2). Hence, by Tulcea's theorem, without any topological assumption on the product space, $(\Omega = S^{\mathbb{Z}_+}, \mathcal{F}^{\otimes \mathbb{Z}_+})$ there is a unique probability P under which the sequence of coordinate projections $\{X_n : n \in Z_+\}$ is a Markov process with transition probability $p(x, dy)$ and initial distribution $\mu(dx)$.

A more enlightening way of stating the (homogeneous) Markov property of a sequence $\{X_n : n = 0, 1, 2, \dots\}$ is that the (conditional) distribution of the future of the process, given its past and present, depends only on the present state. To state this precisely, let P_x denote the distribution of $\{X_n : n = 0, 1, 2, \dots\}$, given $X_0 = x$, i.e., the distribution μ of X_0 in the construction (1.10)–(1.14) is the Dirac measure δ_x. Let $X_{n+} := \{X_n, X_{n+1}, X_{n+2}, \dots\}$, called the after$-n$ process. Then the conditional distribution of X_{n+} given $\sigma\{X_0, X_1, \dots, X_n\}$ (i.e., given $\sigma\{X_0, X_1, \dots, X_n\}$) is P_{X_0}. Here $P_{X_0} = P_x$ on the set $[X_0 = x]$ (Exercise 1).

Remark 1.5. As a stochastic process a Markov process is completely determined by its transition probabilities and its initial distribution. However, given the transition probabilities, it is not uncommon to also view a Markov process as a family of stochastic processes indexed by its possible initial distributions.

Example 7 ***(****Continuous Parameter Markov Process****).*** The extension of the definition (1.3) is as follows.

Definition 1.4. On a probability space $(\Omega, \mathcal{F}, P)$, a stochastic process $\{X_t : t \geq 0\}$ having a state space $(S, \mathcal{S})$ is Markov if for each $0 \leq s < t$, the conditional distribution of X_t, given $\sigma\{X_u, u \leq s\}$, is a function of X_s, alone.

Now the conditional distributions of X_t on the event $[X_s = x]$ in Definition 1.4, again referred to as *transition probabilities*, are denoted by $p(s,t;x,dy)$. That is, for $0 \leq s < t$, $B \in \mathcal{S}$, $x \in S$, $P(X_t \in B|\sigma\{X_u, u \leq s\}) = p(s,t;x,dy)$ on the event $[X_s = x] \in \mathcal{F}$. Here one has (i) $x \to p(s,t;x,B)$ is measurable on $(S,\mathcal{S})$ into $(\mathbb{R},\mathcal{B})$ for all $B \in \mathcal{S}$, and (ii) $B \to p(s,t;x,B)$ is a probability measure on $(S,\mathcal{S})$ for all $x \in \mathcal{S}$. The distribution $\pi(dx)$ of X_0 is referred to as the *initial distribution*. The notation in the case of *homogeneous* (or *stationary*) transition probabilities, i.e., $p(s,t;x,B) \equiv p(0,t-s;x,B)$ is a function of $t-s$, is abbreviated as $p(t-s,x,B)$ for $0 \leq s < t$. Important simple examples are one-dimensional standard Brownian motion for which $p(t;x,dy) = (2\pi t)^{-\frac{1}{2}}e^{-\frac{1}{2t}(y-x)^2}dy$, $t \geq 0, x, y \in \mathbb{R}$. and the Poisson process with intensity $\lambda > 0$, for which $p(t;i,\{j\}) = \frac{(\lambda t)^{j-i}}{(j-i)!}e^{-\lambda t}$, $t \geq 0, 0 \leq i \leq j, i, j \in \mathbb{Z}_+$. Each of these is a stochastic process with stationary, independent increments, from which the Markov property with homogeneous transition probabilities follows since it is quite clear that the conditional distribution of the state at time t, given its past and present states up to time $s < t$, requires only the distribution at time s to compute. Specifically, one simply adds a displacement over time s to t, independent of its state at time s.

While the above examples are tied to additive evolutions, another rich class of examples arises in connection with multiplicative evolutions. The following is a fundamentally important class of such processes.

Example 8 ***(****Discrete Parameter Bienaymé–Galton–Watson Branching Process****).*** While the discrete parameter Bienaymé–Galton–Watson processes are included within the broader framework of Markov processes, as with the case of random walks, they also possess a characteristic structure that gives them a special role in stochastic analysis and modeling. We consider a discrete parameter stochastic process $Y_0, Y_1, \ldots$ with state space $\mathbb{Z}_+$ defined as follows; also see Figure 1.3. The population size is initially given by a non-negative random variable Y_0. Let $p_j, j = 0, 1, 2, \ldots$ be a probability mass function, i.e., $p_j \geq 0, \sum_{j\geq 0} p_j = 1$. Then the first generation population size is given by $Y_1 = 0$ if $Y_0 = 0$, else $Y_1 = \sum_{j=1}^{Y_0} N_{0,j}$, where $N_{0,1}, N_{0,2}, \ldots,$ is an i.i.d. sequence distributed according to the offspring distribution p_j, $j \geq 0$, independently of Y_0. From here the definition

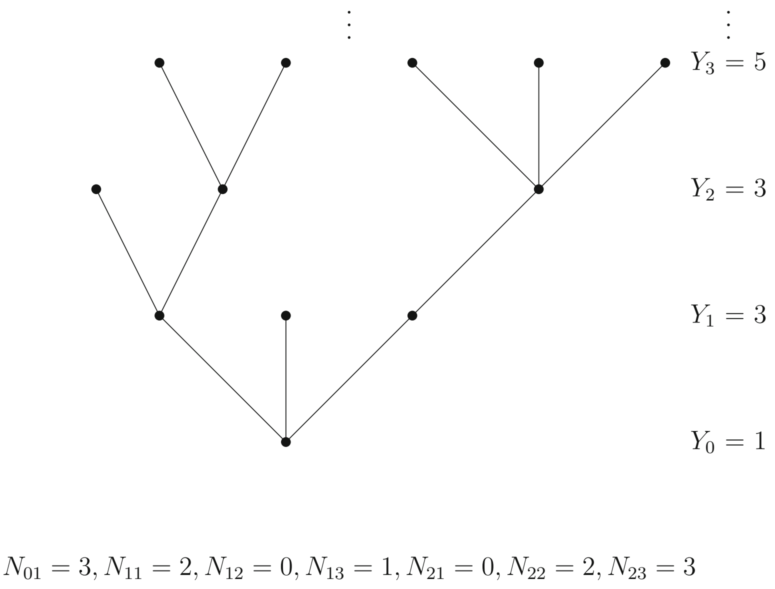


Fig. 1.3 Branching Genealogy

of Y_2 is recursive, with the same rule applied to Y_1 in place of Y_0, and i.i.d. $N_{1,1}, N_{1,2}, \ldots$, distributed as p_j, $j \geq 0$, independently of $N_{0,1}, N_{0,2}, \ldots$, and Y_0. More generally, $Y_{n+1} = 0$ if $Y_n = 0$, else

$$Y_{n+1} = \sum_{j=1}^{Y_n} N_{n,j}, \tag{1.18}$$

where $N_{n,1}, N_{n,2}, \ldots$ is an i.i.d. sequence distributed according to p_j, $j \geq 0$, independently of $N_{m,1}, N_{m,2}, \ldots$, and Y_0, for $0 \leq m \leq n-1$, $n \geq 1$.

As far as first principles go, one may observe that if there is a finite mean number of offspring $\mu = \sum_{j=0}^{\infty} jp_j$, then one has $\mathbb{E}Y_{n+1} = \mu\mathbb{E}Y_n$, $n = 0, 1, 2, \ldots$. In particular, the mean population sizes μ^n, $n \geq 1$, grow or decay according to $\mu > 1$ or $\mu < 1$, respectively. The case $\mu = 1$ is *critical* in the sense that the mean population size remains constant over time. For non-negative integer-valued random variables Y the simple inequality (Exercise 8),

$$P(Y > 0) \leq \mathbb{E}Y \tag{1.19}$$

can be quite useful along the following lines: If $\mu < 1$, then

$$\sum_{n=1}^{\infty} P(Y_n > 0) \leq \sum_{n=1}^{\infty} \mathbb{E}Y_n \leq \sum_{n=1}^{\infty} \mu^n < \infty. \tag{1.20}$$

It follows from the Borel–Cantelli lemma that with probability one $Y_n = 0$ for all but finitely many n. That is, *extinction is certain to occur*, and the total progeny $\sum_{n=0}^{\infty} Y_n$ is almost surely finite; see Exercise 9 for a simple extension of this to non-homogeneous branching. The probability of extinction will be more generally computed in Chapter 9 of the present text for the (homogeneous) Bienaymé–Galton–Watson process.

The previous definition clearly requires a large product space to define the infinite sequences of sequences of i.i.d. random variables. An alternative approach is to exploit the previous construction of a Markov chain with initial distribution of Y_0 and one-step transition probabilities $p(i, \{j\}) = p^{*i}(\{j\})$, $j = i, i+1, \ldots$, where p^{*i} denotes the i-fold convolution of p_j, $j \geq 0$, with $p^{*0}(\{j\}) = \delta_0(\{j\})$, $p_j^{*1} = p_j$, and $p_j^{*(i+1)} = \sum_{k=0}^{j} p_{j-k}^{*i} p_k$ for $j, i = 0, 1, \ldots$. A third approach involves coding the entire genealogy of the branching process as a *random tree graph* as depicted in Figure 1.3. This approach will be developed in later chapters devoted to naturally occurring functions of the branching genealogy.

Queuing theory concerns another class of stochastic processes that naturally arise in a variety of settings as illustrated in the next example.

Example 9 *(Discrete Parameter Queuing Process).* Consider a server that can serve one customer per unit time, but with a random number $X_n, n = 1, 2, \ldots,$ of new customer arrivals during the successive service periods. Assume that $X_1, X_2 \ldots$ are i.i.d. non-negative integer-valued random variables with distribution p_j, $j \geq 0$. Then the number C_n of customers in the system at end of n-th service times $n = 1, 2, \ldots$ may be expressed as

$$C_n = X_n \mathbf{1}[C_{n-1} = 0] + (C_{n-1} - 1 + X_n)\mathbf{1}[C_{n-1} \geq 1], \quad n = 1, 2, \ldots. \tag{1.21}$$

Alternatively, $C_0, C_1, \ldots$ may be more simply defined as a Markov chain with initial distribution of C_0, assumed independent of X_n, $n \geq 1$, and one-step transition probabilities $p(0, j) = p_j$, $j = 0, 1, \ldots$, $p(i, j) = p_{j-i+1}$, $j = i-1, i, i+1 \ldots, i \geq 1$.

Example 10 *(Martingale Processes).* Martingales[5] are defined by an important and essential "statistical dependence"property that can be further exploited in the analysis of Markov and other processes that represent observed models of evolution of random phenomena with time. The defining property requires that predictions of a future state at time t, say, by its conditional expectation on past states up to time $s < t$, be precisely the state at time s. A familiar metaphor is *The prediction of tomorrow's weather, given the past, is today's weather.* In gambling contexts it would be viewed as a *fairness* property, while in analysis and PDEs its counterpart resembles a *harmonic function* property. The previous examples of martingales provide a number of important martingale processes, see Exercise 10, ranging from (i) *mean zero random walks*, (ii) *standard Brownian motion*, (iii) a *Poisson process*

[5]The theory of martingales is primarily due to Doob (1953).

centered on its mean, and (iv) the *critical branching process*. However, the breadth and significance of martingale theory far exceeds this brief list.

Exercises

1. (a) Use Kolmogorov's existence theorem on a Polish state space to prove that the distribution of a process $(X_0, X_1, \ldots, X_n, \ldots)$ is determined by the distributions of $(X_0, X_1, \ldots, X_n)$ for $n = 0, 1, 2, \ldots$ (See Remark 1.3); and use Tulcea's theorem to show that this holds for Markov processes on arbitrary state spaces.
   (b) Prove that the Markov property (Definition 1.3) implies that the conditional distribution of $(X_n, X_{n+1}, \ldots, X_{n+m})$, given $\sigma\{X_0, \ldots, X_n\}$, is the same as the distribution of $(X_0, X_1, \ldots, X_m)$, starting at X_n, that is, the latter distribution is that of $(x, X_1, \ldots, X_m)$ on the set $[X_n = x]$ [*Hint*: Use successive conditioning with respect to $\sigma\{X_0, \ldots, X_{n+m-1}\}$, $\sigma\{X_0, \ldots, X_{n+m-2}\}, \ldots, \sigma\{X_0, \ldots, X_n\}$.]
   (c) Show that the conditional distribution of the after$-n$ process $X_{n+} = (X_n, X_{n+1}, \ldots)$, given $\sigma\{X_0, \ldots, X_n\}$, is P_{X_n} (i.e., P_x on $[X_n = x]$). [*Hint*: Use (b).]
2. Show that for fixed $t > 0$, $N(t)$ in Example 3 has a Poisson distribution with mean θt.
3. Prove that a Gaussian process with constant means and covariances $\sigma_{ij} = \sigma(i - j)$, $i, j \geq 1$, is a stationary process in the sense of Definition 1.2.
4. Suppose that the transition probabilities $p(x, dy)$ and initial distribution $\mu(dx)$ of a Markov process Y have the invariance property that $\int_S p(x, B)\mu(dx) = \mu(B)$ for all $B \in \mathcal{S}$. Show that Y is a stationary process in the sense of Definition 1.2.
5. Verify the discrete parameter Markov property for the coordinate projection process in Example 6 by comparing the two indicated conditional distributions of X_{n+1}.
6. (a) Show that a process having independent increments is a Markov process.
   (b) Verify the continuous parameter Markov property of Brownian motion and the Poisson process.
7. Verify the computation of the mean population size for the Bienaymé–Galton–Watson process in Example 8.
8. Verify the inequality (1.19) for non-negative integer random variables.
9. (*Inhomogeneous Bienaymé–Galton–Watson branching process*) Suppose that the offspring distribution for the Bienaymé–Galton–Watson process is modified to be dependent upon the generation. That is, replace p_j, $j = 0, 1, 2, \ldots$ by $p_j^{(n)}$, $j = 0, 1, 2, \ldots, n = 0, 1, 2, \ldots$, as the offspring distribution for particles

in the n-th generation. Let $\mu_n = \sum_{j=1}^{\infty} j p_j^{(n)}$ denote the mean number of offspring of a single particle in the n-th generation. Show that extinction[6] is certain, i.e., with probability one, $Y_n = 0$ eventually for all n sufficiently large, if $\sum_{n=1}^{\infty} \prod_{j=1}^{n} \mu_j < \infty$.

10. Verify the indicated conditional expectations for the martingale property in each of the following examples.

    (i) For the random walk with mean zero displacements, show that $\mathbb{E}(S_{n+1}|\sigma\{S_j, j \leq n\}) = S_n, n = 0, 1, 2, \ldots$.
    (ii) For the standard Brownian motion, show that $\mathbb{E}(B(t)|\sigma\{B(u), u \leq s\}) = B(s), s \leq t$.
    (iii) For the homogeneous Poisson process with intensity parameter $\theta > 0$, show that $\mathbb{E}(N(t)|\sigma\{N(u), u \leq s\}) = N(s), s \leq t$
    (iv) For the discrete parameter Bienaymé–Galton–Watson process with (critical) mean offspring $\mu = 1$, show that $\mathbb{E}(Y_{n+1}|\sigma\{Y_j, j \leq n\}) = Y_n, n = 0, 1, 2, \ldots$

[6] Under some extra conditions on the offspring distribution, it is shown in Agresti (1975) that the condition $\sum_{n=0}^{\infty}(\mathbb{E}Y_n)^{-1} = \infty$ is both necessary and sufficient for extinction in the non-homogeneous case.

Chapter 2
The Simple Random Walk I: Associated Boundary Value Distributions, Transience, and Recurrence

The simple random walk is the generic example of a discrete time temporal evolution on an integer state space. In this chapter it is defined and simple combinatorics are provided in the computation of its distribution. Two possible characteristic long-time properties, (point) recurrence and transience, are identified in the course of the analysis. Recurrence is a form of "stochastic periodicity" in which the process revisits a state (or arbitrarily small neighborhood) infinitely often, while transience refers to the phenomena in which there are at most finitely many returns.

The simple random walk was introduced in Chapter 1, Example 2. Here we consider some of its properties in detail. Think of a particle moving randomly among the integers according to the following rules. At time $n = 0$ the particle is at the origin. Suppose that X_n denotes the displacement of the particle at the nth step from its position S_{n-1} at time $n-1$, where $\{X_n\}_{n=1}^{\infty}$ is an i.i.d. sequence with $P(X_n = +1) = p$, $P(X_n = -1) = q = 1 - p$ for each $n \geq 1$. The *position process* $\{S_n\}_{n=0}^{\infty}$ is then given by

$$S_n := X_1 + \cdots + X_n, \quad S_0 = 0. \tag{2.1}$$

The stochastic process $\{S_n : n = 0, 1, 2, \ldots\}$ is called the *simple random walk*. The related process $S_n^x = S_n + x$, $n = 0, 1, 2, \ldots$ is called the *simple random walk starting at x*.

The simple random walk is often used by physicists as an approximate model of the fluctuations in the position of a relatively large solute molecule immersed in a pure fluid. According to Einstein's diffusion theory, the solute molecule gets kicked around by the smaller molecules of the fluid whenever it gets within the

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_2

range of molecular interaction with fluid molecules. Displacements in any one direction (say, the vertical direction) due to successive collisions are small and taken to be independent. We shall frequently return to this physical model. One may also think of X_n as a gambler's gain in the nth game of a series of independent and stochastically identical games: a negative gain means a loss. Then $S_0^x = x$ is the gambler's initial capital, and S_n^x is the capital, positive or negative, after n plays of the game.

The first problem is to calculate the distribution of S_n^x at a fixed time n. To calculate the probability of $[S_n^x = y]$, count the number u of $+1$'s in a path from x to y in n steps. Since $n - u$ is then the number of -1's, one must have $u - (n - u) = y - x$, or $u = (n + y - x)/2$. For this, n and $y - x$ must be both even or both odd, and $|y - x| \le n$. Hence

$$P(S_n^x = y) = \begin{cases} \binom{n}{\frac{n+y-x}{2}} p^{(n+y-x)/2} q^{(n-y+x)/2} \\ \qquad \text{if } |y-x| \le n, \text{ and } y-x,\ n \text{ have the same parity} \\ 0 \quad \text{otherwise.} \end{cases} \tag{2.2}$$

Again one may note that the *existence* of the simple random walk $\{S_n\}_{n=0}^\infty$ rests on the existence of a sequence of i.i.d. Bernoulli ± 1-valued random variables $\{X_n\}_{n=1}^\infty$, for a given probability parameter $p \in [0, 1]$, defined on a probability space $(\Omega, \mathcal{F}, P)$; see Examples 1, 2 of Chapter 1.

The simple random walk is an example of a *Markov chain* of great importance. Markov chains on countable state spaces and, more generally, Markov processes on general state spaces, were defined in Chapter 1, along with examples.

As already hinted in Chapter 1, in view of the central limit theorem, a plot of the sample paths of the simple symmetric random walk on the time scale $0, \frac{1}{n}, \frac{2}{n}, \ldots$ with spatial increments $\pm n^{-\frac{1}{2}}$ is, for large n, indistinguishable from a continuous space-time process distributed as *Brownian motion*. This latter process and its connection to random walk will be treated in-depth in forthcoming chapters.

We next consider the asymptotic behavior of the simple random walk as time progresses. Let us view the manner in which a particle escapes from an interval. Let T_y^x denote the first time that the process starting at x reaches y, i.e.,

$$T_y^x := \inf\{n \ge 0 : S_n^x = y\}. \tag{2.3}$$

To avoid trivialities, assume $0 < p < 1$. For integers c and d with $c < d$, denote

$$\varphi(x) := P(T_d^x < T_c^x). \tag{2.4}$$

In other words, $\varphi(x)$ is the probability that the particle starting at x reaches d before it reaches c. Alternatively it is the probability of being at "d" upon reaching the *boundary* $\{c, d\}$ of the (discrete) *interval* denoted $\overline{c, d} := \{c, c+1, \ldots, d-1, d\}$ when started at $x \in \overline{c, d}$.

Proposition 2.1 *(Boundary Distribution Under Nonzero Drift).* Assume $0 < p < 1$, $p \neq q$. Let $c, d \in \mathbb{Z}$, $c < d$. Then

$$P(T_d^x < T_c^x) = \frac{1-(q/p)^{x-c}}{1-(q/p)^{d-c}} \qquad \text{for } c \le x \le d,\ \ p \neq q,$$

$$P(T_c^x < T_d^x) = \frac{1-(p/q)^{d-x}}{1-(p/q)^{d-c}} \qquad \text{for } c \le x \le d,\ \ p \neq q.$$

Proof. Since in one step the particle moves to $x+1$ with probability p, or to $x-1$ with probability q, one has (Exercise 6)

$$\varphi(x) = p\varphi(x+1) + q\varphi(x-1) \tag{2.5}$$

so that

$$\begin{aligned}\varphi(x+1) - \varphi(x) &= \frac{q}{p}[\varphi(x) - \varphi(x-1)], \quad x \in \overline{c+1, d-1}\\ \varphi(c) = 0, &\qquad \varphi(d) = 1.\end{aligned} \tag{2.6}$$

Thus, $\varphi(x)$ is the solution to the *discrete boundary value problem* (2.6). For $p \neq q$, Eq. (2.6) yields

$$\begin{aligned}\varphi(x) &= \sum_{y=c}^{x-1}[\varphi(y+1) - \varphi(y)] = \sum_{y=c}^{x-1}\left(\frac{q}{p}\right)^{y-c}[\varphi(c+1) - \varphi(c)]\\ &= \varphi(c+1)\sum_{y=c}^{x-1}\left(\frac{q}{p}\right)^{y-c} = \varphi(c+1)\frac{1-(q/p)^{x-c}}{1-q/p}.\end{aligned} \tag{2.7}$$

To determine $\varphi(c+1)$ take $x = d$ in (2.7) to get

$$1 = \varphi(d) = \varphi(c+1)\frac{1-(q/p)^{d-c}}{1-q/p}.$$

Then

$$\varphi(c+1) = \frac{1-q/p}{1-(q/p)^{d-c}}$$

so that the first relation holds. The second follows by symmetry or by the same general solution method. ■

Now let

$$\psi(x) := P(T_c^x < T_d^x). \tag{2.8}$$

Note that $\varphi(x) + \psi(x) = 1$, proving that the particle starting in the interior of $\overline{c, d}$ will eventually reach the two-sided boundary (i.e., either c or d) with probability 1. For semi-infinite intervals one obtains the following in the limits as either $d \to \infty$ or $c \to -\infty$.

Corollary 2.2. Assume $0 < p < 1$, $p \neq q$. Let $c < x$ be integers.

$$P(\{S_n^x\}_{n=0}^{\infty} \text{ will ever reach } c) = P(T_c^x < \infty) = \begin{cases} \left(\frac{q}{p}\right)^{x-c}, & \text{if } p > \frac{1}{2} \\ 1, & \text{if } p < \frac{1}{2}, \end{cases}$$

and for integers $x < d$,

$$P(\{S_n^x\}_{n=0}^{\infty} \text{ will ever reach } d) = P(T_d^x < \infty) = \begin{cases} 1, & \text{if } p > \frac{1}{2} \\ \left(\frac{p}{q}\right)^{d-x}, & \text{if } p < \frac{1}{2}. \end{cases}$$

Remark 2.1. As one might guess as the result of the informal remarks made earlier connecting simple random walk and Brownian motion via space-time scaling limits, in the case of Brownian motion the linear second order boundary value difference equations will take the form of differential equations.

Observe that one gets from these calculations the distribution function for the *extremes* $M^x = \sup_n S_n^x$ and $m^x = \inf_n S_n^x$ (Exercise 9). Note also that by the strong law of large numbers,

$$P\left(\frac{S_n^x}{n} = \frac{x + S_n}{n} \to p - q \qquad \text{as } n \to \infty\right) = 1. \tag{2.9}$$

Hence, if $p > q$, then the random walk drifts to $+\infty$ (i.e., $S_n^x \to +\infty$) with probability 1. In particular, the process is certain to reach $d > x$ if $p > q$. Similarly, if $p < q$, then the random walk drifts to $-\infty$ (i.e., $S_n^x \to -\infty$), and starting at $x > c$ the process is certain to reach c if $p < q$. In either case, no matter what the integer y is,

$$P(S_n^x = y \text{ i.o.}) = 0, \qquad \text{if } p \neq q, \tag{2.10}$$

where i.o. is shorthand for "infinitely often." For if $S_n^x = y$ for integers $n_1 < n_2 < \cdots$ through a sequence going to infinity, then

$$\frac{S_{n_k}^x}{n_k} = \frac{y}{n_k} \to 0 \qquad \text{as } n_k \to \infty,$$

the probability of which is zero by (2.9).

Definition 2.1. A state y for which (2.10) holds with $x = y$ is called *transient*. If all states are transient, then the stochastic process is said to be a *transient process*.

Now let us turn to the symmetric case.

Proposition 2.3 ***(Boundary Distribution under Zero Drift).*** Assume $p = q = 1/2$ and let $c \le x \le d$ be integers.

$$P(T_d^x < T_c^x) = \frac{x-c}{d-c}, \quad x \in \overline{c,d}, \ p = q = \frac{1}{2},$$

$$P(T_c^x < T_d^x) = \frac{d-x}{d-c}, \quad x \in \overline{c,d}, \ p = q = \frac{1}{2}.$$

Proof. In the case $p = q = \frac{1}{2}$, according to the boundary value problem (2.6), the graph of $\varphi(x)$ is along the line of constant slope between the points $(c, 0)$ and $(d, 1)$. Thus the first relation follows. The second relation then follows by a symmetry argument or by the same method of calculation. ■

Again we have

$$\varphi(x) + \psi(x) = 1. \tag{2.11}$$

Corollary 2.4. Let $p = q = 1/2$. For integers x, y

$$P(S_n^x = y \text{ i.o.}) = 1. \tag{2.12}$$

Proof. Given any initial position $x > c$,

$$\begin{aligned} &P(\{S_n^x\}_{n=0}^{\infty} \text{ will eventually reach } c) \\ &\quad = P(T_c^x < \infty) \\ &\quad = \lim_{d\to\infty} P(\{S_n^x\}_{n=0}^{\infty} \text{ will reach } c \text{ before it reaches } d) \\ &\quad = \lim_{d\to\infty} \frac{d-x}{d-c} = 1. \end{aligned} \tag{2.13}$$

Similarly, whatever the initial position $x < d$,

$$P(\{S_n^x\}_{n=0}^{\infty} \text{ will eventually reach } d) = P(T_d^x < \infty) = \lim_{c\to-\infty} \frac{x-c}{d-c} = 1. \tag{2.14}$$

Thus, no matter where the particle may be initially, it will eventually reach any given state y with probability 1. After having reached y for the first time, it will move to $y+1$ or to $y-1$. From either of these positions the particle is again bound to reach y with probability 1, and so on. (Exercise 10). ■

Remark 2.2. Although the argument conditioning on the first step X_1 of the random walk used to derive the crucial equation (2.5) is correct (Exercise 6), such arguments are more generally justified by the so-called *strong Markov property* discussed in Chapter 7.

Definition 2.2. A state y for which (2.12) holds with $x = y$ is called *recurrent*. If all states are recurrent, then the stochastic process is called a *recurrent process.*

We conclude this chapter with the following simple observation with regard to the *time η_x of the first return to x* obtained by conditioning on the first step.

Proposition 2.5. Define $\eta_x := \inf\{n \geq 1 : S_n^x = x\}$. Then, $P(\eta_x < \infty) = 2\min(p, q)$.

Proof. Suppose that $p > q$. Conditioning on X_1 and using Corollary 2.2 yields

$$\begin{aligned} P(\eta_x < \infty) &= \mathbb{E}P(\eta_x < \infty | X_1) \\ &= P(T_x^{x+1} < \infty)p + P(T_x^{x-1} < \infty)q \\ &= \frac{q}{p}p + 1q = 2q = 2\min(p, q). \end{aligned}$$

The cases $p > q$ and $p = q$ are handled similarly using Corollary 2.2 and Corollary 2.4, respectively. ■

Corollary 2.6. In the asymmetric case $p \neq q$, the simple random walk is transient.

Proof. This follows directly from Proposition 2.5 and Corollary 2.2. ■

Exercises

1. Let $\{S_n^x\}_{n=0}^{\infty}$ be the simple random walk starting at $x \in \mathbb{Z}$. Show that for $0 = n_0 < n_1 < \cdots < n_m$, $P\left(S_0^x = x, S_{n_1}^x = y_1, \ldots, S_{n_m}^x = y_m\right) = p_{xy_1}^{(n_1)} \cdot p_{y_1y_2}^{(n_2-n_1)} \cdots p_{y_{m-1}y_m}^{(n_m-n_{m-1})}$ where

$$p_{ab}^{(k)} = \begin{cases} \binom{k}{\frac{k+b-a}{2}} p^{\frac{k+b-a}{2}} q^{\frac{k-b+a}{2}}, & |b-a| \leq k, \;\; a, b \text{ have same parity} \\ 0 & \text{else.} \end{cases}$$

2. (*Lazy Random Walk*) The lazy simple random walk is a modification of the simple random walk with displacement probabilities $p, 1-p$ in which the displacements are allowed to be zero with positive probability $\theta \in (0, 1-p)$. That is $\{X_n : n \geq 1\}$ is an i.i.d. sequence with $P(X_n = 1) = p$, $P(X_n = -1) = 1 - p - \theta$, $P(X_n = 0) = \theta, n = 1, 2, \ldots$, and $S_0^x = x \in \mathbb{Z}$, $S_n^x = x + X_1 \cdots + X_n, n \geq 1$. Calculate $P(S_n^x = y)$ for the lazy simple random

walk. (Notice that for $0 < \theta < 1$ one avoids the parity issues intrinsic to simple random walk.)

3. Show that the simple random walk $S^x = \{S_0^x, S_1^x, \dots\}$ on the integers is a Markov chain and identify the the initial distribution and transition probabilities.
4. (*Continuous Time Simple Random Walk*) Suppose that $\{S_n^x : n = 0, 1, 2, \dots\}$ and $\{N_t : t \geq 0\}$ are, respectively, a simple random walk on $\mathbb{Z}$ starting at x with parameter $0 \leq p \leq 1$, and an independent Poisson process on $[0, \infty)$ with parameter $\lambda > 0$. (a) Show that two processes may be defined on a common probability space $(\Omega, \mathcal{F}, P)$ as independent stochastic processes. (b) Define a *continuous time simple random walk* starting at x on $\mathbb{Z}$ by $X_t^x = S_{N_t}^x, t \geq 0$. Calculate $P(X_t^x = y), y \in \mathbb{Z}$. (c) Fix $0 = t_0 < t_1 < \cdots < t_m, y_1, \dots, y_m \in \mathbb{Z}$ and calculate $P(\{\omega \in \Omega : X_{t_j}^x(\omega) = y_j\})$ for $y_j \in \mathbb{Z}, j = 0, \dots, m$. (d) Calculate $\mathbb{E}X_t^x$, $\mathrm{Cov}(X_s^x, X_t^x)$.
5. Suppose that G is an additive abelian group with a topology that makes $(x, y) \to x + y$ a continuous map from $G \times G$ to G for the product topology on $G \times G$. Such a group is said to be a topological abelian group. (i) Formulate a definition of a random walk on G, with its Borel sigma-field, and express the distribution of S_n in terms of a convolution of the distributions of increments $X_n = S_n - S_{n-1}, n \geq 1$. (ii) Characterize all the random walks on the group $G = \{0, 1\}$, with addition modulo two, as two-state Markov chains. (iii) In the remaining parts (a)–(d) of this exercise, consider a random walk $S_n, n \geq 1, S_0 = 0$ on the additive group $G = \mathbb{R}$ having i.i.d. increments $X_n = S_n - S_{n-1}$.

   (a) Assume X_1 is Gaussian with mean zero and unit variance. Show that 0 is not (pointwise) recurrent, i.e., $P(S_n = 0) = 0 n \geq 1$, while for $\epsilon > 0$, but for any interval containing $(-\epsilon, \epsilon)$ about 0, one has *interval recurrence* $P(S_n \in (-\epsilon, \epsilon) i.o.) = 1$. [*Hint*: Use the Chung–Fuchs recurrence criteria[1]]
   (b) Show that if $\mathbb{E}|X_1| < \infty$, and $\mathbb{E}X_1 = 0$ then 0 is interval-recurrent. [*Hint*: Apply the Taylor expansion[2] and Chung–Fuchs criteria.]
   (c) Consider a random walk on $\mathbb{R}$ having i.i.d. symmetric stable[3] increments with characteristic function $\varphi(\xi) = \mathbb{E}e^{i\xi X_1} = e^{-|\xi|^\alpha}$ with $0 < \alpha < 2$. Show that $P(S_n \in (-\epsilon, \epsilon) i.o.) = 1$ for $1 \leq \alpha \leq 2$. [*Hint*: $\frac{1}{1-\varphi(\xi)} \sim \frac{1}{|\xi|^\alpha}$.]
   (d) Give an example of a random walk on $\mathbb{R}$ for which 0 is interval-recurrent, and for which the increments are symmetrically distributed but whose mean does not exist. [*Hint*: Consider $\alpha = 1$, noting that each X_j has the symmetric Cauchy distribution.]

[1] See BCPT p. 124, Cor. 6.15.

[2] See BCPT p. 125, Lemma 3.

[3] See BCPT p. 129.

6. Show that the probabilities $\varphi(x) = P(T_d^x < T_c^x)$, $x \in \mathbb{Z}$, satisfy the difference equation (2.5), $\varphi(x) = p\varphi(x+1) + q\varphi(x-1)$; $c < x < d$. [*Hint*: The conditional distribution of the sequence $(x + X_1, x + X_1 + X_2, \dots)$, given X_1, is the same as the distribution of a random walk starting at $x + X_1$.]
7. (a) Show that for a finite state Markov chain at least one state must be recurrent.
   (b) Give an example of a Markov chain for which all states are transient.
8. Verify the details for the proof of the corollary to Proposition 2.1
9. Calculate $P(M^x > y)$ and $P(m^x \le y)$ for $p \ne \frac{1}{2}$, where $M^x = \sup_n S_n^x$, $m^x = \inf_n S_n^x$, for the simple asymmetric random walk starting at x.
10. Show that $P(S_n^x = y \text{ i.o.}) = 1$ for the simple symmetric random walk $\{S_n^x\}_{n=0}^\infty$.
11. Suppose that two particles are initially located at points x and y, respectively. At each unit of time a particle is selected at random, both being equally likely to be selected, and is displaced one unit to the right or left with probabilities p and q, respectively. Calculate the probability that the two particles will eventually meet.
12. (*A Gambler's Ruin*) A gambler wins or loses 1 unit with probabilities p and $q = 1 - p$, respectively, at each play of a game. The gambler has an initial capital of x units and the adversary has an initial capital of $d > x$ units. The game is played repeatedly until one of the players is broke.

    (i) Calculate the probability that the gambler will eventually go broke.
    (ii) What is the expected duration of the game in the case $p = q$?

13. (*Range of Random Walk*) Let $\{S_n\}_{n=0}^\infty$ be a simple random walk starting at 0 and define the range R_n in time n by $R_n = \#\{S_0 = 0, S_1, \dots, S_n\}$. R_n represents the number of distinct states[4] visited by the random walk in time 0 to n.

    (i) Show that $\mathbb{E}(R_n/n) \to |p - q|$ as $n \to \infty$. [*Hint*: Write $R_n = 1 + \sum_{k=1}^n \mathbf{1}[S_k \ne S_j, j = 0, \dots, k-1]$ and reduce the problem to the calculation of $\lim_{n\to\infty} P(S_n - S_j \ne 0, j = 0, 1, \dots, n-1)$ in terms of an equivalent probability for the successive partial sums of the i.i.d. displacements expressed from time n backward.]
    (ii) Verify that $R_n/n \to 0$ in probability as $n \to \infty$ for the symmetric case $p = q = \frac{1}{2}$. [*Hint*: Use (i) and the Markov inequality.]

14. Let $\{S_n\}_{n=0}^\infty$ be a simple random walk starting at 0. Show the following.

    (i) If $p \ne \frac{1}{2}$, then $\sum_{n=0}^\infty P(S_n = 0) = (1 - 4pq)^{-1/2} = |p - q|^{-1}$. [*Hint*: Apply the Taylor series generalization of the Binomial theorem to $\sum_{n=0}^\infty z^n P(S_n = 0)$ noting that

$$\binom{2n}{n}(pq)^n = (-1)^n (4pq)^n \binom{-\frac{1}{2}}{n},$$

[4] This exercise treats a very special case of a more elaborate contemporary theory of random walk on graphs initiated by Dvoretsky and Erdos (1951).

where $\binom{x}{n} := x(x-1)\cdots(x-n+1)/n!$ for $x \in \mathbb{R}$, $n = 0, 1, 2, \ldots$].

(ii) Give another proof of the transience of 0 using (ii) for $p \neq \frac{1}{2}$. [*Hint*: Use the Borel–Cantelli Lemma.]

15. (*A Geometric Random Walk*) Let $S_0 > 0$ denote today's price of a certain stock and suppose that the *yields* defined by $Y_n = (S_n - S_{n-1})/S_{n-1}$, $n = 1, 2, \ldots$, on the successive prices $S_1, S_2, \ldots$ are i.i.d. with values in $(-1, \infty)$. That is,

$$S_n = \prod_{j=1}^{n}(1 + Y_j) \cdot S_0, \quad n = 1, 2, \ldots .$$

(i) Show that $S_n \to 0$ a.s. if $\mathbb{E}Y_1 < 0$.

(ii) Show that $S_n \to 0$ a.s. if $\mathbb{E}Y_1 = 0$ and $P(Y_1 > 0) > 0$.

(iii) Give an example with $\mathbb{E}Y_j > 0$ such that $\limsup_n S_n = +\infty$, $\liminf_n S_n = 0$. [*Hint*: Consider $Y_j = e^{Z_j} - 1$ with $Z_1, Z_2, \ldots$ i.i.d. symmetric Bernoulli ± 1-valued.]

Chapter 3
The Simple Random Walk II: First Passage Times

In view of recurrence vs transience phenomena, the time T_y^0 to reach a fixed integer state y starting at, say, the origin, may or may not be a finite random variable. Nevertheless, one may consider the possibly defective distribution of T_y^0. An important stochastic analysis tool, referred to as the reflection principle, is used to make this calculation. With this analysis another important refinement of the recurrence property is identified for symmetric random walk, referred to as null recurrence, showing that while the walker is certain to reach y, the expected time $\mathbb{E}T_y^0 = \infty$. This refinement involves an application of Stirling's asymptotic formula for $n!$, for which a proof is also provided. An extension to random walks on the integers that do not skip integer states to the left is also given.

Consider the random variable $T_y := T_y^0$ representing the first time the simple random walk starting at zero reaches the level (state) y. We will calculate the distribution of T_y by means of an analysis of the sample paths of the simple random walk. Let $F_{N,y} := [T_y = N]$ denote the event that the particle reaches state y for the first time at the Nth step. Then,

$$F_{N,y} = [S_n \neq y \qquad \text{for } n = 0, 1, \dots, N-1, S_N = y]. \tag{3.1}$$

Note that "$S_N = y$" means that there are $(N + y)/2$ plus 1's and $(N - y)/2$ minus 1's among $X_1, X_2, \dots, X_N$; see (2.2). Therefore, we assume that $|y| \leq N$ and $N + y$ is even. Now there are as many paths leading from $(0, 0)$ to (N, y) as there are ways of choosing $(N + y)/2$ plus 1's among $X_1, X_2, \dots, X_N$, namely $\binom{N}{\frac{N+y}{2}}$. Each of these choices has the same probability of occurrence, specifically $p^{(N+y)/2}q^{(N-y)/2}$. Thus,

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_3

$$P(F_{N,y}) = Lp^{(N+y)/2}q^{(N-y)/2}, \tag{3.2}$$

where L is the number of paths from $(0,0)$ to (N, y) that do not touch or cross the level y prior to time N. To calculate L consider the complementary number L' of paths that do reach y prior to time N,

$$L' = \binom{N}{\frac{N+y}{2}} - L. \tag{3.3}$$

First consider the case of $y > 0$. If a path from $(0,0)$ to (N, y) has reached y prior to time N, then either (a) $S_{N-1} = y+1$ or (b) $S_{N-1} = y-1$ and the path from $(0,0)$ to $(N-1, y-1)$ has reached y prior to time $N-1$. The contribution to L' from (a) is $\binom{N-1}{\frac{N+y}{2}}$. We need to calculate the contribution to L' from (b) (Figure 3.1).

Proposition 3.1 *(A Reflection Principle).* Let $y > 0$. The collection of all paths from $(0,0)$ to $(N-1, y-1)$ that touch or cross the level y prior to time $N-1$ is in one-to-one correspondence with the collection of all possible paths from $(0,0)$ to $(N-1, y+1)$.

Proof. Given a path γ from $(0,0)$ to $(N-1, y+1)$, there is a first time τ at which the path reaches level y. Let γ' denote the path which agrees with γ up to time τ but is thereafter the mirror reflection of γ about the level y Then γ' is a path from $(0,0)$ to $(N-1, y-1)$ that touches or crosses the level y prior to time $N-1$. Conversely, a path from $(0,0)$ to $(N-1, y-1)$ that touches or crosses the level y prior to time

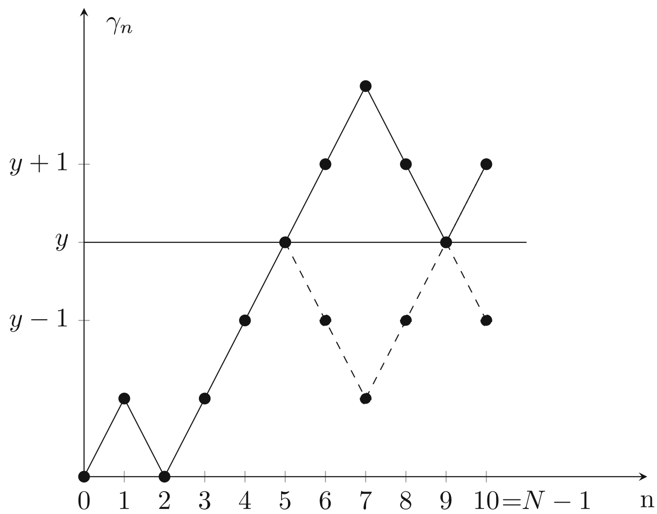


Fig. 3.1 Reflection Principle

$N-1$ may be reflected to get a path from $(0,0)$ to $(N-1, y+1)$. This reflection transformation establishes the one-to-one correspondence. ■

Proposition 3.2.

$$P(T_y = N) = \frac{|y|}{N}\binom{N}{\frac{N+y}{2}} p^{(N+y)/2} q^{(N-y)/2} = \frac{|y|}{N} P(S_N = y) \tag{3.4}$$

for $N = |y|, |y|+2, |y|+4, \ldots$.

Proof. It follows from the reflection principle that the contribution to L' from (b) is $\binom{N-1}{\frac{N+y}{2}}$. Hence

$$L' = 2\binom{N-1}{\frac{N+y}{2}}. \tag{3.5}$$

Therefore, by (3.3), (3.2),

$$\begin{aligned} P(T_y = N) & \\ &= P(F_{N,y}) = \left[\binom{N}{\frac{N+y}{2}} - 2\binom{N-1}{\frac{N+y}{2}}\right] p^{(N+y)/2} q^{(N-y)/2} \\ &= \frac{y}{N}\binom{N}{\frac{N+y}{2}} p^{(N+y)/2} q^{(N-y)/2} \quad \text{for } N \geq y,\ \ y+N \text{ even},\ \ y > 0. \end{aligned} \tag{3.6}$$

To calculate $P(T_y = N)$ for $y < 0$, simply relabel $+1$ as -1 and -1 as $+1$. Using this new code, the desired probability is given by replacing y by $-y$ and interchanging p, q in (3.6), i.e.,

$$P(T_y = N) = -\frac{y}{N}\binom{N}{\frac{N-y}{2}} q^{(N-y)/2} p^{(N+y)/2}.$$

Thus, for all integers $y \neq 0$, one obtains the asserted formula. ■

In the special case $p = q = 1/2$ the first passage distribution takes the form

$$P(T_y = N) = \frac{|y|}{N}\binom{N}{\frac{N+y}{2}} \frac{1}{2^N} \qquad \text{for } \ N = |y|, |y|+2, |y|+4, \ldots. \tag{3.7}$$

Corollary 3.3. For simple random walk with $0 < p < 1$, assuming y and N have the same parity,

$$P(T_y = N|S_N = y) = \frac{|y|}{N}. \tag{3.8}$$

Another application of the reflection principle, sometimes referred to as the *method of images* in this context, is to determine the distribution of a simple random walk in the presence of *boundaries*. Here is an example to illustrate ideas; see Exercise 1 for additional examples. In Chapter 18 a number of additional interesting identities are given as a consequence of reflection, translation, and symmetry transformations.

Corollary 3.4 *(Method of Images).* Fix a positive integer b.

i. Define $\overline{S}_n = S_{n\wedge T_{-b}}, n = 0, 1, \dots,$ where $S_n, n \geq 0$, is simple symmetric random walk starting at 0 and T_{-b} the first passage time to $-b$. Then,

$$P(\overline{S}_{2n} = 0) = P(S_{2n} = 0) - P(S_{2n} = 2b), n \geq 1.$$

ii. Fix a positive integer a. Then,

$$P(\overline{S}_{2n} = 0, T_a \leq 2n) = P(S_{2n} = 2a) - 2P(S_{2n} = 2b + 2a), n \geq 1.$$

iii. Define $\overline{\overline{S}}_n = S_{n\wedge T_{-b}\wedge T_a}$. Then,

$$\begin{aligned} P(\overline{\overline{S}}_{2n} = 0) = P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = -2b) \\ +2\sum_{k=1}^{\infty}\{P(S_{2n} = k(2a + 2b)) - P(S_{2n} = 2a + k(2a + 2b))\}, \end{aligned}$$

and in the case $a = b$,

$$P(\overline{\overline{S}}_{2n} = 0) = \sum_{k=-\infty}^{\infty} \{P(S_{2n} = 4ka) - P(S_{2n} = 2a + 4ka)\}.$$

Proof. It is suggested to sketch sample paths as an aid to following the various reflections. (i) One has $P(\overline{S}_{2n} = 0) = P(S_{2n} = 0, T_{-b} > 2n) = P(S_{2n} = 0) - P(S_{2n} = 0, T_{-b} \leq 2n) = P(S_{2n} = 0) - P(S_{2n} = -2b) = P(S_{2n} = 0) - P(S_{2n} = 2b)$, where the second to last equality is obtained by reflection of the random walk paths about $y = -b$, and the last equality uses symmetry of S_{2n} about 0.

Part (ii) is proven similarly but uses double reflections. Note that existence of double reflections is assured because the boundaries and boundary crossings reflect with the paths after the first single reflection. They are invertible by reversing the order of reflections.

$$
\begin{aligned}
&P(\overline{S}_{2n} = 0, T_a \le 2n) \\
&\quad = P(S_{2n} = 0, T_a \le 2n, T_{-b} > 2n) \\
&\quad = P(S_{2n} = 0, T_a \le 2n) - P(S_{2n} = 0, T_a \le 2n, T_{-b} \le 2n) \\
&\quad = P(S_{2n} = 0, T_a \le 2n) - P(S_{2n} = 0, T_{-b} \le 2n, T_a < T_{-b}) \\
&\qquad -P(S_{2n} = 0, T_a \le 2n, T_{-b} < T_a) \\
&\quad = P(S_{2n} = 2a) - P(S_{2n} = 2b + 2a) \\
&\qquad -P(S_{2n} = -2b - 2a) \\
&\quad = P(S_{2n} = 2a) - 2P(S_{2n} = 2b + 2a),
\end{aligned}
$$

where the first term is by reflection about $y = a$ and the second term, for paths ω with $T_a(\omega) < T_{-b}(\omega), \omega_0 = \omega_{2n} = 0$, is by use of double reflections from $y = a$ followed by a reflection of this path about $y = 2a + b$ at $(T'_{2a+b}, 2a + b)$ to define ω' such that $\omega'_0 = 0, \omega'_{2n} = 2a + 2b$. The third term, for $T_{-b}(\omega) < T_a(\omega)$, is by double reflections about $y = -b$ at $(T_{-b}, -b)$ followed by a reflection of this path about $y = -2b - a$ at $(T'_{-2b-a}, -2b - a)$, to define ω' such that $\omega'_0 = 0, \omega'_{2n} = -2a - 2b$ The final formula follows by symmetry of S_{2n} about 0.

Finally, for (iii) note that

$$
\begin{aligned}
&P(\overline{\overline{S}}_{2n} = 0) = P(S_{2n} = 0, T_{-b} > 2n, T_a > 2n) \\
&\quad = P(S_{2n} = 0, T_a > 2n) - P(S_{2n} = 0, T_a > 2n, T_{-b} \le 2n) \\
&\quad = P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = 0, T_{-b} \le 2n) \\
&\qquad +P(S_{2n} = 0, T_a \le 2n, T_{-b} \le 2n) \\
&\quad = P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = 2b) \\
&\qquad +P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n).
\end{aligned}
\tag{3.9}
$$

The first difference incorporates the $k = 0$ term of the asserted series to be proved. The problem is to calculate $P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n)$. For this first note that, as in part (ii), a path ω such that $\omega_0 = 0, \omega_{2n} = 0, T_{-b}(\omega) \le 2n, T_a(\omega) \le 2n$ reflects across $y = a$ at (T_a, a) to a path that must touch $y = 2a + b$ at $(T_{2a+b}, 2a + b)$. The second reflection yields a path $\omega'_0 = 0, \omega'_{2n} = 2a + 2b$. This doubly reflected path ω' may or may not touch the next level $y = 3a + 2b$ $(k = 1)$. Thus, there is a maximal $k \ge 1$ such that the iterated double reflections yield a path ω' such that (i) If $T_{-b}(\omega) < T_a(\omega)$, then $\omega'_0 = 0, \omega'_{2n} = k(2a + 2b), T_{k(2a+2b)+a}(\omega') > 2n, T_{-b}(\omega') \le 2n$, and (ii) if $T_{-b}(\omega) > T_a(\omega)$, then $\omega'_0 = 0, \omega'_{2n} = k(2a + 2b), T_{k(2a+2b)+a}(\omega') > 2n$. Similarly for double reflections across $y = -b$ and then $y = -2b - a$ to $-2b - 2a$, the new path may or may not touch $y = -3b - 2a$ $(k = -1)$, so that there is a maximal $|k|$. For $k = 0$ (no reflections) the path ω must touch both $y = a$ and $y = -b$. The double reflections define one-to-one and invertible relations by reversing the reflections. Since the double reflection $|k|$-times defines *two* relabeling of a path ω such that $\omega_0 = \omega_{2n} = 0, T_{-b} \le 2n, T_a \le 2n$ defined by maximal $|k| = |k(\omega)|$, in order to make the relation well-defined (function) one must specify direction of reflections. Say the maximal reflection is *upward* for $k \ge 1$ and *downward* for $k \le -1$. We adopt the following rule: For $T_{-b}(\omega) > T_a(\omega)$ use maximal upward double reflections from $y = a$ to define ω' such that $\omega'_0 = 0, \omega'_{2n} = k(2a + 2b), T_{a+k(2a+2b)}(\omega') > 2n, T_a(\omega') < T_{2a+b}(\omega')$. For $T_{-b}(\omega) < T_a(\omega)$ use maximal downward double reflections from $y = -b$ to define ω' such that $\omega'_0 = 0, \omega'_{2n} = k(2a + 2b), T_{-b+k(2a+2b)}(\omega') > 2n(k \le -1), T_{-b}(\omega') < T_{-2b-a}(\omega')$. Note that "maximality" of k is defined by the respective conditions $[S_{2n} = k(2a + 2b), T_{a+k(2a+2b)} > 2n)], (k \ge 1)$, $[S_{2n} = k(2a + 2b), T_{-b+k(2a+2b)} > 2n)], (k \le -1)$, and this sets up one-to-one correspondences with $\omega \in [T_{-b} > T_a]$ and $\omega \in [T_{-b} < T_a]$, respectively, and such that $\omega_0 = \omega_{2n} = 0$, $T_{-b}(\omega) \le 2n$, $T_a(\omega) \le 2n$. By the skip-free property of simple random walk paths, the conditions that $T_a(\omega') < T_{2a+b}(\omega')$, $T_{-b}(\omega') < T_{-2b-a}(\omega')$ are redundant since for each path ω', $x \to T_x(\omega'), T_0(\omega') = 0$, is increasing as a function of $x > 0$ and decreasing for $x < 0$. With this rule one has

$$
\begin{aligned}
&P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n) \\
&\quad = P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n, T_{-b} > T_a) \\
&\qquad + P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n, T_{-b} < T_a) \\
&\quad = \sum_{k=1}^{\infty} P(S_{2n} = k(2a + 2b), T_{a+k(2a+2b)} > 2n) \\
&\qquad + \sum_{k=-\infty}^{-1} P(S_{2n} = k(2a + 2b), T_{-b+k(2a+2b)} > 2n) \\
&\quad = \sum_{k=1}^{\infty} \{P(S_{2n} = k(2a + 2b)) - P(S_{2n} = 2a + k(2a + 2b))\} \\
&\qquad + \sum_{k=-\infty}^{-1} \{P(S_{2n} = -k(2a + 2b)) - P(S_{2n} = -2b + k(2a + 2b))\} \\
&\quad = 2\sum_{k=1}^{\infty} \{P(S_{2n} = k(2a + 2b)) - P(S_{2n} = 2a + k(2a + 2b))\}.
\end{aligned}
$$

Thus, one has using (3.9) incorporating $k = 0$,

$$
\begin{aligned}
&P(\overline{\overline{S}}_{2n} = 0) \\
&\quad = P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = -2b) + P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n) \\
&\quad = P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = -2b) + P(S_{2n} = 0, T_{-b} \le 2n, T_a \le 2n)
\end{aligned}
$$

$$= P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = -2b)$$
$$+2\sum_{k=1}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}$$

using symmetry of S_{2n} with respect to the origin. To complete the proof let us check that in the case $a = b$,

$$P(S_{2n} = 0) - P(S_{2n} = 2a) - P(S_{2n} = -2b)$$
$$+2\sum_{k=1}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}$$
$$= \sum_{k=-\infty}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}.$$

Namely, consider the difference with the desired series. That is,

$$\sum_{k=-\infty}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}$$
$$-2\sum_{k=1}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}$$
$$= P(S_{2n} = 0) - P(S_{2n} = 2a) + \sum_{k=-\infty}^{-1}\{P(S_{2n} = k(2a+2b))$$
$$-P(S_{2n} = 2a + k(2a+2b))\}$$
$$-\sum_{k=1}^{\infty}\{P(S_{2n} = k(2a+2b)) - P(S_{2n} = 2a + k(2a+2b))\}$$
$$= P(S_{2n} = 0) - P(S_{2n} = 2a) + \sum_{k=1}^{\infty}\{P(S_{2n} = 2a + k(2a+2b))$$
$$-P(S_{2n} = 2a - k(2a+2b))\}.$$

Telescoping occurs in the last series if $a = b$. Remarkably,

$$\sum_{k=1}^{\infty}\left\{P(S_{2n} = 2a + k(2a+2b)) - P(S_{2n} = 2a - k(2a+2b))\right\} = -P(S_{2n} = 2b),$$

which completes the proof for this case as well. ■

Remark 3.1. In the proofs of parts (ii) and (iii) of Corollary 3.4 two different types of double reflections are used, namely upward/downward or downward/upward in (ii), and upward/upward or downward/downward in (iii). Also see Exercise 2.

The following proposition records a frequently used asymptotic formula for $n!$ from which one may easily deduce that the *expected time to reach* y is infinite; see Proposition 3.6 below. Stirling's formula is such a prevalent tool for limit theorems of probability that we include a proof here.

Proposition 3.5 *(Stirling's Formula).* The sequence $s_n := \frac{n!}{\sqrt{2\pi n}n^n e^{-n}}, n = 1, 2, \ldots$ is monotonically decreasing and

i $\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}n^n e^{-n}} = 1,$
ii $1 \le \frac{n!}{(2\pi n)^{1/2} n^n e^{-n}} < \frac{e}{(2\pi)^{1/2}}, \quad n = 1, 2, \ldots.$

Proof. To see that the sequence of ratios $\{s_n\}_{n=1}^{\infty}$ is monotonically decreasing, observe that

$$\begin{aligned}
\log \frac{n!}{(2\pi n)^{1/2} n^n e^{-n}} &= \log n! - \frac{1}{2}\log n - n\log n + n - \log(2\pi)^{\frac{1}{2}} \\
&= \left\{\sum_{j=1}^{n} \log j - \frac{1}{2}\log n\right\} - \{n\log n - n\} - \log(2\pi)^{\frac{1}{2}} \\
&= \left\{\sum_{j=2}^{n} \frac{\log(j-1) + \log(j)}{2} - \int_1^n \log x\, dx\right\} + \log\left(\frac{e}{\sqrt{2\pi}}\right), \qquad (3.10)
\end{aligned}$$

where the integral term may be checked by integration by parts. The point is that the term defined by

$$T_n = \sum_{j=2}^{n} \frac{\log(j-1) + \log(j)}{2} \qquad (3.11)$$

provides the inner trapezoidal approximation to the area under the curve $y = \log x$, $1 \le x \le n$, noting the concavity of $x \to \log x$. Thus, in particular,

$$0 \le \int_1^n \log x\, dx - T_n$$

is monotonically increasing and the asserted upper bound on s_n holds. To obtain the asymptotic approximation (and lower bound) one notes by integration by parts that $n! = \Gamma(n+1)$, where $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt = 2\int_0^\infty s^{2x-1}e^{-s^2}ds, x > 0.$

Making a change of variables of the form $s = \sqrt{x} + u$ yields

$$\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_0^\infty e^{x-s^2}(\frac{s}{\sqrt{x}})^{2x-1}ds = 2\int_{-\sqrt{x}}^\infty e^{-2u\sqrt{x}}(1+\frac{u}{\sqrt{x}})^{2x-1}e^{-u^2}du. \tag{3.12}$$

Define

$$g(x,u) := \begin{cases} 0 & \text{if } u \le -\sqrt{x} \\ e^{-2u\sqrt{x}}(1+\frac{u}{\sqrt{x}})^{2x-1} & \text{if } u \ge -\sqrt{x}. \end{cases} \tag{3.13}$$

Then, the Taylor expansion for $\log(1+y) = \sum_{m=1}^\infty (-1)^{m+1}y^m/m$ applied to $\log(1+u/\sqrt{x})$ for large x yields $\log g(x,u) = -u^2 + O(x^{-1/2})$, and therefore $\lim_{x\to\infty} g(x,u) = e^{-u^2}$ as $x \to \infty$. Now check that $g(x,u)$ achieves a maximum value at $u = -\frac{1}{2\sqrt{x}}$ with $g(x,u) \le \mathbb{E}(1-\frac{1}{2x})^{2x-1}$. It now follows from (3.12), using Lebesgue's dominated convergence theorem,

$$\lim_{x\to\infty}\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_{-\infty}^\infty e^{-2u^2}du = 2\sqrt{\pi/2} = \sqrt{2\pi}. \tag{3.14}$$

Letting $x = n$ above and noting that $n! = \Gamma(n+1) = n\Gamma(n)$, one arrives at the desired limit in Proposition 3.5(i). The lower bound in (ii) also follows from the fact that the sequence within curly brackets in (3.10) is decreasing. This completes the proof of the proposition and provides the basic asymptotic formula

$$n! = \sqrt{2\pi n}n^n e^{-n}(1+\delta_n), \tag{3.15}$$

where $\delta_n \to 0$ as $n \to \infty$. ■

Proposition 3.6. For the simple symmetric random walk starting at an integer x, one has for every state $y \ne x$,

$$\mathbb{E}T_y^x = \infty.$$

Proof. Without loss of generality one may take $x = 0$ and apply Stirling's formula to compute the asymptotic form of the terms $NP(T_y = N) = O(N^{-\frac{1}{2}})$ as $N \to \infty$ from which divergence of the series defining $\mathbb{E}T_y$ follows. ■

To conclude this chapter we make note of the surprising fact that the relation between the first passage time distribution and position conveyed by Proposition 3.2 is actually true for random walks[1] on the integers defined by $S_n = Y_1 + Y_2 + \cdots, Y_n, n \ge 1, S_0 = 0$, for $i.i.d.$ integer-valued displacement random variables Y_i

[1]The proof presented here is due to Hofstad van der and Keane (2008) and also applies to symmetrically dependent walks, i.e., to partial sum sequences having exchangeable increments.

such that $P(Y_1 \geq -1) = 1$, i.e., with the property that it is impossible to skip states in moving to the left.[2]

An application to formulae for the distribution of total progeny for a particular critical Bienaymé–Galton–Watson branching process is given in the Example 22.2 of Chapter 22. Kemperman's formula may also be viewed as a *generalized ballot theorem* (see Exercise 4).

Proposition 3.7 *(Kemperman's Formula).* Let $S_0 = 0$, $S_n = Y_1 + \cdots + Y_n, n \geq 1$ be a random walk on the integers starting at zero, and such that it does not skip states to the left, i.e., $P(S_{n+1} - S_n \geq -1) = 1 \forall n \geq 0$. For a positive integer y, let $T_{-y} = \inf\{n \geq 1 : S_n = -y\}$. Then

$$P_0(T_{-y} = N) = \frac{y}{N} P_0(S_N = -y), \quad N = 1, 2, \ldots,$$

or equivalently, letting $T_0 = \inf\{n \geq 1 : S_n = 0\}$ if $S_0 = y > 0$,

$$P_y(T_0 = N) = \frac{y}{N} P_y(S_N = 0), \quad N = 1, 2, \ldots.$$

Proof. In anticipation of the string of calculations to follow let us first note that $\mathbb{E}(Y_1|S_N = -y) = \mathbb{E}(Y_j|S_N = -y) = -\frac{y}{N}$. (This last equality follows by summing over $j = 1, \ldots N$.) Now, using the definitions of S and T,

$$\begin{aligned}
P_0(T_{-y} = N) &= \sum_{x=-1}^{\infty} P_0(T_{-y} = N | Y_1 = x) P_0(Y_1 = x) \\
&= \sum_{x=-1}^{\infty} P_0(T_{-y-x} = N - 1) P_0(Y_1 = x) \\
&= \sum_{x=-1}^{\infty} \frac{y + x}{N - 1} P_0(S_{N-1} = -y - x) P_0(Y_1 = x),
\end{aligned}$$

where the last equality is obtained from an induction hypothesis on the validity of the formula for $N - 1$. It follows by reorganizing the conditioning and using the anticipated conditional expectation above that

$$\begin{aligned}
&P_0(T_{-y} = N) \\
&= \frac{1}{N - 1} \sum_{x=-1}^{\infty} (y + x) P_0(S_{N-1} = -y - x) P_0(Y_1 = x)
\end{aligned}$$

[2]This result is generally attributed to Kemperman (1950).

$$= \frac{1}{N-1}\sum_{x=-1}^{\infty}(y+x)P_0(S_N=-y|Y_1=x)P_0(Y_1=x)$$

$$= \frac{1}{N-1}\{yP_0(S_N=-y)+\sum_{x=-1}^{\infty}xP_0(Y_1=x|S_N=-y)P_0(S_N=-y)\}$$

$$= \frac{1}{N-1}P_0(S_N=-y)\{y+\mathbb{E}(Y_1|S_N=-y)\}$$

$$= \frac{1}{N-1}P_0(S_N=-y)\{y-\frac{y}{N}\}=\frac{y}{N}P_0(S_N=-y). \tag{3.16}$$

For the equivalent statement simply consider the respective events with S_n replaced by S_n+y, $n=0,1\dots$. ■

The property that the walk on the integers does not skip states to the left described in Proposition 3.7 will be referred to as the *left-skip free* property.

Exercises

1. (*Method of Images*) Let $a,b>0$, $0<x<a$. Define a simple random walk with absorbing boundaries at $-b,a$ by $\overline{\overline{S}}_n = S_{n\wedge T_{-b}\wedge T_a}, n\geq 0$, where $S_n, n\geq 0$, is the (unrestricted) simple symmetric random walk starting at 0. Extend Corollary 3.4 by showing
   (i) $P(S_{2n}=x, T_{-b}>2n)=P(S_{2n}=x)-P(S_{2n}=2b-x), n\geq 1$.
   (ii) $P(S_{2n}=x, T_{-b}>2n, T_a\leq 2n)=P(S_{2n}=2a+x)-2P(S_{2n}=2b+2a-x), n\geq 1$.
   (iii) In the case $a=b$, $P(S_{2n}=x, T_{-a}>2n, T_a>2n)=\sum_{k=-\infty}^{\infty}\{P(S_n=x+4ka)-P(S_n=2a-x+4ka)\}, n\geq 1$.
2. Show that for $a\neq b$, the formula $P(S_n=0, T_{-b}>n, T_a>n)=P(S_{2n}=0)-P(S_{2n}=2a)-P(S_{2n}=2b)+2\sum_{k=1}^{\infty}\{P(S_n=k(2a+2b)-P(S_n=2a+k(2a+2b))\}, n\geq 1$, fails to reduce to the corresponding doubly infinite series. [*Hint*: The probability is 2^{-2n} in the case $a=2, b=1$.]
3. (*Bertrand's classic ballot theorem*) Candidates A and B have probabilities p and $q=1-p$ $(0<p<1)$ of winning any particular vote, respectively, independently among voters. If A scores a votes and B scores b votes, $a>b$, then show that $\frac{a-b}{a+b}$ is the (conditional) probability that A will maintain a lead throughout the process of sequentially counting all $a+b$ votes cast. [*Hint*: Reformulate as a hitting time problem for symmetric simple random walk and apply the reflection principle, independently of p.]

4. (*Generalized ballot theorem*[3]) Let $\{S_n = Y_1 + \cdots + Y_n : n \geq 1\}$, $S_0 = 0$, be a random walk on the integers having non-negative integer valued displacement random variables. Show that for a nonnegative integer $0 \leq y < N$, $P_0(S_m < m, m = 1, 2, \ldots, N | S_N = y) = \frac{N-y}{N}$. [*Hint*: Apply the Kemperman formula for hitting zero starting at $N - y > 0$ to the left-skip free random walk $\tilde{S_m} = \tilde{Y}_1 + \cdots + \tilde{Y}_m, m \geq 1$, where $\tilde{S}_m = N - y - m + S_N - S_{N-m}, m = 1, \ldots, N$, $\tilde{S}_0 = N - y > 0$, and $\tilde{S}_m - \tilde{S}_{m-1} = Y_{N-m+1} - 1 \geq -1$.]
5. (*A Reflection Property*) For the simple symmetric random walk $\{S_n\}_{n=0}^{\infty}$ starting at 0 use the fact that, for a positive integer $y > 0$, $N = 1, 2, \ldots$.

$$P\left(\max_{n \leq N} S_n \geq y\right) = 2P(S_N \geq y) - P(S_N = y),$$

to prove the algebraic identity

$$\sum_{y \leq m \leq N}^{*} \frac{y}{m} \binom{m}{\frac{m+y}{2}} 2^{-m} = 2 \sum_{x \geq y}^{*} \binom{N}{\frac{N+x}{2}} 2^{-N} - \binom{N}{\frac{N+y}{2}} 2^{-N},$$

where the respective asterisks in the sums indicate that the summation is over values of m such that $m + y$ is even, and x such that $N + x$ is even, respectively.
6. Let $\{S_n\}_{n=0}^{\infty}$ be the simple symmetric random walk starting at 0 and let

$$M_N = \max\{S_n : n = 0, 1, 2, \ldots, N\}; \quad m_N = \min\{S_n : n = 0, 1, 2, \ldots, N\}.$$

(i) Calculate the distribution of M_N.
(ii) Calculate the distribution of m_N.
(iii) Show $P(M_N \geq z, S_N = y) = \begin{cases} P(S_N = y), & y \geq z \\ P(S_N = 2z - y) & y \leq z. \end{cases}$

7. Suppose that the points of the state space $S = \mathbb{Z}$ are painted blue with probability ρ or green with probability $1 - \rho$, $0 \leq \rho \leq 1$, independently of each other and of a simple random walk $\{S_n\}_{n=0}^{\infty}$ starting at 0[4] Let B denote the random set of states (integer sites) colored blue and let $N_n(\rho)$ denote the amount of time (*occupation time*) that the random walk spends in the set B prior to time n, i.e.,

$$N_n(\rho) = \sum_{k=0}^{n} \mathbf{1}_B(S_k).$$

[3] Hofstad van der and Keane (2008) attribute this formulation to Konstantopoulos (1995).

[4] See Spitzer (1976) for this and related problems in the so-called potential theory of random walk.

Show that regardless of p, q, $\mathbb{E}N_n(\rho) = (n+1)\rho$. [*Hint*: $\mathbb{E}\mathbf{1}_B(S_k) = \mathbb{E}\{\mathbb{E}[\mathbf{1}_B(S_k) \mid S_k]\}$.]

8. Establish the following analytic identities as consequences of the probabilistic results of this chapter.

(i) $$\sum_{\substack{N \geq |y|,\\ N+y \text{ even}}} \left\{ \frac{|y|}{N} \binom{N}{\frac{N+y}{2}} 2^{-N} \right\} = 1 \qquad \text{for all } y \neq 0.$$

(ii) For $p > q$,

$$\sum_{\substack{N \geq |y|,\\ N+y \text{ even}}} \left\{ \frac{|y|}{N} \binom{N}{\frac{N+1}{2}} p^{(N+y)/2} q^{(N-y)/2} \right\} = \begin{cases} 1 & \text{for } y > 0 \\ \left(\frac{p}{q}\right)^y & \text{for } y < 0. \end{cases}$$

Chapter 4
Multidimensional Random Walk

The simple symmetric random walk on the integers readily extends to that of a simple symmetric random walk on the k-dimensional integer lattice, in which at each step the random walk moves with equal probability to one of its $2k$ neighboring states on the lattice. A celebrated theorem of Pólya provides a role for dimension k in distinguishing between recurrent and transient properties of the random walk. Namely, it is shown by combinatorial methods that the simple symmetric random walk on the k-dimensional integer lattice $\mathbb{Z}^k$ is recurrent for $k = 1, 2$ and transient for $k \geq 3$.

The *k-dimensional unrestricted simple symmetric random walk* describes the motion of a particle moving randomly on the integer lattice $\mathbb{Z}^k$ according to the following rules. Starting at a site $\mathbf{x} = (x_1, \ldots, x_k)$ with integer coordinates, the particle moves to a neighboring site in one of the $2k$ coordinate directions randomly selected with probability $1/2k$, and so on, independently of previous displacements. The *displacement* at the nth step is a random variable $\mathbf{X}_n$ whose possible values are vectors of the form $\pm\mathbf{e}_i$, $i = 1, \ldots, k$, where the jth component of $\mathbf{e}_i$ is 1 for $j = i$ and 0 otherwise. $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are i.i.d. with

$$P(\mathbf{X}_n = \mathbf{e}_i) = P(\mathbf{X}_n = -\mathbf{e}_i) = 1/2k \qquad \text{for } i = 1, \ldots, k. \tag{4.1}$$

The corresponding *position process* is defined by

$$\mathbf{S}_0^{\mathbf{x}} = \mathbf{x}, \quad \mathbf{S}_n^{\mathbf{x}} = \mathbf{x} + \mathbf{X}_1 + \cdots + \mathbf{X}_n, \quad n \geq 1. \tag{4.2}$$

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_4

The case $k = 1$ is already treated in the preceding chapters with $p = q = \frac{1}{2}$. In particular, for $k = 1$ we know that the simple symmetric random walk is *recurrent.*

Consider the coordinates of $\mathbf{X}_n = (X_n^1, \ldots, X_n^k)$. Although X_n^i and X_n^j are *not independent*, notice that they are *uncorrelated* for $i \neq j$. Likewise, the coordinates of the position vector $\mathbf{S}_n^{\mathbf{x}} = (S_n^{\mathbf{x},1}, \ldots, S_n^{\mathbf{x},k})$ are uncorrelated. In particular,

$$\mathbb{E}\mathbf{S}_n^{\mathbf{x}} = \mathbf{x},$$

$$\operatorname{Cov}(S_n^{\mathbf{x},i}, S_n^{\mathbf{x},j}) = \begin{cases} \frac{n}{k}, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{4.3}$$

Therefore the covariance matrix of $\mathbf{S}_n^{\mathbf{x}}$ is $\frac{n}{k}\mathbf{I}$ where $\mathbf{I}$ is the $k \times k$ identity matrix.

The problem of describing the recurrence properties of the simple symmetric random walk in k dimensions is solved by the theorem of Pólya below. The proof will be preceded by the following preliminary lemma.

Lemma 1. Let $\mathbf{S}_0 = 0$ and

$$\begin{aligned} r_n &= P(\mathbf{S}_n = \mathbf{0}) \\ f_n &= P(\mathbf{S}_n = \mathbf{0} \quad \text{for the first time after time 0 at } n), \quad n \geq 1. \end{aligned} \tag{4.4}$$

Let $\hat{r}(s)$ and $\hat{f}(s)$ denote the respective probability generating functions of $\{r_n\}_{n=0}^{\infty}$ and $\{f_n\}_{n=0}^{\infty}$ defined by

$$\hat{r}(s) = \sum_{n=0}^{\infty} r_n s^n, \quad \hat{f}(s) = \sum_{n=0}^{\infty} f_n s^n \quad (0 < s < 1). \tag{4.5}$$

Then

$$\hat{r}(s) = \frac{1}{1 - \hat{f}(s)}. \tag{4.6}$$

The probability γ of eventual return to the origin,

$$\gamma := \sum_{n=1}^{\infty} f_n = \hat{f}(1^-), \tag{4.7}$$

satisfies $\gamma < 1$ if and only if $\beta := \hat{r}(1^-) < \infty$.

Proof. One has the convolution equation

$$r_n = \sum_{j=0}^{n} f_j r_{n-j} \qquad \text{for } n = 1, 2, \ldots, \qquad r_0 = 1, \quad f_0 = 0, \tag{4.8}$$

which transforms as

$$\hat{r}(s) = 1 + \sum_{n=1}^{\infty}\sum_{j=0}^{n} f_j r_{n-j} s^j s^{n-j} = 1 + \sum_{j=0}^{\infty}\left(\sum_{m=0}^{\infty} r_m s^m\right) f_j s^j = 1 + \hat{f}(s)\hat{r}(s). \tag{4.9}$$

Therefore (4.6) follows. Note that by the monotone convergence theorem, $\hat{r}(s) \nearrow \hat{r}(1^-)$ and $\hat{f}(s) \nearrow \hat{f}(1^-)$ as $s \nearrow 1$. If $\hat{f}(1^-) < 1$, then $\hat{r}(1^-) = \lim s \uparrow 1(1 - \hat{f}(s))^{-1} < \infty$. If $\hat{f}(1^-) = 1$, then $\hat{r}(1^-) = \lim_{s\nearrow 1}(1 - \hat{f}(s))^{-1} = \infty$. Therefore, $\gamma < 1$ if and only if $\beta := \hat{r}(1^-) < \infty$. ■

***Theorem 4.1** (Pólya).* $\{\mathbf{S}_n^{\mathbf{x}}\}_{n=0}^{\infty}$ is recurrent for $k = 1, 2$ and transient for $k \geq 3$.

Proof. We will simply compute the criteria of Lemma 1. Namely, we note that $\mathbf{0}$ is transient (i.e., $\gamma < 1$) or recurrent (i.e., $\gamma = 1$) if and only if $\beta := \hat{r}(1^-) < \infty$ or $\hat{r}(1^-) = \infty$), respectively. The result has already been obtained for $k = 1$ by a different method; also see Exercise 1. It is sufficient to consider recurrence/transience of $\mathbf{0}$ for $\{\mathbf{S}_n = \mathbf{S}_n^{\mathbf{0}}\}_{n=0}^{\infty}$ (Exercise 2). This criterion is applied to the case $k = 2$ as follows. Since a return to $\mathbf{0}$ is possible at time $2n$ if and only if the numbers of steps among the $2n$ in the positive horizontal and vertical directions equal the respective numbers of steps in the negative directions,

$$\begin{aligned} r_{2n} &= 4^{-2n}\sum_{j=0}^{n}\frac{(2n)!}{j!j!(n-j)!(n-j)!} = \frac{1}{4^{2n}}\binom{2n}{n}\sum_{j=0}^{n}\binom{n}{j}^2 \\ &= \frac{1}{4^{2n}}\binom{2n}{n}\sum_{j=0}^{n}\binom{n}{j}\binom{n}{n-j} = \frac{1}{4^{2n}}\binom{2n}{n}^2. \end{aligned} \tag{4.10}$$

The combinatorial identity used to get the last line of (4.10) follows by considering the number of ways of selecting samples of size n from a population of n objects of type 1 and n objects of type 2 (Exercise 3). Apply Stirling's formula to (4.10) to get $r_{2n} = O(1/n) > c/n$ for some $c > 0$. Therefore, $\beta = \hat{r}(1) = +\infty$ and so $\mathbf{0}$ is recurrent in the case $k = 2$. Let us include the case $k = 3 = 2 + 1$ in a general inductive argument for all $k \geq 3$ to show that

$$r_{2n} \equiv r_{2n}^{(k)} \leq c_k n^{-k/2}, \quad k \geq 3. \tag{4.11}$$

For this we use induction on the dimension k, but first condition on the binomially distributed number N_{2n} of times the $(k+1)$-st coordinate is selected among paths in the event $[S_{2n} = 0]$, to obtain the recursive relation

$$r_{2n}^{(k+1)} = \sum_{j=0}^{n} P(S_{2n} = 0|N_{2n} = 2n - 2j)P(N_{2n} = 2n - 2j)$$

$$= \sum_{j=0}^{n} \binom{2n}{2j} (\frac{1}{k+1})^{2n-2j} (\frac{k}{k+1})^{2j} r_{2n-2j}^{(1)} r_{2j}^{(k)}, \tag{4.12}$$

i.e., given that there are $2n - 2j$ selections of the $(k+1)$-st coordinate in the event $[S_{2n} = 0]$, the one-dimensional random walk of the $(k+1)$-st coordinate viewed at times of those selections must return to zero in $2n - 2j$ steps, and the k-dimensional random walk viewed at the other times must independently return to zero in $2j$ steps. One may easily check that the terms $j = 0, n$ are $O(n^{-\frac{k+1}{2}})$; recall that $a_n = O(b_n)$ means that there is a constant c (independent of n) such that $|a_n| \le c|b_n|$ for all n. To simplify notation let us use c to signify a constant which does not depend on n, though we will not otherwise keep track of its revised values in subsequent estimates. Also note that $r_{2n-2j}^{(1)} = O((2n-2j)^{-\frac{1}{2}}) = O(\frac{n^{\frac{1}{2}}}{2n-2j+1})$ and by induction $r_{2j}^{(k)} = O((2j)^{\frac{k}{2}})$. To (inductively) bound the expression for $r_{2n}^{(k+1)}$, separately, consider cases of even and odd $k = 2m, 2m+1$, say, respectively. In the even case (k = 2m), for example, the $r_{2j}^{(k)}$-term may be further bounded by a constant (independent of n) times $(2j)^{-m} \le c/(2j+1)\cdots(2j+m)$. Use this and adjust the binomial coefficient $\binom{2n}{2j}$ accordingly to obtain $r_{2n}^{(k+1)} \le c\frac{n^{\frac{1}{2}}(2n)!}{(2n+m+1)!} \sum_{j=1}^{n-1} \binom{2n+m+1}{2j+m} (\frac{1}{k+1})^{2n-2j+1} (\frac{k}{k+1})^{2j+m} \le c\frac{n^{\frac{1}{2}}(2n)!}{(2n+m+1)!} \le c\frac{n^{\frac{1}{2}}}{n^{m+1}} = cn^{-\frac{k+1}{2}}$. Note that this includes the case $k = 2$ so that, in particular we have proven $r_{2n}^{(3)} \le cn^{-\frac{3}{2}}$. The general case of odd $k = 2m+1$ follows directly from the above bound for $k = 2m$. That is, $r_{2j}^{(2m+1)} \equiv r_{2j}^{(k+1)} \le cn^{-\frac{k+1}{2}} = cn^{-\frac{2m+1}{2}}$. To complete the proof simply note that for $k \ge 3$,

$$\sum_n r_n^{(k)} < \infty. \tag{4.13}$$

In particular, 0 is recurrent in dimensions $k = 1, 2$ and transient for all higher dimensions $k \ge 3$. ■

Exercises

1. Show for one-dimensional simple symmetric random walk that $\sum_{n=0}^{\infty} P(S_n = 0)$ diverges.
2. Show that the k-dimensional simple symmetric random walk $\{\mathbf{S}_n^{\mathbf{x}} : n = 0, 1, 2, \dots\}$ is recurrent or transient according to whether $\mathbf{0}$ is a recurrent or transient state for $\{\mathbf{S}_n : n = 0, 1, 2, \dots\}$. [*Hint*: For arbitrary states $\mathbf{x}, \mathbf{y}$ consider the number of visits to $\mathbf{y}$ starting from $\mathbf{x}$ as a sum of indicators.]
3. Verify the combinatorics required for the last equality in (4.10).

4. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be i.i.d. random vectors with values in $\mathbb{Z}^k$ $(k \geq 1)$. Suppose that $\mathbb{E}|X_1^{(j)}| < \infty$, $j = 1, \ldots k$, where $\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(k)})$, $n = 1, 2, \ldots$. Let $\mathbf{S}_n := \mathbf{X}_1 + \cdots + \mathbf{X}_n, n = 1, 2, \ldots$. Show that if $\boldsymbol{\mu} = \mathbb{E}\mathbf{X}_1 = (\mathbb{E}X_1^{(1)}, \ldots, \mathbb{E}X_1^{(k)}) \neq \mathbf{0}$ then $P(\mathbf{S}_n = \mathbf{0} \text{ i.o.}) = 0$. [*Hint*: Use the strong law of large numbers].
5. (i) Show that for the 2-dimensional simple symmetric random walk, the probability of a return to (0, 0) at time $2n$ is the same as that for two independent walkers, one along the horizontal and the other along the vertical, to be at (0, 0) at time $2n$. Also verify this by a geometric argument based on two independent walkers with step size $1/\sqrt{2}$ and viewed along the axes rotated by 45°.
   (ii) Show that relations (4.8) hold for a general random walk on the integer lattice in any dimension.
6. (i) Show that the result of Exercise 5(i) above does not hold in $k = 3$ dimensions.
   (ii) Show that the motion of three independent simple symmetric random walkers starting at (0, 0, 0) in $\mathbb{Z}^3$ is transient.
7. Calculate the probability that the simple symmetric k-dimensional random walk will return i.o. to a previously occupied site.[1] [*Hint*: The conditional probability, given $\mathbf{S}_0, \ldots, \mathbf{S}_n$, that $\mathbf{S}_{n+1} \notin \{\mathbf{S}_0, \ldots, \mathbf{S}_n\}$ is at most $(2k-1)/2k$. Check that

$$P(\mathbf{S}_{n+1}, \ldots, \mathbf{S}_{n+m} \in \{\mathbf{S}_0, \ldots, \mathbf{S}_n\}^c) \leq \left(\frac{2k-1}{2k}\right)^m$$

for each $m \geq 1$.]

[1] For a more detailed perspective on contemporary problems of this flavor see Lawler and Limic (2010).

Chapter 5
The Poisson Process, Compound Poisson Process, and Poisson Random Field

Poisson processes broadly refer to stochastic processes that are the result of counting occurrences of some random phenomena (points) in time or space such that occurrences of points in disjoint regions are statistically independent, and counts of two or more occurrences in an infinitesimally small region are negligible. This chapter provides the definition and some characteristic properties of both homogeneous and inhomogeneous Poisson processes, and more general random fields; the latter refers to occurrences in non-linearly ordered (e.g., non-temporal) spaces. The compound Poisson process is a fundamentally important example from the perspective of both applications and general representations of processes with independent increments. As such it may be viewed as a continuous parameter generalization of the random walk.

The *law of rare events* provides the Poisson distribution, $p_k = \frac{\nu^k}{k!}e^{-\nu}, k = 0, 1, 2, \ldots,$ with parameter $\nu > 0$ as an approximation to the Binomial distribution, $\binom{n}{k}p^k(1-p)^{n-k}, k = 0, 1, 2 \ldots, n,$ with parameters n, p in the limit as $n \to \infty, p \downarrow 0, np = \nu > 0$. This is a rather simple calculus exercise to verify (see Exercise 1) also referred to as the *Poisson approximation to the binomial distribution*.

A probabilistic derivation that also provides error bounds was given by Lucien Le Cam[1] using a remarkable concept referred to as *coupling*. For this, let $\mathbf{X} = \{X_1, X_2, \ldots X_n\}$ be i.i.d. Bernoulli $0-1$-valued random variables with parameter $p = P(X_i = 1), i \geq 1$, and let $\mathbf{Y} = \{Y_1, Y_2, \ldots Y_n\}$ be an i.i.d. sequence of

[1] Le Cam (1960a).

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_5

Poisson distributed random variables with parameter ν/n. Then $S = \sum_{j=1}^{n} X_j$ has the desired Binomial distribution, and $T = \sum_{j=1}^{n} Y_j$ has the desired Poisson distribution.

***Lemma 1** (Coupling Inequality).* For any set A of non-negative integers one has

$$|P(S \in A) - P(T \in A)| \leq P(S \neq T).$$

Proof. One has

$$\begin{aligned} &|P(S \in A) - P(T \in A)| \\ &\quad = |P(S \in A, T \in A) + P(S \in A, T \in A^c) - P(S \in A, T \in A) - P(S \in A^c, T \in A)| \\ &\quad \leq P(S \neq T). \end{aligned}$$

■

The innovation involved in coupling is that of reducing the chance of $S \neq T$ by introducing correlations (statistical dependence) between the random variables. First observe that

$$P(S \neq T) = P(\sum_{j=1}^{n} X_j \neq \sum_{j=1}^{n} Y_j) \leq P(\cup_{j=1}^{n}[X_j \neq Y_j]) \leq nP(X_1 \neq Y_1).$$

The probability is reduced by the construction given in the following proof.

***Proposition 5.1** (Poisson Approximation to Binomial Distribution).* For $\nu = np$,

$$|\binom{n}{k} p^k (1-p)^{n-k} - \frac{\nu^k}{k!} e^{-\nu}| \leq np^2.$$

Proof. Let $U_1, U_2, \ldots U_n$ be i.i.d. uniformly distributed random variables on the interval [0, 1]. Define

$$X_j = \begin{cases} 0 & 0 \leq U_j < 1-p \\ 1 & 1-p \leq U_j < 1, \end{cases} \tag{5.1}$$

and

$$Y_j = \begin{cases} 0 & 0 \leq U_j < e^{-p} \\ 1 & e^{-p} \leq U_j < e^{-p} + pe^{-p} \\ 2 & e^{-p} + pe^{-p} \leq U_j < e^{-p} + pe^{-p} + \frac{p^2}{2!}e^{-p} \\ \vdots & \text{etc.} \end{cases} \tag{5.2}$$

This provides a coupled construction of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ with the desired distributions. Moreover, using $e^{-p} \geq 1 - p$, one has

$$
\begin{aligned}
P(S \neq T) &\leq nP(X_1 \neq Y_1) \\
&= n\{P(X_1 = 0, Y_1 \geq 1) + P(X_1 = 1, Y_1 = 0) + P(X_1 = 1, Y_1 \geq 2)\} \\
&= n\{P(U_j < 1 - p, U_j \geq e^{-p}) + P(U_j \geq 1 - p, U_j < e^{-p}) \\
&\quad + P(U_j \geq e^{-p} + pe^{-p}, U_j \geq 1 - p)\} \\
&= n\{0 + (e^{-p} - 1 + p) + (1 - e^{-p} - pe^{-p})\} \\
&= np(1 - e^{-p}) \leq np^2.
\end{aligned}
$$

■

An inspection of the proof shows that in fact one may approximate the sum of independent Bernoulli random variables with parameters $p_j = P(X_j = 1), 1 \leq j \leq n$, by a Poisson distribution with parameter $\nu = \sum_{j=1}^{n} p_j$ with an error at most $\sum_{j=1}^{n} p_j^2 \leq \max_{1 \leq j \leq n} p_j \sum_{j=1}^{n} p_j$; see Exercise 2.

The Poisson process is a *renewal counting process* with i.i.d. exponential inter-arrival times; see Chapters 25, 26. Let $T_0, T_1, \ldots$ be i.i.d. exponentially distributed random variables with parameter $\lambda > 0$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Fixing a time origin at $t = 0$, T_0 denotes the time to the first occurrence, T_1 the time between the first and second occurrences, and so on. Thus T_k may be viewed as an *inter-arrival time* , or as a *holding time* in the state k, for a particle moving on $\{0, 1, 2, \ldots\}$, one step at a time. For $t \geq 0$, the total number of occurrences by time t is counted by

$$
X_t = \begin{cases} 0 & \text{if } T_0 > t \\ \sup\{n \geq 1 : T_0 + \cdots + T_{n-1} \leq t\} & \text{else.} \end{cases} \tag{5.3}
$$

Here the indexing set is $\Lambda = [0, \infty)$ and a priori $S = \mathbb{Z}_+ \cup \{\infty\}$ with $\mathcal{S} = 2^S$, the power set of S. Note that $P(X_t = 0) = P(T_0 > t) = e^{-\lambda t}, t \geq 0$. Also, for $k \geq 1$, the k-fold convolution of exponential densities $\lambda e^{-\lambda t}, t \geq 0$, is the Gamma density $\lambda^k t^{k-1} e^{-\lambda t}/(k-1)!, t \geq 0$, with parameters k, λ. In particular, therefore, the time $A_k := T_0 + \cdots + T_{k-1}$ of the kth arrival has a Gamma distribution with parameters k, λ and

$$
\begin{aligned}
Q_t(\{k\}) = P(X_t = k) &= P(T_0 + \cdots + T_{k-1} \leq t < T_0 + \cdots + T_k) \\
&= P(T_0 + \cdots + T_k > t) - P(T_0 + \cdots + T_{k-1} > t) \\
&= \frac{(\lambda t)^k}{k!} e^{-\lambda t} + P(T_0 + \cdots + T_{k-1} > t) \\
&\quad - P(T_0 + \cdots + T_{k-1} > t)
\end{aligned}
$$

$$= \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{5.4}$$

where the first term of the third line results from integration by parts of the Gamma density; see Exercise 8. More generally let us show that $\{X_t : t \geq 0\}$ is a process with *independent* and *stationary* (or *homogeneous*) Poisson distributed increments in a sense made precise as follows.

Proposition 5.2. For any $0 = t_0 < t_1 < \cdots < t_m, k_i \in \mathbb{Z}_+, 1 \leq i \leq m$,

$$P(X_{t_i} - X_{t_{i-1}} = k_i,\ i = 1, \ldots, m) = \prod_{i=1}^{m} P(X_{t_i} - X_{t_{i-1}} = k_i) \tag{5.5}$$

with

$$P\left(X_{t_i} - X_{t_{i-1}} = k_i\right) = \frac{[\lambda(t_i - t_{i-1})]^{k_i}}{k_i!} e^{-\lambda(t_i - t_{i-1})} = P\left(X_{t_i - t_{i-1}} = k_i\right). \tag{5.6}$$

To verify (5.5), (5.6) we will first derive an interesting and useful property concerning the conditional distribution of the successive arrival times in the period from 0 to t given the number of arrivals in $[0, t]$. Proposition 5.2 will essentially follow as a corollary.

Proposition 5.3 *(Order Statistic Property (o.s.p.)).* Let $A_1 = T_0$, $A_2 = T_0 + T_1, \ldots, A_j = T_0 + T_2 + \cdots + T_{j-1}, \ldots$, denote successive arrival times of the process $\{X_t : t \geq 0\}$. Then the conditional distribution of $(A_1, A_2, \ldots, A_k)$ given $[X_t = k]$ is the same as that of k increasingly ordered independent random variables each having the uniform distribution on $(0, t]$.

Proof. Let $U_0, U_1, \ldots, U_{k-1}$ be k i.i.d. random variables uniformly distributed on $(0, t]$. Let $U_{(0)}$ be the smallest of $\{U_0, U_1, \ldots, U_{k-1}, U_{(1)}\}$ the next smallest, etc., so that with probability one $U_{(0)} < U_{(1)} < \cdots < U_{(k-1)}$. Since each realization of the *order statistic* defined by $(U_{(0)}, U_{(1)}, \ldots, U_{(k-1)})$ corresponds to exactly one of $k!$ permutations of $(U_0, U_1, \ldots, U_{k-1})$, each with the same probability density $1/t^k$, the joint density of $(U_{(0)}, U_{(1)}, \ldots, U_{(k-1)})$ is given by

$$g(s_0, \ldots, s_{k-1}) = \begin{cases} \frac{k!}{t^k} & \text{if } 0 < s_0 < s_1 < \cdots < s_{k-1} \leq t \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

For the remainder of the proof of Proposition 5.3 we simply compute the conditional density of $(A_1, \ldots, A_k)$ given $[X_t = k]$ and compare it to (5.7). First note that T_0, $T_1, \ldots, T_k$ are i.i.d. with joint density given by

$$f(x_0, x_1, \ldots, x_k) = \begin{cases} \lambda^{k+1} \exp\left\{-\lambda \sum_{i=0}^{k} x_i\right\} & \text{if } x_i > 0 \text{ for all } i, \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

Since the Jacobian of the transformation

$$(x_0, x_1, \ldots, x_k) \to (x_0, x_0 + x_1, \ldots, x_0 + x_1 + \cdots + x_k)$$

is 1, the joint density of $T_0, T_0 + T_1, \ldots, T_0 + T_1 + \cdots + T_k$ is obtained from (5.8) as

$$h(t_0, t_1, \ldots, t_k) = \lambda^{k+1} e^{-\lambda t_k} \qquad \text{for } 0 < t_0 < t_1 < t_2 < \cdots < t_k. \tag{5.9}$$

The density of the conditional distribution of $T_0, T_0 + T_1, \ldots, T_0 + T_1 + \cdots + T_k$ given $[X_t = k] \equiv [T_0 + T_1 + \cdots + T_{k-1} \le t < T_0 + T_1 + \cdots + T_k]$ is therefore

$$q(s_0, s_1, \ldots s_k) = \begin{cases} \dfrac{\lambda^{k+1} e^{-\lambda s_k}}{P(T_0 + T_1 + \cdots + T_{k-1} \le t < T_0 + T_1 + \cdots + T_k)} \\ \quad \text{if } 0 < s_0 < s_1, \ldots < s_k \text{ and } s_{k-1} \le t < s_k, \\ 0 \quad \text{otherwise.} \end{cases} \tag{5.10}$$

Therefore, using (5.4),

$$q(s_0, s_1, \ldots, s_k) = \begin{cases} \dfrac{\lambda^{k+1} e^{-\lambda s_k}}{e^{-\lambda t} \frac{(\lambda t)^k}{k!}} = \dfrac{\lambda k! e^{-\lambda s_k}}{e^{-\lambda t} t^k} \\ \quad \text{if } 0 < s_0 < \cdots < s_k, s_{k-1} \le t < s_k, \\ 0 \quad \text{otherwise.} \end{cases} \tag{5.11}$$

Integrating this over s_k we get the conditional density of $T_0, T_0 + T_1, \ldots, T_0 + T_1 + \cdots + T_{k-1}$ given $[X_t = k]$ as

$$p(s_0, s_1, \ldots, s_{k-1}) = \begin{cases} \dfrac{(\lambda) k!}{t^k e^{-\lambda t}} \displaystyle\int_t^\infty e^{-\lambda s_k} ds_k = \dfrac{k!}{e^{-\lambda t} t^k} e^{-\lambda t} = \dfrac{k!}{t^k} \\ \quad \text{if } 0 < s_0 < s_1 < \cdots < s_{k-1} \le t, \\ 0 \quad \text{otherwise.} \end{cases} \tag{5.12}$$

Thus, as asserted, p is the same as g in (5.7). ■

The verification of (5.5), (5.6) can now be made using Proposition 5.3 and (5.4) as follows:

Proof. (of Proposition 5.2) Writing $k = k_1 + \cdots + k_m$, one uses (5.4) to get

$$P(X_{t_i} - X_{t_{i-1}} = k_i, \ i = 1, \ldots, m)$$

$$= P(X_{t_i} - X_{t_{i-1}} = k_i,\ i = 1, \ldots, m \mid X_{t_m} = k) \cdot e^{-\lambda t_m} \frac{(\lambda t_m)^k}{k!}. \quad (5.13)$$

By Proposition 5.3, the conditional probability above equals the probability that of k independent random variables each uniformly distributed on $(0, t_m]$, k_1 are in $(0, t_1)$, k_2 are in $(t_1, t_2], \ldots, k_m$ are in $(t_{m-1}, t_m]$. This latter probability is

$$\frac{k!}{k_1! k_2! \cdots k_m!} \left(\frac{t_1}{t_m}\right)^{k_1} \left(\frac{t_2 - t_1}{t_m}\right)^{k_2} \cdots \left(\frac{t_m - t_{m-1}}{t_m}\right)^{k_m}. \quad (5.14)$$

Using this in (5.13), one arrives at (5.5), (5.6). ■

Outside a subset of Ω of zero probability X_t takes values in $\mathbb{Z}_+$ for all $t \geq 0$. Such a subset may be removed and $S = \mathbb{Z}_+$ may then be taken as the state space. Also, for each $\omega \in \Omega$, the function $t \to X_t(\omega)$, referred to as a *sample path* of the process $\{X_t : t \geq 0\}$, defined by (5.3) is a right-continuous non-decreasing step function. This stochastic process is called a *Poisson process*. Note that by (5.5) the increments $X_{t_i} - X_{t_{i-1}}$, $i = 1, \ldots, m$, are independent with Poisson distributions having means $\lambda(t_i - t_{i-1})$, respectively. The parameter $\lambda > 0$ is called the *intensity parameter*.

As noted above, for fixed $\omega \in \Omega$, the sample paths $t \to X_t(\omega)$ are right-continuous step functions. Moreover, for any $t > 0$, $P(X_t < \infty) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = 1$ so that there can be at most finitely many occurrences in any finite time interval. In this regard one says that the process is *non-explosive*; see Exercise 13 for contrast. By (5.3), one may view the stochastic process $\{X_t : t \geq 0\}$ as a random path in the *path space* Γ of right-continuous non-decreasing step functions on $[0, \infty)$. The *distribution* of the stochastic process $X := \{X_t : t \geq 0\}$ is the induced probability measure $Q = P \circ X^{-1}$ on Γ which is uniquely specified by the probabilities assigned to *finite-dimensional events* as follows: Let $\mathcal{G}$ be the σ-field generated by finite-dimensional subsets of Γ of the form

$$F = \{g \in \Gamma : g(t_i) = k_i, \quad i = 1, \ldots, m\}, \quad (5.15)$$

where $k_i \in \mathbb{Z}_+, 0 = t_0 < t_1 < \cdots < t_m$; alternative see Exercise 4. Then $Q = P \circ X^{-1}$ is a probability on the path space $(\Gamma, \mathcal{G})$ and, in particular, the probability of a finite-dimensional event has the following formula

$$Q(F) = P(\{\omega \in \Omega : X_{t_i}(\omega) = k_i, \quad i = 1, \ldots, m\}) \quad (5.16)$$

$$= \prod_{j=1}^{m} \frac{[\lambda(t_j - t_{t-1})]^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} e^{-\lambda(t_j - t_{j-1})}, \quad (5.17)$$

for $0 = k_0 \leq k_1 \leq k_2 \leq \cdots \leq k_m \in \mathbf{Z}_+$, $t_0 = 0$, see Exercise 9. The *finite-dimensional distributions* of $\{X_t : t \geq 0\}$ refer to these joint distributions of $(X_{t_1}, \ldots, X_{t_m})$ for fixed but arbitrary $0 \leq t_1 < t_2 < \cdots < t_m$.

Remark 5.1. In view of the above developments, an equivalent alternative definition of the homogeneous Poisson process X may be given by simply specifying a process having right-continuous and non-decreasing unit jump step functions as sample paths starting at $X_0 = 0$, and having stationary independent Poisson distributed increments $X_t - X_s, s < t$, with mean $\lambda(t-s)$, for some $\lambda > 0$.

Remark 5.2. A martingale characterization of the Poisson process is given in Chapter 15.

The implicit *order* structure of the real numbers is both natural and essential for the previous definition of the Poisson process on $\Lambda = [0, \infty)$ in terms of a sequence of i.i.d. exponential inter-arrival times. However, one may extract certain basic structure, in particular the independence of the numbers of points in disjoint regions, for an extension of the model to the random occurrence of points in higher dimensional (unordered) spaces. To motivate the extension, consider again the Poisson process $X = \{X_t : t \geq 0\}$ on $\Lambda = [0, \infty)$ in terms of the corresponding *random counting measure* M defined on the Borel σ-field $\mathcal{B}(\Lambda)$ by regarding X as its distribution function, i.e.,

$$M(a, b] = X_b - X_a, \quad 0 \leq a \leq b. \tag{5.18}$$

Equivalently $M(dt) = \sum_{n=1}^{\infty} \delta_{A_n}(dt)$, where $A_1 = T_0, A_2 = T_0 + T_1, \ldots,$ (Exercise 6). That is, the random points $A_1, A_2, \ldots$ distributed in Λ may be represented in terms of the random atoms of the counting measure M.

To proceed with the general definition of a Poisson random field, let $(\Lambda, \mathcal{L}, \rho)$ be a measure space with an arbitrary non-negative sigma-finite measure $\rho(ds)$. Let $\mathcal{M}$ denote the space of all non-negative, integer-valued measures on $(\Lambda, \mathcal{L})$ and give $\mathcal{M}$ the σ-field $\mathcal{F}_{\mathcal{M}}$ generated by sets of the form $\{\mu \in \mathcal{M} : \mu(B) = n\}$ for $B \in \mathcal{L}$, and integral $n \geq 0$ or $n = \infty$.

Definition 5.1. The random field $M = \{M(B) : B \in \mathcal{L}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is called the *Poisson random field with intensity* $\rho(dx)$ if

1. $M : \Omega \to \mathcal{M}$ is measurable with respect to the designated σ-fields $\mathcal{F}$ and $\mathcal{F}_{\mathcal{M}}$, respectively.
2. For each $B \in \mathcal{L}$ such that $\rho(B) < \infty$, $P(M(B) = n) = \frac{\rho(B)^n}{n!} e^{-\rho(B)}, n = 0, 1, \ldots$. If $\rho(B) = \infty$, then $M(B) = \infty$ a.s.
3. For pairwise disjoint sets $B_k, k = 1, \ldots, n$ in $\mathcal{L}$, $n \geq 1$, the random variables $M(B_1), \ldots, M(B_n)$ are independent.

As a special case this definition includes the *inhomogeneous Poisson process* on $\Lambda = [0, \infty)$, or even $\Lambda = \mathbf{R}$, with non-constant intensity $\rho(dx) = \lambda(x)dx$ for a non-negative (measurable) function λ on Λ; see Exercise 14.

The problem now is to show that a Poisson random field exists on a given σ-finite measure space $(\Lambda, \mathcal{L}, \rho)$.

Theorem 5.4. Let $(\Lambda, \mathcal{L}, \rho)$ be a σ-finite measure space,$\rho \neq 0$. Then there is a Poisson random field M on Λ with intensity ρ.

Proof. Since ρ is σ-finite there is a collection of disjoint measurable subsets $\Lambda_1, \Lambda_2, \ldots$ of Λ with $0 < \rho(\Lambda_j) < \infty, j \geq 1$, and $\Lambda = \cup_{j=1}^{\infty} \Lambda_j$. Using the Kolmogorov extension theorem, construct a probability space $(\Omega, \mathcal{F}, P)$ with a sequence $X_1, X_2, \ldots$ of independent Poisson distributed random variables with respective means $\rho(\Lambda_j), j \geq 1$, and, for each $j \geq 1$, a sequence of i.i.d. random variables $A_1^{(j)}, A_2^{(j)}, \ldots$, also independent of $X_1, X_2, \ldots$, such that $A_i^{(j)}$ is distributed on Λ_j according to the probability distribution $\rho(dx)/\rho(\Lambda_j)$. Now, for $B \in \mathcal{L}$ define

$$M(B) = \sum_{j=1}^{\infty} \mathbf{1}[X_j \geq 1] \sum_{k=1}^{X_j} \mathbf{1}[A_k^{(j)} \in B \cap \Lambda_j].$$

Note that $M(\cdot) = \lim_{N \to \infty} \sum_{j=1}^{N} \mathbf{1}[X_j \geq 1] \sum_{k=1}^{X_j} \delta_{A_k^{(j)}}(\cdot)$ is the a.s. limit of $\mathcal{F}_{\mathcal{M}}$ measurable maps of Ω into $\mathcal{M}$, and is therefore measurable. It is also a simple calculation to see that for disjoint sets $B_1, \ldots, B_m$ in $\mathcal{L}$ one has for any positive $r_1, \ldots, r_m$ that (Exercise 5)

$$\mathbb{E} \exp\{-\sum_{i=1}^{m} r_i M(B_i)\} = \prod_{i=1}^{m} \exp\{(e^{-r_i} - 1)\rho(B_i)\}, \tag{5.19}$$

since the factors $\exp\{-r_i M(B_i)\}$ are independent for $i = 1, 2, \ldots, m$. ■

Example 1. Suppose that M is a non-homogeneous Poisson process on $\Lambda = \mathbb{R}$ with intensity measure $\rho(dx) = e^{-x}dx$. Then let us observe that since $\int_0^{\infty} e^{-x}dx < \infty$, there is a right-most occurrence, say $T_{\max}$, for M. Specifically one has

$$\mathbb{E}M(0, \infty) = \int_0^{\infty} \rho(x)dx < \infty,$$

and therefore $M(0, \infty) < \infty$ with probability one. Since there is a.s. at most a finite number of occurrences to the right of 0, there must be a largest one. The distribution is easily calculated by

$$P(T_{\max} \leq x) = P(M(x, \infty) = 0) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

This is referred to as the standard *Gumbel* or *Frechet* extreme value distribution.

Although time-inhomogeneous Poisson processes occur naturally in various counting scenarios, for example, as a result of seasonality in rain shower counts, rush-hours in traffic accident counts, etc., the inhomogeneity may be removed by a change of time-scale.

Proposition 5.5 *(Homogenization of the Poisson Process).* Suppose that $\mathbf{X} = \{X_t : t \geq 0\}$ is an inhomogeneous Poisson process with intensity measure $\rho(dt) = \rho(t)dt$ on $\Lambda = [0, \infty)$. If $\rho(t)$ is strictly positive and $\int_0^\infty \rho(s)ds = \infty$, then there is a *time-change* function $\tau : [0, \infty) \to [0, \infty)$ such that $\tilde{\mathbf{X}} = \{X_{\tau(t)} : t \geq 0\}$ is a Poisson process with (constant) unit intensity with respect to Lebesgue measure.

Proof. The function $\gamma(t) = \int_0^t \rho(s)ds, t \geq 0$, is continuous, one-to-one, and onto $[0, \infty)$. Define $\tau(t) = \gamma^{-1}(t), t \geq 0$. From here it is simple to check that $\tilde{\mathbf{X}}$ has non-decreasing sample paths with non-negative, independent Poisson distributed increments with, $\tilde{X}_0 = 0, \mathbb{E}(\tilde{X}_t - \tilde{X}_s) = t - s, 0 \leq s < t$. ■

In another direction, the Poisson process leads to a natural continuous parameter generalization of the random walk as follows.

Definition 5.2. Let $N = \{N_t : t \geq 0\}$ be a homogeneous Poisson process and let $Y_1, Y_2, \ldots$ be an i.i.d. sequence in $\mathbb{R}^n$, independent of X. Then the process $X_t = \sum_{j=1}^{N_t} Y_j$ is referred to as a *compound Poisson process* on $\mathbb{R}^n$.

The sample paths of the compound Poisson process can be depicted as in Figure 1.1, but with the $+1$ unit increments now replaced by realizations of $Y_1, Y_2, \ldots$ in the up or down directions according to their positive or negative values, respectively.

The compound Poisson process enjoys many interesting properties inherited from the random walk and the Poisson process that are delineated in exercises. In particular, one can readily check the following property (Exercise) of the distribution Q_t of X_t for each t:

Definition 5.3. A probability distribution Q on the Borel σ-field of $\mathbb{R}$ is said to be *infinitely divisible* if for each integer $n \geq 1$ there is a probability distribution Q_n such that Q is the n-fold convolution $Q = Q^{*n}$.

Proposition 5.6. The compound Poisson process X is a process with stationary, independent increments. For each $t > 0$, the distribution of X_t is infinitely divisible.

Along these same lines note that, given N_t, the conditional distribution of X_t is given by an N_t-fold convolution of the distribution Q of Y_1. In particular, denoting the (unconditional) distribution of X_t by Q_t, one has the "continuous convolution property"

Proposition 5.7.

$$Q_{t+s} = Q_t * Q_s, \quad s, t \geq 0. \tag{5.20}$$

Proof. Simply condition on N_{t+s}, apply the binomial theorem, and make a change of variable in the sum as follows: For $B \in \mathcal{B}(\mathbb{R}^n)$,

$$Q_{t+s}(B) = \mathbb{E}Q^{*N_{t+s}}(B)$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} Q^{*k}(B)\frac{\lambda^k (t+s)^k}{k!} e^{-\lambda(t+s)} \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} Q^{*k}(B) \sum_{j=0}^{k} \binom{k}{j} t^j e^{-\lambda t} s^{k-j} e^{-\lambda s} \\
&= \Big(\sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} Q^{*j} e^{-\lambda t} * \sum_{i=0}^{\infty} \frac{(\lambda s)^i}{i!} Q^{*i} e^{-\lambda s}\Big)(B) \\
&= Q_t * Q_s(B). \qquad (5.21)
\end{aligned}
$$

Here the factorization $Q^{*k} = Q^{*j} Q^{*(k-j)}$ is used as well. ■

This structure leads to a special representation of the characteristic function of X_t, referred to as a *Lévy–Khinchine formula*, as follows (Exercise 16).

Proposition 5.8. For the compound Poisson process X one has

$$
\mathbb{E}e^{i\xi X_t} = \exp\{\int_{\mathbb{R}^k} (e^{i\xi y} - 1)\nu_t(dy)\}, \quad \xi \in \mathbb{R}^n,
$$

where $\nu_t(dy) = \lambda t Q(dy)$, and Q is the distribution of Y_1.

Exercises

1. (*A Simple Law of Rare Events*) Suppose that X_n has a binomial distribution with parameters n and $p_n = \nu/n$ for some $\nu > 0$. Use calculus to establish that $\lim_{n\to\infty} P(X_n = k) = \nu^k e^{-\nu}/k!, k = 0, 1, 2, \ldots$.[*Hint*: Use the asymptotic formula $\lim_{n\to\infty}(1 + \frac{x_n}{n})^n = e^x$ whenever $\lim_{n\to\infty} x_n = x$.]
2. (*Poisson approximation to independent Bernoulli sums*) Suppose that $X_1, \ldots, X_n$ are independent Bernoulli $0-1$-valued random variables with $P(X_j = 1) = p_j, 1 \le j \le n$. Let Y have Poisson distribution with parameter $\nu = \sum_{j=1}^n p_j$. Then $|P(\sum_{j=1}^n X_j = k) - P(Y = k)| \le 2\sum_{j=1}^n p_j^2$.
3. (*Infinitesimal description of the Poisson process*) Suppose that $N = \{N_t : t \ge 0\}$ is a non-negative integer-valued (counting) stochastic process having non-decreasing paths with $N_{0+} = 0$. Assume that for any $0 = t_0 < t_1 < \cdots < t_m$, the counts $N_{t_j} - N_{t_{j-1}}, 1 \le j \le m$ are independent, and homogeneous in the sense that the distribution of $N_{t+\Delta} - N_{s+\Delta}$ does not depend on $\Delta > 0$. Assuming infinitesimal probabilities $P(N_\Delta = 1) = \lambda\Delta + o(\Delta)$, and $P(N_\Delta \ge 2) = o(\Delta)$ as $\Delta \downarrow 0$, for some $\lambda > 0$, show: (a) $\frac{d}{dt}P(N_t = 0) = -\lambda P(N_t = 0)$, $P(N_{0+} = 0) = 1$; (b) $\frac{d}{dt}P(N_t = k) = -\lambda P(N_t = k) + \lambda P(N_t = k-1), k \ge 1$; (c) Show that the unique solution to this system of differential equations is given by $P(N_t = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}, t \ge 0$. [*Hint*: To derive (a), (b)

consider the Newton quotients defining the indicated derivative and express the events $[N_{t+\Delta} = k]$ jointly in terms of counts of 0, 1 or more than 2 occurrences in the interval $[t, t + \Delta]$.]

4. Show that one may equivalently define the σ-field $\mathcal{G}$ on Γ as the smallest σ-field such that random variable $\omega \to \gamma(\omega)$, given by $\gamma(\omega)(t)X_t(\omega)$ in (5.3), is a measurable function from Ω to Γ.
5. Verify the formula (5.19) for the multivariate moment generation function of the Poisson random field.
6. Show that the definition of the one-dimensional Poisson random measure given by (5.18) is a.s. purely atomic with atoms at the arrival times $A_1, A_2, \ldots$.
7. Suppose that $X = \{X_t : t \geq 0\}$ is a time-homogeneous Poisson process with intensity parameter $\lambda > 0$. Show that $\lambda = \lim_{n\to\infty} \frac{X_n}{n}$ a.s. [*Hint*: Express X_n in terms of i.i.d. increments.]
8. The *Gamma distribution* with parameters $\lambda > 0$, $\nu > 0$ has density $f_{\nu,\lambda}(x) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$, $x \geq 0$. Verify that

$$\int_t^\infty f_{\nu,\lambda}(s)ds = \lambda^{-1} f_{\nu,\lambda}(t) + \int_t^\infty f_{\nu-1,\lambda}(s)ds.$$

9. Let $\{X_t : t \geq 0\}$ be the Poisson process defined by (5.3). Show for $0 \leq t_1 < t_2 < \cdots < t_m, m \geq 1$,

$$P\left(X_{t_1} = k_1, \ldots, X_{t_m} = k_m\right) = \prod_{j=1}^m \frac{[\lambda(t_j - t_{t-1})]^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} e^{-\lambda(t_j - t_{j-1})},$$

for $k_0 = 0 \leq k_1 \leq k_2 \leq \cdots \leq k_m \in \mathbf{Z}_+$, $t_0 = 0$.

10. Let $(U_{(1)}, U_{(2)}, \ldots U_{(k)})$ be the order statistic from k i.i.d. random variables uniformly distributed on $[0, t]$. Show that $U_{(j)}$ has probability density $f_j(x) = j\binom{k}{j}(\frac{x}{t})^{j-1}(1 - \frac{x}{t})^{k-j} \cdot \frac{1}{t}$, $0 < x \leq t$.
11. (*Thinning*) Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. Bernoulli $0 - 1$ valued random variables, $P(\varepsilon_i = 1) = p = 1 - P(\varepsilon_i = 0)$, $i \geq 1$, and independent of the Poisson process $\{X_t : t \geq 0\}$ defined by (5.3). Let $N_0 = 0$, $N_1 = \inf\{n \geq 1 : \varepsilon_n = 1\}$, $N_r = \inf\{n \geq N_{r-1} + 1 : \varepsilon_n = 1\}$, $r = 2, 3, \ldots$. Define

$$\tilde{T}_0 = T_0 + \cdots + T_{N_1 - 1}$$
$$\tilde{T}_r = T_0 + \cdots + T_{N_{r+1}-1} = \tilde{T}_{r-1} + T_r + \cdots + T_{N_{r+1}-1}.$$

That is, the occurrences of the process $\{X_t : t \geq 0\}$ are one-by-one and independently accepted with probability p or rejected with probability $1 - p$; referred to as *independent thinning* or *independent splitting* the process. Define

$$\tilde{X}_t = \begin{cases} 0 & \text{if } \tilde{T}_0 > t \\ \sup\{n \geq 1 : \tilde{T}_0 + \cdots + \tilde{T}_{n-1} \leq t\}. & \end{cases}$$

Show that $\{\tilde{X}_t : t \geq 0\}$ is a Poisson process[2] with intensity parameter $\tilde{\lambda} = p\lambda$. [*Hint*: Use moment generating functions or characteristic functions. For independence of inter-arrivals show factorization and argue that this is sufficient by the uniqueness theorem for such multivariate transforms.]

12. (*Age and Residual Lifetime*) Let $X_t : t \geq 0$ be a time-homogenous Poisson process with intensity $\lambda > 0$, and arrival times $T_0, T_1, \ldots$. The *age* of the last occurrence at time $t > 0$ is defined by the time since the last occurrence prior to t, $a(t) = t - T_{N(t)-1}$. The *residual lifetime* is defined by $r(t) = T_{N(t)} - t$. [Note: The shift in $N(t)$ accounts for the *first* occurrence being T_0.] Show that $a(t)$ and $r(t)$ are independent, and $a(t) \leq t$ has an exponential distribution with parameter $\lambda > 0$ truncated at t, and $r(t)$ has an (unrestricted) exponential distribution with parameter λ. [*Hint*: For $0 \leq s_1 < t, s_2 > 0$, express $[a(t) \leq s_1, r(t) \leq s_2]$ in terms of the increments of X_t. The age does not exceed $s_1 < t$ if and only if there is at least one occurrence between time $t - s_1$ and t, while the residual time does not exceed $s_2 > 0$ if and only if there is at least one occurrence in time t to $t + s_2$.]
13. (*Feller's Non-explosion Criteria*) Suppose that $T_0, T_1, T_2, \ldots$ are independent exponentially distributed random variables with parameters $\lambda_0, \lambda_1, \lambda_2, \ldots$, respectively. Define

$$\zeta = T_0 + T_1 + \cdots + T_n + \cdots .$$

   (i) Show that $P(\zeta = \infty) = 1$ if and only if $\sum_{j=0}^{\infty} \frac{1}{\lambda j} = \infty$. The event $[Y_t = \infty]$ is referred to as *explosion*[3] in finite time. [*Hint*: $\mathbb{E}\zeta = \sum_{j=0}^{\infty} \frac{1}{\lambda_j}$. So the condition for $P(\zeta < \infty) = 1$ is obvious. Consider $\mathbb{E}e^{-\zeta} = \prod_{j=0}^{\infty} \frac{\lambda_j}{1+\lambda_j} = 1/\prod_{j=0}^{n}(1+\lambda_j^{-1})$, and note that $\prod_{j=0}^{n}(1+\lambda_j^{-1}) \leq e^{\sum_{j=0}^{n} \frac{1}{\lambda_j}}$.]

   (ii) (ii) *Newton's divided differences* are recursively defined for a function f at $n+1$ distinct points $x_0, x_1, \ldots, x_n$ by $[x_0, \ldots, x_n]_f = \frac{[x_0,\ldots,x_{n-1}]_f - [x_1,\ldots,x_n]_f}{x_0 - x_n}$, where $[x_j]_f = f(x_j)$, $j = 0, 1, \ldots n$. Show by induction that $[x_0, \ldots, x_n]_f = \sum_{k=0}^{n} \frac{f(x_k)}{D_n'(x_k)}$, where $D_n(x) = \prod_{j=0}^{n}(x - x_j)$ and $D_n'(x)$ its derivative.

   (iii) Let $e_j(t) = \lambda_j e^{-\lambda_j t}$, $t \geq 0$. Show that for $\lambda_i \neq \lambda_j$, $0 \leq i, j \leq n$, the pdf $f_n = e_0 * \cdots * e_n$ of $T_0 + \cdots + T_n$ is given by

[2] For illustration of a contemporary application of the splitting property to "micromobility" see Fields (2020).

[3] A clever calculation using differential equations of the probability of explosion in a specified time t is given in Feller (1968), Vol I.

$$f_n(t) = (-1)^{n+1} t^n \prod_{j=0}^{n} \lambda_j [\lambda_0 t, \ldots, \lambda_n t]_{\exp}$$

$$= (-1)^{n+1} \prod_{j=0}^{n} \lambda_j \sum_{k=0}^{n} \frac{e^{-\lambda_k t}}{\prod_{j \neq k} (\lambda_k - \lambda_j)}, \quad \exp(x) = e^{-x}, x \geq 0.$$

(iv) (*Yule Branching Process*) Starting with a single progenitor $Y_0 = 1$, after an exponentially distributed time with parameter $\lambda > 0$, the parent particle dies and produces two offspring. The offspring independently of one another and the parent follow the identically distributed exponential lifetime before death and reproduction, from generation to generation. Let Y_t denote the number of particles alive at time t. Show that the process is non-explosive. [*Hint*: Represent as above with $\lambda_n = (n+1)\lambda, n \geq 0$.]

14. Suppose that $\mathbf{X} = \{X_t : t \geq 0\}$ is an inhomogeneous Poisson process with intensity function $\rho(t)$ on $S = [0, \infty)$.

(i) Show that if $\int_0^\infty \rho(s)ds < \infty$, then $P(\sup_{t \geq 0} X_t < \infty) = 1$, [*Hint*: The sample paths are non-decreasing so that $\sup_{t \geq 0} X_t = \lim_{t \to \infty} X_t$ a.s.] i.e., in contrast to explosion, throughout all time at most a finite number of occurrences can occur.

(ii) Calculate the distribution of the first arrival time A_0 of $\mathbf{X}$. In particular, what is $P(A_0 = \infty)$?

(iii) Show that the conditional distribution of A_0 given $[A_0 < \infty]$ in the case $\rho(t) = e^{-t}, t \geq 0$ is the *Frechet extreme value distribution* $e^{e^{-t}}, t \geq 0$.

15. Show that the compound Poisson process has stationary and independent increments. In particular, show that the distribution Q_t of X_t is infinitely divisible.

16. (*Lévy–Khinchine formula: special case*) Show that the characteristic function of the compound Poisson process may be represented in the form $\mathbb{E}e^{i\xi X_t} = \exp\{\int_{\mathbb{R}}^n (e^{i\xi y} - 1)\nu_t(dy)\}, \xi \in \mathbb{R}^n$, where $\nu_t(dy) = \lambda t Q(dy)$, and Q is the distribution of Y_1.

Chapter 6
The Kolmogorov–Chentsov Theorem and Sample Path Regularity

While constructions of probability distributions of stochastic processes indexed by uncountable parameter spaces, e.g., intervals, can be readily achieved via the Kolmogorov extension theorem, the regularity of the sample paths is not mathematically accessible in such constructions, for the simple reason that sample path properties that depend on uncountably many time points, e.g., continuity, do not define measurable events in the Kolmogorov model. In this chapter we consider a criterion, also due to Kolmogorov (Published in Slutsky (1937)) and Chentsov (1956) for Hölder continuous versions of processes and random fields. In addition to providing a tool for construction of k-dimensional Brownian motion, it yields the construction of continuous random fields such as the *Brownian sheet*. The chapter concludes with a demonstration of nowhere differentiability of the (continuous) Brownian paths.

Definition 6.1. A stochastic process (or random field) $Y = \{Y_u : u \in \Lambda\}$ is a *version* of $X = \{X_u : u \in \Lambda\}$ taking values in a metric space if Y has the same finite dimensional distributions as X.

Theorem 6.1 *(Kolmogorov–Chentsov Theorem).* Suppose $X = \{X_u : u \in \Lambda\}$ is a stochastic process (or random field) with values in a complete metric space (S, ρ), indexed by a bounded rectangle $\Lambda \subset \mathbb{R}^k$ and satisfying

$$\mathbb{E}\rho^{\alpha}(X_u, X_v) \le c|u - v|^{k+\beta}, \quad \text{for all } u, v \in \Lambda,$$

where c, α, and β are positive numbers. There is a version $Y = \{Y_u : u \in \Lambda\}$ of X, which is a.s. Hölder continuous of any exponent γ such that $0 < \gamma < \frac{\beta}{\alpha}$.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_6

Proof. Without essential loss of generality, we take $\Lambda = [0, 1]^k$ and the norm $|\cdot|$ to be the *maximum norm* given by $|u| = \max\{|u_i| : 1 \le i \le k\}$, $u = (u_1, \ldots, u_k)$. For each $N = 1, 2, \ldots$, let L_N be the finite lattice $\{j2^{-N} : j = 0, 1, \ldots 2^N\}^k$. Write $L = \cup_{N=1}^{\infty} L_N$. Define $M_N = \max\{\rho(X_u, X_v) : (u, v) \in L_N^2, |u - v| \le 2^{-N}\}$. Since (i) for a given $u \in L_N$, there are no more than 3^k points in L_N such that $|u - v| \le 2^{-N}$ ($v_i = u_i, u_i - 2^{-N}$, or $u_i + 2^{-N}$ for each i), (ii) there are $(2^N + 1)^k$ points in L_N, and (iii) for every given pair (u, v), the condition of the theorem holds, one has by Chebyshev's inequality that

$$P(M_N > 2^{-\gamma N}) \le c3^k (2^N + 1)^k (\frac{2^{-N(k+\beta)}}{2^{-\alpha\gamma N}}). \tag{6.1}$$

In particular, since $\gamma < \beta/\alpha$,

$$\sum_{N=1}^{\infty} P(M_N > 2^{-\gamma N}) < \infty. \tag{6.2}$$

Thus there is a random positive integer $N^* \equiv N^*(\omega)$ and a set Ω^* with $P(\Omega^*) = 1$, such that

$$M_N(\omega) \le 2^{-\gamma N} \quad \text{for all} N \ge N^*(\omega), \omega \in \Omega^*. \tag{6.3}$$

Fix $\omega \in \Omega^*$, and let $N \ge N^*(\omega)$. We will see by induction that, for all $m \ge N$, one has

$$\rho(X_u, X_v) \le 2 \sum_{j=N}^{m} 2^{-\gamma j}, \quad \text{for all } u, v \in L_m, |u - v| \le 2^{-N}. \tag{6.4}$$

For $m = N$, this follows from (6.3). Suppose, as an induction hypothesis, that (6.4) holds for $m = N, N + 1, \ldots, n$. Let $u, v \in L_{n+1}$, $|u - v| \le 2^{-N}$. Write $u = (i_1 2^{-n-1}, \ldots, i_k 2^{-n-1})$, $v = (j_1 2^{-n-1}, \ldots, j_k 2^{-n-1})$, where $i_\nu, j_\nu, 1 \le \nu \le k$, belong to $\{0, 1, 2, \ldots, 2^{n+1}\}$. We will find $u^*, v^* \in L_n$ such that $|u - u^*| \le 2^{-n-1}$, $|v - v^*| \le 2^{-n-1}$, and $|u^* - v^*| \le 2^{-N}$. For this, let the ν-th coordinate, say $i_\nu^* 2^{-n-1}$ of u^*, be the same as that of u if i_ν is even; and $i_\nu^* = i_\nu - 1$ if i_ν is odd and $i_\nu \ge j_\nu$, and $i_\nu^* = i_\nu + 1$ if i_ν is odd and $i_\nu < j_\nu$, $\nu = 1, \ldots, k$. Then $|u^* - u| \le 2^{-n-1}$, and $u^* \in L_n$ (since i_ν^* is even and $i_\nu^* 2^{-n-1} = (i_\nu^*/2)2^{-n}$). Similarly, define v^* with the roles of i_ν and j_ν interchanged, to get $v^* \in L_n$ and $|v - v^*| \le 2^{-n-1}$, with, moreover, $|u^* - v^*| \le |u - v| \le 2^{-N}$. Then by (6.3) and the induction hypothesis,

$$\rho(X_u, X_v) \le \rho(X_u, X_{u^*}) + \rho(X_{u^*}, X_{v^*}) + \rho(X_{u^*}, X_v)$$

$$\leq 2^{-\gamma(n+1)} + 2\sum_{\nu=N}^{n} 2^{-\gamma\nu} + 2^{-\gamma(n+1)} = 2\sum_{\nu=N}^{n+1} 2^{-\gamma\nu}, \tag{6.5}$$

completing the induction argument for (6.4), for all $\omega \in \Omega^*$, $m \geq N+1$, $N \geq N^*(\omega)$. Since $2\sum_{\nu=N}^{\infty} 2^{-\gamma\nu} = 2^{-\gamma(N-1)}(1-2^{-\gamma})^{-1}$, and $L = \cup_{m=N+1}^{\infty} L_m$, for all $N \geq N^*(\omega)$, it follows that

$$\begin{aligned} &\sup\{\rho(X_u, X_v) : u, v \in L, |u-v| \leq 2^{-N}\} \\ &= \sup\{\rho(X_u, X_v) : u, v \in \cup_{m=N+1}^{\infty} L_m, |u-v| \leq 2^{-N}\} \\ &\leq 2^{\gamma} 2^{-\gamma N} 2^{\gamma}(1-2^{-\gamma})^{-1}, \quad N \geq N^*(\omega), \omega \in \Omega^*. \end{aligned} \tag{6.6}$$

This proves that on Ω^*, $u \to X_u$ is uniformly continuous (from L into (S, ρ)) and is Hölder continuous with exponent γ. Now define $Y_u := X_u$ if $u \in L$ and otherwise $Y_u := \lim X_{u_N}$, with $u_N \in L$ and $u_N \to u$, if $u \notin L$. Because of uniform continuity of $u \to X_u$ on L (for $\omega \in \Omega^*$) and completeness of (S, ρ), the last limit is well-defined (Exercise 3). For all $\omega \notin \Omega^*$, let Y_u be a fixed element of S for all $u \in [0,1]^k$. Finally, letting $\gamma_j \uparrow \beta/\alpha$, $\gamma_j < \beta/\alpha$, $j \geq 1$, and denoting the exceptional set above as Ω_j^*, one has the Hölder continuity of Y for every $\gamma < \beta/\alpha$ on $\Omega^{**} := \cap_{j=1}^{\infty} \Omega_j^*$ with $P(\Omega^{**}) = 1$.

That Y is a version of X may be seen as follows. For any $r \geq 1$ and r vectors $u_1, \ldots, u_r \in [0,1]^k$, there exist $u_{jN} \in L$, $u_{jN} \to u_j$ as $N \to \infty$ $(1 \leq j \leq r)$. Then $(X_{u_{1N}}, \ldots, X_{u_{rN}}) = (Y_{u_{1N}}, \ldots, Y_{u_{rN}})$ a.s., and $(X_{u_{1N}}, \ldots, X_{u_{rN}}) \to (X_{u_1}, \ldots, X_{u_r})$ in probability, $(Y_{u_{1N}}, \ldots, Y_{u_{rN}}) \to (Y_{u_1}, \ldots, Y_{u_r})$ a.s. (Exercise 4). ■

An important consequence of Theorem 6.1 is the construction of Brownian motion defined in Example 5 of Chapter 1 (see also Exercise 1).

***Corollary 6.2** (Brownian Motion).* Let $X = \{X_t : t \geq 0\}$ be a real-valued Gaussian process defined on $(\Omega, \mathcal{F}, P)$, with $X_0 = 0$, $\mathbb{E}X_t = 0$, and $\mathrm{Cov}(X_s, X_t) = s \wedge t$, for all $s, t \geq 0$. Then X has a version $B = \{B_t : t \geq 0\}$ with continuous sample paths, which are Hölder continuous on every bounded interval with exponent γ for every $\gamma \in (0, \frac{1}{2})$.

Proof. Since $\mathbb{E}|X_t - X_s|^{2m} = c(m)(t-s)^m$, $0 \leq s \leq t$, for some constant $c(m)$, for every $m > 0$, the Kolmogorov–Chentsov Theorem 6.1 implies the existence of a version $B^{(0)} = \{B_t^{(0)} : 0 \leq t \leq 1\}$ with the desired properties on $[0,1]$. Let $B^{(n)}$, $n \geq 1$, be independent copies of $B^{(0)}$, independent of $B^{(0)}$. Define $B_t = B_t^{(0)}$, $0 \leq t \leq 1$, and $B_t = B_1^{(0)} + \cdots + B_1^{(n-1)} + B_{t-[t]}^{(n)}$, for $t \in [n, n+1)$, $n = 1, 2, \ldots$. ■

***Corollary 6.3** (Brownian Sheet).* Let $X = \{X_u : u \in [0,\infty)^2\}$ be a real-valued Gaussian random field satisfying $\mathbb{E}X_u = 0$, $\mathrm{Cov}(X_u, X_v) = (u_1 \wedge v_1)(u_2 \wedge v_2)$ for all $u = (u_1, u_2)$, $v = (v_1, v_2)$. Then X has a continuous version on $[0,\infty)^2$, which is Hölder continuous on every bounded rectangle contained in $[0,\infty)^2$ with exponent γ for every $\gamma \in (0, \frac{1}{2})$.

Proof. First let us note that on every compact rectangle $[0, M]^2$, $\mathbb{E}|X_u - X_v|^{2m} \le c(M)|u - v|^m$, for all $m = 1, 2, \ldots$. For this it is enough to check that on each horizontal line $u = (u_1, c)$, $0 \le u_1 < \infty$, X_u is a one-dimensional Brownian motion with mean zero and variance parameter $\sigma^2 = c$ for $c \ge 0$. The same holds on vertical lines. Hence

$$\begin{aligned}
&\mathbb{E}|X_{(u_1,u_2)} - X_{(v_1,v_2)}|^{2m} \\
&\quad \le 2^{2m-1}\big(\mathbb{E}|X_{(u_1,u_2)} - X_{(v_1,u_2)}|^{2m} + \mathbb{E}|X_{(v_1,u_2)} - X_{(v_1,v_2)}|^{2m}\big) \\
&\quad \le 2^{2m-1}c(m)\big(u_2^m|u_1 - v_1|^m + v_1^m|u_2 - v_2|^m\big) \\
&\quad \le 2^{2m-1}c(m)M^m 2|u - v|^m,
\end{aligned}$$

where $u = (u_1, u_2)$, $v = (v_1, v_2)$. ■

Remark 6.1. One may define the *Brownian sheet* on the index set $\Lambda_{\mathcal{R}}$ of all rectangles $R = [u, v)$, with $u = (u_1, u_2), v = (v_1, v_2)$, $0 \le u_i \le v_i < \infty$ $(i = 1, 2)$, by setting

$$X_R \equiv X_{[u,v)} := X_{(v_1,v_2)} - X_{(v_1,u_2)} - X_{(u_1,v_2)} + X_{(u_1,u_2)}. \tag{6.7}$$

Then X_R is Gaussian with mean zero and variance $|R|$, the area of R. Moreover, if R_1 and R_2 are non-overlapping rectangles, then X_{R_1} and X_{R_2} are independent. More generally, $\mathrm{Cov}(X_{R_1}, X_{R_2}) = |R_1 \cap R_2|$ (Exercise 5). Conversely, given a Gaussian family $\{X_R : R \in \Lambda_{\mathcal{R}}\}$ with these properties, one can restrict it to the class of rectangles $\{R = [0, u) : u = (u_1, u_2) \in [0,\infty)^2\}$ and identify this with the Brownian sheet in corollary 6.3. It is simple to check that for all n-tuples of rectangles $R_1, R_2, \ldots, R_n \subset [0,\infty)^2$, the matrix $((|R_i \cap R_j|))_{1\le i,j\le n}$ is symmetric and non-negative definite (Exercise 5). So the finite dimensional distributions of $\{X_R : R \in \Lambda_{\mathcal{R}}\}$ satisfy Kolmogorov's consistency condition.

The estimates derived in the proof of the Kolmogorov–Chentsov theorem easily yield the following useful result, which is made use of in a later chapter to prove a functional central limit theorem.

Proposition 6.4. Let $\Lambda \subset \mathbb{R}^k$ be a bounded rectangle, and let $X^{(n)} = \{X_u^{(n)} : u \in \Lambda\}$, $n \ge 1$, be a sequence of continuous processes with values in a complete metric space (S, ρ) satisfying

$$\mathbb{E}\rho^\alpha(X_u^{(n)}, X_v^{(n)}) \le c|u - v|^{k+\beta}, \quad \text{for all } u, v \in \Lambda, n \ge 1,$$

for some positive numbers c, α, β. Then, for every given $\epsilon > 0$ and $0 < \eta < 1$, there is a $\delta > 0$ such that

$$P(\sup\{\rho(X_u^{(n)}, X_v^{(n)}) : u, v \in \Lambda, |u - v| \leq \delta\} > \epsilon) < \eta, \quad \text{for all } n \geq 1.$$

Proof. Since the estimates obtained for the proof of Theorem 6.1 depend only on the constants c, α, and β, the asserted bound on the probability is proved in the same way as one would prove it for X in place of $X^{(n)}$. Specifically, given $\epsilon > 0$, $\eta < 1$, find $N(\eta)$ such that, for a given $\gamma \in (0, \beta/\alpha)$, $\sum_{N=N(\eta)}^{\infty} \theta(N) < \eta$, where $\theta(N)$ is the right side of the inequality (6.1). This provides the asserted probability bound with $(1 - 2^{-\gamma})^{-1} 2^{\gamma} 2^{-\gamma N(\eta)}$ in place of ϵ; recall the inequalities for the induction argument leading to uniform continuity on Ω^* in the proof of Theorem 6.1. If this last quantity is larger than ϵ, then find $N(\epsilon, \eta) \geq N(\eta)$ such that $(1 - 2^{-\gamma}) 2^{\gamma} 2^{-\gamma N(\epsilon,\eta)} \leq \epsilon$. Then the asserted bound holds with $\delta = 2^{-N(\epsilon,\eta)}$. ■

We conclude this chapter with some basic properties of multidimensional Brownian motions.

Definition 6.2. Let $\mathbf{D}$ be a symmetric non-negative definite $k \times k$ matrix. A *k-dimensional Brownian motion with drift $\boldsymbol{\mu}$ and diffusion coefficient matrix* $\mathbf{D}$ is a stochastic process $\{\mathbf{X}_t : t \geq 0\}$ with state space $\mathbb{R}^k$ having continuous sample paths and independent Gaussian increments with mean and covariance of an increment $\mathbf{X}_{t+s} - \mathbf{X}_s$ being $t\boldsymbol{\mu}$ and $t\mathbf{D}$, respectively. If $\mathbf{X}_0 = \mathbf{x}$, then this Brownian motion is said to *start at* $\mathbf{x}$. A Brownian motion starting at $\mathbf{0}$ with zero drift and diffusion coefficient $\mathbf{D} = \mathbf{I}$ is called the *standard k-dimensional Brownian motion* and denoted $\{\mathbf{B}_t : t \geq 0\}$. The standard Brownian motion started at $\mathbf{x}$ will be denoted $\mathbf{B}_t^{\mathbf{x}} := \mathbf{x} + \mathbf{B}_t, t \geq 0$.

Observe that since uncorrelated jointly distributed Gaussian random variables are independent, the k component processes of a k-dimensional standard Brownian motion are easily checked to be independent one-dimensional standard Brownian motions. Moreover, for a given drift $\boldsymbol{\mu} \in \mathbb{R}^k$ and non-negative definite symmetric matrix $\mathbf{D}$ with “square-root” $\mathbf{D}^{\frac{1}{2}}$, i.e. $\mathbf{D}^{\frac{1}{2}} \mathbf{D}^{\frac{1}{2}^t} = \mathbf{D}$, the process defined by

$$\mathbf{X}_t^{\mathbf{x}} := \mathbf{x} + \boldsymbol{\mu} t + \mathbf{D}^{\frac{1}{2}} \mathbf{B}_t, \quad t \geq 0, \tag{6.8}$$

is k-dimensional Brownian motion started at $\mathbf{x} \in \mathbb{R}^k$ with drift $\boldsymbol{\mu}$ and diffusion matrix $\mathbf{D}$.

Remark 6.2. A glimpse at a fundamentally important connection with analysis and partial differential equations can be observed by checking that, in the case that $\mathbf{D}$ is nonsingular, the pdf $y \to p(t; x, y)$ of X_t starting at $X_0 = x$ solves the heat equation

$$\frac{\partial p}{\partial t} = (2\pi)^{-\frac{k}{2}} (\det \mathbf{D})^{-\frac{1}{2}} \exp\{\sum_{i,j=1}^{k} (y_i - x_i - \mu_i t) c_{ij} (y_j - x_j - \mu_j t)\}$$

$$= \sum_{i=1}^{k} \mu_i \frac{\partial p}{\partial x_i} + \sum_{i,j=1}^{k} D_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j}$$

$$= -\sum_{i=1}^{k} \mu_i \frac{\partial p}{\partial y_i} + \sum_{i,j=1}^{k} D_{ij} \frac{\partial^2 p}{\partial y_i \partial y_j}, \tag{6.9}$$

where $\mathbf{D}^{-1} = ((c_{i,j}))$. As equations in either the so-called spatial *backward* (transition probability) variable x or the *forward* variable y, these two equations are also referred to as *Kolmogorov's backward and forward equations for Brownian motion*, respectively.

The following basic properties of Brownian motion are essentially direct consequences of the definition.

Theorem 6.5. Let $\mathbf{B}$ be a standard k–dimensional Brownian motion. Then

1. *(*Symmetry). $\mathbf{W}_t := -\mathbf{B}_t$, $t \geq 0$, is a standard Brownian motion.
2. *(Homogeneity and Independent Increments).* $\{\mathbf{B}_{t+s} - \mathbf{B}_s : t \geq 0\}$ is a standard Brownian motion independent of $\{\mathbf{B}_u : 0 \leq u \leq s\}$, for every $s \geq 0$,
3. *(Scale Change).* For every $\lambda > 0$, $\{\mathbf{B}_t^{(\lambda)} := \lambda^{-\frac{1}{2}}\mathbf{B}_{\lambda t} : t \geq 0\}$ is a standard Brownian motion.
4. *(Time-Inversion).* $\mathbf{W}_t := t\mathbf{B}_{1/t}$, $t > 0$, $\mathbf{W}_0 = 0$, is a standard Brownian motion.
5. *(Rotation Invariance).* Let $\mathbf{O}$ be a $k \times k$ orthogonal matrix (i.e., $\mathbf{O}\mathbf{O}^t = \mathbf{I}_k$, the $k \times k$ identity matrix). Then $\mathbf{O}\mathbf{B} = \{\mathbf{O}\mathbf{B}_t : t \geq 0\}$ is a standard k–dimensional Brownian motion.

Proof. With the exception of the property of rotation invariance, these properties are left as exercises.[1] For rotation invariance, first, note that since $\mathbf{x} \to \mathbf{O}\mathbf{x}$ is continuous, the paths of the composite map $t \to \mathbf{O}\mathbf{B}_t$ are continuous. Second, one may observe, e.g., using characteristic functions, that since $\mathbf{O}$ is a nonsingular matrix with determinant one, the distribution of the increments $\mathbf{O}\mathbf{B}_{t_{j+1}} - \mathbf{O}\mathbf{B}_{t_j} = \mathbf{O}(\mathbf{B}_{t_{j+1}} - _{t_j})$, $0 = t_0 < t_1 < t_2 \cdots < t_m$, remains independent and Gaussian. Moreover, for each $t >$), $\mathbf{O}\mathbf{B}_t$ has mean vector $\mathbb{E}\mathbf{O}\mathbf{B}_t = \mathbf{O}\mathbb{E}\mathbf{B}_t = \mathbf{O}\mathbf{0} = \mathbf{0}$, and variance–covariance matrix $\mathbb{E}(\mathbf{O}\mathbf{B}_t)(\mathbf{O}\mathbf{B}_t)^t = \mathbb{E}\mathbf{O}\mathbf{B}\mathbf{B}^t\mathbf{O}^t = \mathbf{O}t\mathbf{I}\mathbf{O}^{\mathrm{tr}} = t\mathbf{I}$. ■

Remark 6.3. The continuous but otherwise highly irregular behavior of the sample paths of observed particle suspensions did not go unnoticed by the experimentalists who set out to document the validity of Einstein's model. Most notable among these efforts in the early twentieth century were those of Perrin whose work on this problem led to experimental determination of Avogadro's constant. In his research monograph,[2] Perrin exclaims: "The trajectories are confused and complicated so

[1]The proofs are given in BCPT, p. 141–142, as well.

[2]Perrin (1913) is the original French publication; the quote here is taken from the English translation, Perrin (1929).

often and so rapidly that it is impossible to follow them; the trajectory actually measured is very much simpler and shorter than the real one. Similarly, the apparent mean speed of a grain during a given time varies *in the wildest way* in magnitude and direction, and does not tend to a limit as the time taken for an observation decreases, as may be easily shown by noting, in the camera lucida,[3] the positions occupied by a grain from minute to minute, and then every five seconds, or, better still, by photographing them every twentieth of a second, as has been done by Victor Henri Comandon, and de Broglie when kinematographing the movement. It is impossible to fix a tangent, even approximately, at any point on a trajectory, and we are thus reminded of the continuous underived functions of the mathematicians." That such sample path behavior is intrinsic to the definition of Brownian motion is mathematically confirmed by the fact that, as we will see below, with probability one, the Brownian paths (in one-dimension) are of unbounded variation in every non-degenerate interval.

Observe that the sample paths $t \rightarrow B_t(\omega)$ of the one-dimensional Brownian motion $B^x := \{B_t : t \geq 0\}$ starting at x are required by definition to be elements of the space $C[0, \infty)$ of continuous real-valued functions on $[0, \infty)$. The *distribution* of this process is a probability measure $Q_x = P \circ B^{x\,-1}$ on $C[0, \infty)$ induced by the map $h(\omega) = \{B_t(\omega) : t \geq 0\}$, $\omega \in \Omega$. That is, letting $\mathcal{G}$ be the σ-field of $C[0, \infty)$ generated by finite dimensional subsets of the form $F = \{x \in C[0, \infty) : x(t_i) \in C_i,\ i = 1, \ldots, m\}$, $0 < t_1 < t_2 < \cdots < t_m$, $C_i \in \mathcal{B}$, (Borel σ-field $\mathcal{B}$ on $\mathbb{R}$) $1 \leq i \leq m$, Q_x is uniquely specified by the finite dimensional (path) probabilities

$$Q_x(F) = P(B_{t_i} \in C_i,\ i = 1, \ldots, m). \tag{6.10}$$

Accordingly, the distribution $Q = Q_0$ of standard Brownian motion started at 0 is a probability on the path space $C[0, \infty)$, referred to as *Wiener measure*.

Similarly the distribution of Brownian motion on a finite interval $[0, T]$ may be viewed as a probability measure on the space $C[0, T]$ induced by the map $\omega \rightarrow (B_t(\omega) : 0 \leq t \leq T)$.

From the point of view of weak convergence theory to be applied in a later chapter, it will be useful to view the induced distributions as probabilities on metric spaces $C[0, \infty)$ and $C[0, T]$. A metric for $C[0, T]$ is $d_T(x, y) = \max_{0 \leq t \leq T} |x(t) - y(t)|$ and for $C[0, \infty)$ is $d(x, y) = \sum_{N=1}^{\infty} 2^{-N} \frac{d_N(x,y)}{1+d_N(x,y)}$. Convergence in the latter metric $d(x, y)$ means uniform convergence on compact subintervals $[0, N]$ for all $N \geq 0$. The Borel σ-fields for these metrics coincide with the σ-fields generated by finite dimensional events (Exercise 9).

We close this chapter with a basic mathematical development to complement the empirical observations described in Remark 6.3. Let $\{B_t : t \geq 0\}$ denote a one-dimensional standard Brownian motion starting at zero.

[3]The experiment was repeated with modern upgrades of the experimental apparatus (camera) in Newburgh et al. (2006).

Proposition 6.6. Define

$$V_n = \sum_{i=1}^{2^n} |B_{i/2^n} - B_{(i-1)/2^n}|.$$

Then,

1. $\mathbb{E}V_n = 2^{n/2}\mathbb{E}|B_1|$.
2. $\mathrm{Var}\, V_n = \mathrm{Var}\,|B_1| = 1$.
3. With probability one, $\{B_t : t \geq 0\}$ is of unbounded variation on $0 \leq t \leq 1$.
4. Outside a set of probability zero, every Brownian path ω is of unbounded variation on every non-degenerate interval $[a, b]$, $0 \leq a < b \leq 1$, where a and b may depend on ω.

Proof. The calculation $\mathbb{E}V_n = 2^{n/2}\mathbb{E}|B_1| = 2^{n/2+1}$ follows immediately from stationary increments and scaling properties, and $\mathrm{Var}\, V_n = \mathrm{Var}\,|B_1| = 1$ follows from independent increments and scaling properties. Since the partition into intervals of length $2^{-(n+1)}$ is a refinement of the partition into intervals of length 2^{-n}, one has $V_{n+1} \geq V_n$, $n \geq 1$. Thus $\lim_{n\to\infty} V_n$ exists but may be infinite, almost surely. Using Chebyshev's inequality, one has $P(|V_n - \mathbb{E}V_n| > n) \leq n^{-2}$. Thus, by the Borel–Cantelli lemma I, $P(V_n > 2^{\frac{n}{2+1}} - n\ i.o.) = 0$. It follows that $V_n \to \infty$ with probability one. By scaling applied to all intervals with a, b rational, it therefore follows that outside a set of probability zero, Brownian paths are of unbounded variation on every non-degenerate such interval $[a, b]$, $0 \leq a < b \leq t, t > 0$. Property 4 follows from this (Exercise 10). ■

Exercises

1. Prove that the stochastic process constructed in Corollary 6.2 has independent mean zero Gaussian increments over disjoint time intervals, with the variance of the increment $B_{t+s} - B_t$ being s.
2. Show that positivity of β is necessary for the Kolmogorov–Chentsov theorem by considering the Poisson process.
3. Let f be a uniformly continuous function defined on a dense subset D of a metric space (Λ, d) into a complete metric space (S, ρ).
   (a) Prove that f has a unique extension as a (uniformly) continuous function from Λ into S.
   (b) If f is Hölder continuous of exponent γ on S, show that this is true of its extension to Λ as well.
4. Amplify the last sentence in the proof of the Kolmogorov–Chentsov Theorem 6.1. [*Hint*: By Chebyshev's inequality, the condition of the theorem implies that for any $\epsilon > 0$, $P\big(\rho(X_{u_{jN}}, X_{u_j}) > \epsilon\big) \to 0$, as $N \to \infty$, for

all $j = 1, 2, \ldots, m$. The convergence $Y_{u_{jN}} \to Y_{u_j}$ follows from the well-definedness of the limit.]

5. In the context of Remark 6.1, show that (a) $((|R_i \cap R_j|))_{1 \le i,j \le n}$ is a non-negative definite $n \times n$ matrix. [*Hint*: Note that $\sum_{1 \le i,j \le n} c_i c_j |R_i \cap R_j| = \int_{[0,\infty)^2} \left(\sum c_i \mathbf{1}_{R_i}(x)\right)^2 dx$.] (b) Complete the construction of Brownian sheet on $\Lambda = [0, \infty)^2$.

6. Let $\{B_t : t \ge 0\}$ denote standard Brownian motion starting at 0. Verify that $(B_{t_1}, \ldots, B_{t_m})$ has an m-dimensional Gaussian distribution and calculate the mean and the variance–covariance matrix using the fact that the Brownian motion has independent Gaussian increments.
7. Let $\{X_t : t \ge 0\}$ be a real-valued stochastic process with stationary and independent increments starting at 0 with $\mathbb{E}X_s^2 < \infty$ for $s > 0$. Assume $\mathbb{E}X_t$ and $\mathbb{E}X_t^2$ are continuous functions of t.

   (i) Show that $\mathbb{E}X_t = mt$ for some constant m.
   (ii) Show that $\operatorname{Var} X_t = \sigma^2 t$ for some constant $\sigma^2 \ge 0$. Also $\mathbb{E}X_t^2 = \sigma^2 t + m^2 t^2$ is linear if and only if $m = 0$.

8. Let $\{\mathbf{X}_t^{\mathbf{x}} : t \ge 0\}$ be the k-dimensional Brownian motion with drift $\boldsymbol{\mu}$ and diffusion coefficient matrix $\mathbf{D}$.

   (i) Calculate the mean of $\mathbf{X}_t^{\mathbf{x}}$ and the variance–covariance matrix of $\mathbf{X}_t^{\mathbf{x}}$.
   (ii) In the case $\boldsymbol{\mu} = 0, \mathbf{D} = \mathbb{I}$ of standard Brownian motion, show that for each fixed $t > 0$, $\frac{\mathbf{X}_t}{||\mathbf{X}_t||}$ is uniformly distributed over the k-dimensional sphere. [*Hint*: Recall the rotation invariance property of standard Brownian motion.]

9. (i) For $T > 0$, let $\mathcal{G}_T$ denote the σ–field of subsets of $C[0, T]$ generated by finite dimensional events of the form

$$F = \{x \in C[0, T] : x(t_i) \in B_i, \quad i = 1, \ldots, m\},$$

   where $0 < t_1 < t_2 < \cdots < t_m \le T$, $B_i \in \mathcal{B}$. Show that $\mathcal{G}_T$ coincides with the Borel σ–field on the metric space $C[0, T]$ for the metric $d_T(x, y) = \max_{0 \le t \le T} |x(t) - y(t)|$.
   (ii) Let $\mathcal{G}$ be the σ–field of subsets of $C[0, \infty)$ generated by events of the form

$$F = \{x \in C[0, \infty) : x(t_i) \in B_i, \quad i = 1, \ldots, m\},$$

   where $0 < t_1 < t_2 < \cdots < t_m$, $B_i \in \mathcal{B}$. Show that $\mathcal{G}$ coincides with the Borel σ–field on the metric space $C[0, \infty)$ for the metric $d(x, y) = \sum_{N=1}^{\infty} 2^{-N} \frac{d_N(x,y)}{1+d_N(x,y)}$.
10. Let f be a real valued continuous function on $[c, d]$, $c < d$. Define the *variation* $v(f)$ of f on $[c, d]$ by $v(f) = \sup v(f : \pi_k)$, where the supremum is over all partitions π_k of $[c, d]$ of the form $c = a_0 < a_1 < \ldots a_k = d, k \ge 1$,

and $v(\pi_k) = \sum_{i=1}^{k} \mid f(a_i) - f(a_{i-1}) \mid$. Also, let $V_n(f) = \sum_{j=1}^{2^n} \mid f(c + (d-c)j2^{-n}) - f(c + (d-c)j2^{-n}) - f(c + (j-1)(d-c)2^{-n}) \mid$, $V(f) = \lim_{n-\infty} V_n(f)$. Show that $v(f) = V(f)$. [Hint: Fix $\epsilon > 0$. For any given π_k of the above form, there exists $\delta(\epsilon) > 0$ such that $|f(x) - f(y)| < \frac{\epsilon}{2k}$ if $|x - y| \leq \delta(\epsilon)$. Find $n = n(\epsilon)$ for which each a_i is within a distance $\delta(\epsilon)$ from a point of the form $c + (d-c)j2^{-n}(j = 0, 1, \ldots, 2^{-n})$. Then $V_n(f) \geq v(f : \pi_k) - \epsilon$, so that $V(f) \geq v(f) - \epsilon$.]

Chapter 7
Random Walk, Brownian Motion, and the Strong Markov Property

In this chapter the strong Markov property is derived as an extension of the Markov property to certain random times, called stopping times. A number of consequences of the strong Markov property of Brownian motion and the simple random walk are derived. A derivation of the law of the iterated logarithm for Brownian motion is included in this chapter, from which some fine scale sample path properties of Brownian motion are derived as well.

Discrete parameter Markov processes on general state spaces were introduced in Chapter 1. With this as background, let us turn to a strengthened and more useful version of the (homogeneous) Markov property. For this let P_z denote the distribution of a Markov process $X = \{X_n : n \geq 0\}$, i.e., a probability on the product space $(S^\infty, \mathcal{S}^{\otimes\infty})$, with transition probability $p(x, dy)$ and initial distribution $P(X_0 = z) = 1$. Also an expected value of a real-valued function (on S^∞) with respect to P_z is denoted $\mathbb{E}_z$, $z \in S$.

Definition 7.1. Fix $m \geq 0$. The *after-m* (future) process is defined by $X_m^+ := \{X_{n+m} : n \geq 0\}$.

Definition 7.2 *(Markov Property).* We say that $X = \{X_n : n \geq 0\}$ has the *(homogeneous) Markov Property* if for every $m \geq 0$, the conditional distribution of X_m^+, given the $\sigma-$field $\mathcal{F}_m = \sigma\{X_n : n \leq m\}$, is P_{X_m}, i.e., equals P_y on the set $[X_m = y]$.

Suppose now that $\{Z_n : n \geq 1\}$ is an i.i.d. $\mathbb{R}^k$-valued sequence. Let Z_0 be an $\mathbb{R}^k$-valued random variable, independent of $\{Z_n : n \geq 1\}$. Recall that the stochastic process $S_0 = Z_0$, $S_n = Z_0 + Z_1 + \cdots + Z_n$, $n \geq 1$, defines a *(general) random walk* on $\mathbb{R}^k$. As above, let $\mathcal{F}_m = \sigma(S_j : j = 0, 1, \ldots m)$, $m \geq 0$.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_7

Remark 7.1. As noted earlier, we write $\mathbb{E}_x$ for expected values with respect to P_x. Write $S = \mathbb{R}^k$, and let $\mathcal{S}$ be the Borel σ-field of $\mathbb{R}^k$. Also let $S^\infty = (\mathbb{R}^k)^{\mathbb{Z}_+}$. One proof of the following result may be given by showing that the Markov property of Definition 1.3 of Chapter 1, which may be called the *ordinary Markov property*, implies the strengthened version given by Definition 7.2 (see Chapter 1, Exercise 1). On the other hand, a direct proof is given below.

Proposition 7.1 *(Markov Property of the Random Walk).* A random walk has the Markov property of Definition 7.2.

Proof. Let f be an arbitrary real-valued bounded measurable function on $\tilde{\Omega} = S^\infty$, with σ-field $\mathcal{S}^{\otimes\infty}$. One may express $f(S_m^+) \equiv f(S_m, S_m + Z_{m+1}, S_m + Z_{m+1} + Z_{m+2}, \dots)$ as $\psi(U, V)$, where $U = S_m$, $V = (Z_{m+1}, Z_{m+2}, \dots)$, and ψ is a real-valued function defined by $\psi(x_0, z_1, z_2, \dots) = f(x_0, x_0 + z_1, x_0 + z_1 + z_2, \dots)$. Let $\mathcal{G} = \mathcal{F}_m = \sigma(S_j : 0 \le j \le m)$. Since V is independent of $\mathcal{G}$, it follows from the substitution property for conditional expectations that

$$\mathbb{E}\big[f(S_m^+)|\mathcal{G}\big] \equiv \mathbb{E}\big[\psi(U, V)|\mathcal{G}\big] = h(U) \equiv h(S_m),$$

where $h(y) = \mathbb{E}\psi(y, V) \equiv \mathbb{E}f(y, y + Z_{m+1}, y + Z_{m+1} + Z_{m+2}, \dots) = \mathbb{E}f(y, y + Z_1, y + Z_1 + Z_2, \dots) = \mathbb{E}_y f$. ■

Our next task is to further strengthen the Markov property by introducing an extremely useful concept of a *stopping time*, sometimes also called a *Markov time*. Consider a sequence of random variables $\{X_n : n = 0, 1, \dots\}$, defined on some probability space $(\Omega, \mathcal{F}, P)$. Stopping times with respect to $\{X_n\}_{n=0}^\infty$ are defined as follows. Denote by $\mathcal{F}_n$ the $\sigma-$field $\sigma\{X_0, \dots, X_n\}$ comprising all events in $\mathcal{F}$ that depend only on the random variables $\{X_0, X_1, \dots, X_n\}$.

Definition 7.3. A *stopping time* τ *for the process* $\{X_n\}_{n=0}^\infty$ is a random variable taking non-negative integer values, including possibly the value $+\infty$, such that

$$[\tau \le n] \in \mathcal{F}_n \quad (n = 0, 1, \dots). \tag{7.1}$$

Observe that (7.1) is equivalent to the condition

$$[\tau = n] \in \mathcal{F}_n \quad (n = 0, 1, \dots), \tag{7.2}$$

since $\mathcal{F}_n$ are increasing $\sigma-$fields (i.e., $\mathcal{F}_n \subset \mathcal{F}_{n+1}$) and τ is integer-valued.

Informally, (7.1) says that, using τ, the decision to stop or not to stop by time n depends only on the observations $X_0, X_1, \dots, X_n$. An important example of a stopping time is the *first passage time* τ_B to a (Borel) set $B \subset \mathbb{R}^1$,

$$\tau_B := \inf\{n \ge 0 : X_n \in B\}. \tag{7.3}$$

If X_n does not lie in B for any n, one takes $\tau_B = \infty$. Sometimes $[X_0 \in B] \neq \emptyset$ and the infimum in (7.3) is taken over $\{n \geq 1 : X_n \in B\}$, in which case we call it the *first return time* to B, denoted η_B. A less interesting but useful example of a stopping time is a *constant time* $\tau := m$ where m is a fixed positive integer. It is also useful to make note that if τ_1, τ_2 are stopping times then so is the arithmetic sum $\tau_1 + \tau_2$, as well as $\tau_1 \wedge \tau_2 := \max\{\tau_1, \tau_2\}$, and $\tau_1 \vee \tau_2 := \min\{\tau_1, \tau_2\}$ (Exercise 2).

One may define, for every positive integer r, the rth *passage time* $\tau_B^{(r)}$ to B recursively, by

$$\tau_B^{(1)} := \tau_B, \quad \tau_B^{(r)} := \inf\{n > \tau_B^{(r-1)} : X_n \in B\} \quad (r = 2, \ldots). \tag{7.4}$$

Again, if X_n does not lie in B for any $n > \tau_B^{(r-1)}$, take $\tau_B^{(r)} = \infty$. Also note that if $\tau_B^{(r)} = \infty$ for some r, then $\tau_B^{(r')} = \infty$ for all $r' \geq r$. It is a simple exercise to check that each $\tau_B^{(r)}$ is a stopping time (Exercise 1).

Definition 7.4. Given a stopping time τ, the *pre-τ σ−field* $\mathcal{F}_\tau$ is defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap [\tau = m] \in \mathcal{F}_m, \forall m \geq 0\}. \tag{7.5}$$

The *after-τ process* $X_\tau^+ = \{X_\tau, X_{\tau+1}, X_{\tau+2}, \ldots\}$ is well-defined on the set $[\tau < \infty]$ by $X_\tau^+ = X_m^+$ on $[\tau = m]$.

Remark 7.2. $\mathcal{F}_\tau$ is determined by the value of τ and $X_0, X_1, \ldots, X_\tau$ on the set $[\tau < \infty]$. For if $\tau = m$, then $A \cap [\tau = m] \in \sigma(X_j : 0 \leq j \leq m)$.

Theorem 7.2 ***(****Strong Markov Property*[1]***).*** Let τ be a stopping time for the process $\{X_n : n \geq 0\}$. If this process has the Markov property of Definition 7.2, then, on $[\tau < \infty]$, the conditional distribution of the after-τ process X_τ^+, given the pre-τ σ−field $\mathcal{F}_\tau$, is P_{X_τ}.

Proof. Let f be a real-valued bounded measurable function on $(S^\infty, \mathcal{S}^{\otimes\infty})$, and let $A \in \mathcal{F}_\tau$. Then

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_{[\tau<\infty]}\mathbf{1}_A f(X_\tau^+)) &= \sum_{m=0}^{\infty} \mathbb{E}(\mathbf{1}_{[\tau=m]}\mathbf{1}_A f(X_m^+)) \\
&= \sum_{m=0}^{\infty} \mathbb{E}(\mathbf{1}_{[\tau=m]\cap A}\mathbb{E}_{X_m} f) \\
&= \sum_{m=0}^{\infty} \mathbb{E}(\mathbf{1}_{[\tau=m]\cap A}\mathbb{E}_{X_\tau} f) = \mathbb{E}(\mathbf{1}_{[\tau<\infty]}\mathbf{1}_A\mathbb{E}_{X_\tau} f).
\end{aligned} \tag{7.6}$$

[1] Although informally used by some authors earlier, the precise derivation of the strong Markov property is due independently to Dynkin and Yushekivic (1956), and Blumenthal (1957).

The second equality follows from the Markov property for random walk proven above since $A \cap [\tau = m] \in \mathcal{F}_m$. Since f is an arbitrary bounded Borel measurable function it follows that the conditional distribution of X_τ^+ given $\mathcal{F}_\tau$ is given by P_{X_τ}. ■

With this and Proposition 7.1 one has the following.

Corollary 7.3. The random walk has the strong Markov property.

The following example illustrates the utility of the strong Markov property in a standard calculation. The result was derived in Chapter 2, Corollary 2.4, using Proposition 2.3. Here we present a more detailed proof. See Exercise 6 for another application for simple random walk.

***Example 1** (Recurrence of Simple Symmetric Random Walk).* Consider the simple symmetric random walk $S_0^+ = \{S_n : n \geq 0\}$ on $\mathbb{Z}$ started at x. Let $\tau_y := \inf\{n : S_n = y\}$ for $y \in \mathbb{Z}$. Suppose one wishes to prove that $P_x(\tau_y < \infty) = 1$ for $y \in \mathbb{Z}$. Let $a < x < y$. By the Markov property,

$$\begin{aligned}
\varphi(x) &:= P_x(S_0^+ \text{ reaches } y \text{ before } a) \\
&= P_x(S_1^+ \text{reaches } y \text{ before } a) \\
&= \mathbb{E}_x P_x(S_1^+ \text{reaches } y \text{ before } a) | \sigma(S_1)) \\
&= \mathbb{E}_x \varphi(S_1) \\
&= \mathbb{E}_x \varphi(x + Z_1); \text{ Definition 7.2} \\
&= \frac{1}{2}\varphi(x+1) + \frac{1}{2}\varphi(x-1) \qquad (7.7)
\end{aligned}$$

with boundary values $\varphi(y) = 1$, $\varphi(a) = 0$. Solving one obtains $\varphi(x) = (x-a)/(y-a)$. Thus, for every $x < y$, $P_x(\tau_y < \infty) = 1$ follows by letting $a \to -\infty$ using basic "continuity properties" of probability measures. Similarly, letting $y \to \infty$ in $1 - \varphi(x)$, one gets $P_x(\tau_a < \infty) = 1$ for all $a < x$. Finally, $P_x(\tau_x^{(1)} < \infty) = 1$ is shown by conditioning on S_1 (Exercise 3). Hence $P_x(\tau_y^{(1)} < \infty) = 1$ for all x, y. While this calculation only required the Markov property, next consider the problem of showing that the process will return to y infinitely often. One would like to argue that conditioning on the process up to its return to y, it merely starts over. This of course is the strong Markov property. So let us examine carefully the calculation to show that the $r-$th passage time to y, $\tau_y^{(r)} < \infty$ a.s. for every $r = 1, 2, \ldots$. Letting $\tau_y^{(2)} := \inf\{n \geq 1 : (S_{\tau_y^{(1)}}^+)_n = y\}$ denote the first return time to y of the process $S_{\tau_y^{(1)}}^+$, one has

$$\begin{aligned} P_x(\tau_y^{(2)} < \infty) &= \mathbb{E}_x[P_x(\tau_y^{(2)} < \infty | \mathcal{F}_{\tau_y^{(1)}})] \\ &= \mathbb{E}_x P_z(\tau_y < \infty)|_{z=S_{\tau_y^{(1)}}=y} \text{ (strong Markov property)} \\ &= \mathbb{E}_x P_y(\tau_y^{(1)} < \infty) = 1. \end{aligned} \tag{7.8}$$

Similarly, $S_{\tau_y^{(1)}} = y$ a.s.

Now this argument remains valid if one replaces $\tau_y^{(1)}$ by $\tau_y^{(r-1)}$ and $\tau_y^{(2)}$ by $\tau_y^{(r)}$ and assumes that $\tau_y^{(r-1)} < \infty$ almost surely. Hence, by induction, $P_x(\tau_y^{(r)} < \infty) = 1$ for all positive integers r. This is equivalent to the recurrence of the state y in the sense that

$$P_x(S_n = y \text{ for infinitely many } n) = P_x(\cap_{r=1}^{\infty}[\tau_y^{(r)} < \infty]) = 1.$$

Remark 7.3. The Markov property in Definition 7.2 is sometimes defined with respect to a *filtration*, i.e., an increasing family of $\sigma-$fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$, such that $\sigma(X_j : 0 \le j \le m) \subset \mathcal{F}_m$ for all m. The strong Markov property expressed by Theorem 7.2 then holds with respect to the $\sigma-$fields $\mathcal{F}_\tau$ as defined in (7.2), with this filtration $\{\mathcal{F}_m : m \ge 0\}$ in place of $\{\sigma(X_j : 0 \le j \le m) : m \ge 0\}$. For example, one may take $\mathcal{F}_m$ to be the $P-$completion[2] of $\sigma(X_j : 0 \le j \le m)$, for $m \ge 0$. Another example is $\mathcal{F}_m = \sigma(X_j : 0 \le j \le m) \vee \mathcal{G}$, where $\mathcal{G} \subset \mathcal{F}$ is a $\sigma-$field independent of $\{X_j : j \ge 0\}$.

Let $\{X_t : t \in \Lambda\}$ be a stochastic process defined on $(\Omega, \mathcal{F}, P)$ with parameter space $\Lambda = [0, T]$, or $[0, \infty)$, having state space $(S, \mathcal{S})$, where S is a Polish space and $\mathcal{S}$ is its Borel σ-field. For present purposes it is enough to assume that (1) the map $\omega \to X_t(\omega), t \in \Lambda$, is measurable on Ω into a space Γ of *right-continuous* functions with a σ-field $\mathcal{C}$, (2) the set C of continuous functions on Λ into S is a subset of Γ and the Borel σ-field of C is contained in $\mathcal{C}$, and (3) the finite-dimensional projections $\gamma \to (\gamma(t_1), \ldots, \gamma(t_n))$ on Γ into $(S^{\{t_1,\ldots,t_n\}}, \mathcal{S}^{\otimes\{t_1,\ldots,t_n\}})$, determine $\mathcal{C}$, i.e., $\sigma(\cup \mathcal{S}^{\otimes\{t_1,\ldots,t_n\}} : 0 < t_1 < t_2 < \cdots < t_n, n \ge 1) = \mathcal{C}$.

Remark 7.4. One may, in particular, take Γ to be the *Skorokhod space* [3] of all right-continuous functions with left-hand limits (cadlag in French), which is a Polish space, with $\mathcal{C}$ its Borel σ-field. It is known that C is a closed subset of Γ and the relative topology of C is the uniform topology (on compact subsets of Λ).

---

[2]See BCPT p. 225, for the measure-theoretic completion of σ-fields. In particular, completeness can always be achieved and there is no loss in generality in assuming that the underlying probability space $(\Omega, \mathcal{F}, P)$ is complete from the outset.

[3]Skorokhod (1956). In the same issue, the paper Kolmogorov (1956) introduced an equivalent metric which also makes it a complete metric space A detailed account of the Skorokhod topology is given in Billingsley (1968).

Suppose now that $\{X_t : t \in \Lambda\}$ has the Markov property as stated in Definition 1.4 of Chapter 1. The following is a strengthened version of that property, most aptly described on the parameter space $\Lambda = [0, \infty)$. For the statement let $\mathcal{F}_s = \sigma(X_u : 0 \le u \le s)$, $X_s^+ = (X_{s+u} : u \ge 0)$, i.e., the *after-s process*. Let P_x denote the distribution of the process $\{X_t : t \in \Lambda\}$ when $X_0 = x$. Note that P_x is the probability measure on $(\Gamma, \mathcal{C})$ induced by the map $\omega \to X_t(\omega), t \in \Lambda$, with $X_0(\omega) = x$, for P-a.s. all $\omega \in \Omega$. More generally, P_μ is the distribution of the process when X_0 has (initial) distribution μ. Unless otherwise specified, we will assume that the Markov process $\{X_t : 0 \le t < \infty\}$ is *time-homogeneous*, i.e., the conditional distribution of X_{s+t} given $\mathcal{F}_s$, is $p(t; x, dy)$ on $[X_s = x]$. Note also that the Markov property in Definition 1.4 refers to the collection of probability measures $\{P_\mu : \mu$ is a probability measure on$(S, \mathcal{S})\}$, including all $P_x = P_{\delta_x}, x \in S$. As a convention, the process $\{X_t : t \ge 0\}$ is generally referred to as a Markov process whatever be the initial state X_0, unless one is specified.

Proposition 7.4 *(Markov Property for Right-Continuous Stochastic Processes)***.** Suppose that $\{X_t : 0 \le t < \infty\}$ is a stochastic process satisfying conditions (1)–(3) of the paragraph before Remark 7.4 above, with the Markov property of Definition 1.4 (Chapter 1). Then the conditional distribution of X_s^+ given $\mathcal{F}_s$ is P_{X_s}, i.e., P_x on $[X_s = x]$.

Proof. Let $0 \le s_1 < s_1 < \cdots < s_m = s$, and $t_j = s + u_1 + \cdots + u_j$, $1 \le j \le n$, for $u_i > 0, 1 \le i \le n$. We first show that the conditional distribution of $(X_{t_1}, \ldots, X_{t_n})$ given $\sigma(X_{s_1}, X_{s_2}, \ldots, X_{s_m})$ is the P_{X_s}-distribution of $(X_{u_1}, X_{u_1+u_2}, \ldots, X_{u_1+u_2+\cdots+u_n})$, namely, $p(u_1; x, dy_1)p(u_2; y_1, dy_2) \cdots p(u_n; y_{n-1}, dy_n)$. But this follows essentially as in the proof for discrete parameter case (Chapter 1), replacing $p(x, dy)$ by $p(u_1; x, dy_1)$, $p(u_2; y_1, dy_2)$, $\ldots, p(u_n; y_{n-1}, dy_n)$, successively (Exercise 7). By the property (3) above, this proves the assertion (Exercise 7). ■

Corollary 7.5 *(Markov Property of Brownian Motion)***.** The conditional distribution of $(B_s)^+ = \{B_{s+t} : t \ge 0\}$ given $\mathcal{F}_s$ is P_{B_s}.

Proof. Although the strengthened Markov property of Brownian motion on $\mathbb{R}^k$, with arbitrary drift parameter μ and diffusion matrix Σ, is an immediate consequence of Proposition 7.4, we give a direct proof here due to its importance. Without loss of generality, first consider the standardized case $\mu = 0$, $\Sigma = \mathbf{1}_k$, the k-dimensional identity matrix. The general case follows easily from the standardized case (Exercise 7). We can mimic the proof of Proposition 7.1. Let f be a real-valued bounded measurable function on $C([0, \infty) : \mathbb{R}^k)$. Then $\mathbb{E} f\big((B_s)^+ | \mathcal{F}_s\big) = \mathbb{E}\big(\psi(U, V) | \mathcal{F}_s\big)$, where $U = B_s$, $V = \{B_{s+t} - B_s : t \ge 0\}$, $\psi(y, \omega) := f(\omega_y)$, $y \in \mathbb{R}$, $\omega \in C[0, \infty)$, and $\omega_y \in C[0, \infty)$ by $\omega_y(t) = \omega(t) + y$. By the substitution property for conditional expectation, one has

$$\mathbb{E}\big(\psi(U, V) | \mathcal{F}_s\big) = h(U) = h(B_s),$$

where

$$h(y) = \mathbb{E}\psi(y, V) = \mathbb{E}\psi(y, \{B_t : t \geq 0\}) = \mathbb{E}f(B + y) = \int_{C[0,\infty)} f\,dP_y.$$

■

As will be illustrated by examples in this chapter, it is sometimes useful to extend the definition of standard Brownian motion as follows.

Definition 7.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_t, t \geq 0$, a filtration. The *$k-$dimensional standard Brownian motion with respect to this filtration* is a stochastic process $\{B_t : t \geq 0\}$ on $(\Omega, \mathcal{F}, P)$ having (i) stationary, independent Gaussian increments $B_{t+s} - B_s$ with mean zero and covariance matrix $(t-s)I_k$; (ii) a.s. continuous sample paths $t \to B_t$ on $[0, \infty) \to \mathbb{R}^k$; and (iii) for each $t \geq 0$, B_t is $\mathcal{F}_t$-measurable and $B_t - B_s$ is independent of $\mathcal{F}_s$, $0 \leq s < t$. Taking $B_0 = 0$ a.s., then $B^x := \{x + B_t : t \geq 0\}$, is referred to as the *standard Brownian motion started at* $x \in \mathbb{R}^k$ (with respect to the given filtration).

For example, one may take the completion $\mathcal{F}_t = \overline{\sigma}(B_s : s \leq t), t \geq 0$, of the $\sigma-$field generated by the coordinate projections $t \to \omega(t), \omega \in C[0, \infty)$. Alternatively, one may have occasion to use $\mathcal{F}_t = \sigma(B_s, s \leq t) \vee \mathcal{G}$ where $\mathcal{G}$ is some $\sigma-$field independent of $\mathcal{G}$. Recall also the right-continuous filtration $\mathcal{F}_{t+}, t \geq 0$, introduced in the previous chapter in connection with first passage times. The definition of the Markov property can be modified accordingly as follows.

Proposition 7.6. The Markov property of Brownian motions B^x on $\mathbb{R}^k$ defined on $(\Omega, \mathcal{F}, P)$ holds with respect to the filtration

$$\mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\epsilon}, \qquad (t \geq 0), \tag{7.9}$$

where $\mathcal{F}_t = \mathcal{G}_t := \sigma(B_u : 0 \leq u \leq t)$, or $\mathcal{F}_t$ is the P–completion of $\mathcal{G}_t$.

Proof. It is enough to prove that $B_{t+s} - B_s$ is independent of $\mathcal{F}_{s+}$ for every $t > 0$. Let $G \in \mathcal{F}_{s+}$ and $t > 0$. For each $\varepsilon > 0$ such that $t > \varepsilon$, $G \in \mathcal{F}_{s+\varepsilon}$, so that if $f \in C_b(\mathbb{R}^k)$ one has

$$\mathbb{E}(\mathbf{1}_G f(B_{t+s} - B_{s+\varepsilon})) = P(G) \cdot \mathbb{E}f(B_{t+s} - B_{s+\varepsilon}).$$

Letting $\varepsilon \downarrow 0$ on both sides,

$$\mathbb{E}(\mathbf{1}_G f(B_{t+s} - B_s)) = P(G)\mathbb{E}f(B_{t+s} - B_s).$$

■

One may define the "past up to time τ" as the $\sigma-$field of events $\mathcal{F}_\tau$ given by

$$\mathcal{F}_\tau := \sigma(Z_{t\wedge\tau} : t \geq 0). \tag{7.10}$$

The stochastic process $\{\tilde{Z}_t : t \geq 0\} := \{Z_{t\wedge\tau} : t \geq 0\}$ is referred to as the *process stopped at* τ. Events in $\mathcal{F}_\tau$ depend only on the process stopped at τ. The stopped process contains no further information about the process $\{Z_t : t \geq 0\}$ beyond the time τ. Alternatively, a description of the past up to time τ which is often more useful for checking whether a particular event belongs to it may be formulated as follows.

Definition 7.6. Let τ be a stopping time with respect to a filtration $\mathcal{F}_t, t \geq 0$. The *pre-τ σ−field* is

$$\mathcal{F}_\tau = \{F \in \mathcal{F} : F \cap [\tau \leq t] \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

For example, using this definition it is simple to check that

$$[\tau \leq t] \in \mathcal{F}_\tau, \forall\ t \geq 0, \qquad [\tau < \infty] \in \mathcal{F}_\tau. \tag{7.11}$$

Remark 7.5. We will always use[4] Definition 7.6, and not (7.10). Note, however, that $t \wedge \tau \leq t$ for all t, so that $\sigma(X_{t\wedge\tau} : t \geq 0\}$ is surely contained in $\mathcal{F}_\tau$ (see Exercises 2 and 13(i).).

The future relative to τ is the *after-τ process* $Z_\tau^+ = \{(Z_\tau^+)_t : t \geq 0\}$ obtained by viewing $\{Z_t : t \geq 0\}$ from time $t = \tau$ onwards, for $\tau < \infty$. This is,

$$(Z_\tau^+)_t(\omega) = Z_{\tau(\omega)+t}(\omega), \qquad t \geq 0, \qquad \text{on } [\tau < \infty]. \tag{7.12}$$

It is useful to record the following definition.

Definition 7.7. Let S be a metric space with Borel σ-field $\mathcal{B}(S)$. The transition probabilities $p(t, x, dy), x \in S, t \geq 0$, on $\mathcal{B}(S)$, are said to have the *Feller property* if $x \to p(t, x, dy)$ is (weakly) continuous for each $t \geq 0$. Equivalently, the linear operators S defined for bounded, measurable functions $f : S \to \mathbb{R}$ by $f(\cdot) \to \int_S f(y)p(t; \cdot, dy), t \geq 0$, map bounded continuous functions to bounded continuous functions.

Remark 7.6. The equivalent statement in Definition 7.7 is merely the definition of weak convergence of $p(t; z, dy) \Rightarrow p(t; x, dy)$, as $z \to x$, for each $x \in S$.

Theorem 7.7 ***(Strong Markov Property for Right-Continuous Markov Processes).*** Let $\{X_t : t \geq 0\}$ be a Markov process as in Proposition 7.4, satisfying conditions (1)–(3) above. Assume also that the transition probabilities are Feller continuous. Let τ be a stopping time. On $[\tau < \infty]$, the conditional distribution of X_τ^+ given $\mathcal{F}_\tau$

[4] The proof of the equivalence of (7.10) and that of Definition 7.6 may be found in (Stroock and Varadhan, 1980, p. 33).

is the same as the distribution of the $\{X_t : t \geq 0\}$ starting at X_τ. In other words, this conditional distribution is P_{X_τ} on $[\tau < \infty]$.

Proof. First assume that τ has countably many values ordered as $0 \leq s_1 < s_2 < \cdots$. Consider a finite-dimensional function of the after-τ process of the form

$$h(X_{\tau+t_1'}, X_{\tau+t_2'}, \ldots, X_{\tau+t_r'}), \qquad [\tau < \infty], \tag{7.13}$$

where h is a bounded continuous real-valued function on S^r and $0 \leq t_1' < t_2' < \cdots < t_r'$. It is enough to prove

$$\mathbb{E}\left[h(X_{\tau+t_1'}, \ldots, X_{\tau+t_r'})\mathbf{1}_{[\tau<\infty]} \mid \mathcal{F}_\tau\right] = [\mathbb{E}_y h(X_{t_1'}, \ldots, X_{t_r'})]_{y=X_\tau}\mathbf{1}_{[\tau<\infty]}. \tag{7.14}$$

That is, for every $A \in \mathcal{F}_\tau$ we need to show that

$$\mathbb{E}(\mathbf{1}_A h(X_{\tau+t_1'}, \ldots, X_{\tau+t_r'})\mathbf{1}_{[\tau<\infty]}) = \mathbb{E}\left(\mathbf{1}_A \left[\mathbb{E}_y h(X_{t_1'}, \ldots, X_{t_r'})\right]_{y=X_\tau} \mathbf{1}_{[\tau<\infty]}\right). \tag{7.15}$$

Now

$$[\tau = s_j] = [\tau \leq s_j] \cap [\tau \leq s_{j-1}]^c \in \mathcal{F}_{s_j},$$

so that $A \cap [\tau = s_j] \in \mathcal{F}_{s_j}$. Express the left side of (7.15) as

$$\sum_{j=1}^{\infty} \mathbb{E}\big(\mathbf{1}_{A\cap[\tau=s_j]} h(X_{s_j+t_1'}, \ldots, X_{s_j+t_r'})\big). \tag{7.16}$$

By the Markov property, the j–th summand in (7.16) equals

$$\begin{aligned}&\mathbb{E}(\mathbf{1}_A \mathbf{1}_{[\tau=s_j]}[\mathbb{E}_y h(X_{t_1'}, \ldots, X_{t_r'})]_{y=X_{s_j}})\\ &\quad = \mathbb{E}(\mathbf{1}_A \mathbf{1}_{[\tau=s_j]}[\mathbb{E}_y h(X_{t_1'}, \ldots, X_{t_r'})]_{y=X_\tau}).\end{aligned}$$

Summing this over j, one obtains the desired relation (7.15). This completes the proof in the case τ has countably many values $0 \leq s_1 < s_2 < \cdots$.

The case of more general τ may be dealt with by approximating it by stopping times assuming countably many values. Specifically, for each positive integer n define

$$\tau_n = \begin{cases} \frac{j}{2n} & if\ \frac{j-1}{2^n} < \tau \leq \frac{j}{2^n}, \quad j = 0, 1, 2, \ldots \\ \infty & if\ \tau = \infty. \end{cases} \tag{7.17}$$

Since

$$[\tau_n = \tfrac{j}{2^n}] = [\tfrac{j-1}{2^n} < \tau \le \tfrac{j}{2^n}] = [\tau \le \tfrac{j}{2^n}] \backslash [\tau \le \tfrac{j-1}{2^n}] \in \mathcal{F}_{j/2^n},$$
$$[\tau_n \le t] = \bigcup_{j: j/2^n \le t} [\tau_n = \tfrac{j}{2^n}] \in \mathcal{F}_t \qquad \text{for all } t \ge 0.$$

Therefore, τ_n is a stopping time for each n and $\tau_n(\omega) \downarrow \tau(\omega)$ as $n \uparrow \infty$ for each $\omega \in \Omega$. Also one may easily check $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ from definition (see Exercise 2). Let h be a bounded *continuous* function on S^r. Define

$$\varphi(y) \equiv \mathbb{E}_y h(X_{t_1'}, \ldots, X_{t_r'}). \tag{7.18}$$

In view of Feller continuity of $p(t; x, dy)$, φ is continuous (Exercise 8). Let $A \in \mathcal{F}_\tau (\subset \mathcal{F}_{\tau_n})$. One has

$$\mathbb{E}(\mathbf{1}_A h(X_{\tau_n + t_1'}, \ldots, X_{\tau_n + t_r'}) \mathbf{1}_{[\tau_n < \infty]}) = \mathbb{E}(\mathbf{1}_A \varphi(X_{\tau_n}) \mathbf{1}_{[\tau_n < \infty]}). \tag{7.19}$$

Since h, φ are continuous, $\{X_t : t \ge 0\}$ has right-continuous sample paths, and $\tau_n \downarrow \tau$ as $n \to \infty$, Lebesgue's dominated convergence theorem may be used on both sides of (7.19) to get

$$\mathbb{E}(\mathbf{1}_A h(X_{\tau + t_1'}, \ldots, X_{\tau + t_r'}) \mathbf{1}_{[\tau < \infty]}) = \mathbb{E}(\mathbf{1}_A \varphi(X_\tau) \mathbf{1}_{[\tau < \infty]}). \tag{7.20}$$

This establishes (7.15), and therefore (7.14). Since finite-dimensional distributions determine a probability on $(\Gamma, \mathcal{C})$, the proof is complete. ■

In view of the sample path continuity and Markov property of Brownian motion it now follows that

***Corollary 7.8** (Strong Markov Property for Brownian Motion).* Brownian motion on $\mathbb{R}^k$ with drift vector μ and diffusion matrix Σ has the strong Markov property.

Using the construction of the Poisson process, as well as the compound Poisson process, as a right-continuous process with stationary independent increments given in Chapter 5 one obtains the following (Exercise 9).

Corollary 7.9. The compound Poisson process has the strong Markov property.

The examples below further illustrate the usefulness of Theorem 7.7 in typical computations. In all these examples $B \equiv B^0 = \{B_t : t \ge 0\}$ is a one-dimensional standard Brownian motion starting at zero. For $\omega \in C([0, \infty) : \mathbb{R})$ define, for every $a \in \mathbb{R}$,

$$\overline{\tau}_a(\omega) := \inf\{t \ge 0 : \omega(t) = a\},$$
$$\tau_a := \overline{\tau}_a(B), \tag{7.21}$$

with the usual convention that the infimum of an empty set of numbers is ∞.

Example 2 *(Independent Increments and Distribution of the First Passage Process).* A key result for this application is the reflection principle, to now be obtained as an application of the strong Markov property Theorem 7.7. Reflection of paths of Brownian motion starting at x about a horizontal line at level $a \neq x$ after first contact provides an important transformation under which the starting point x and sample path continuity are clearly preserved. In fact, according to the reflection principle, the path distribution is invariant under such a transformation!

Theorem 7.10 *(The Reflection Principle).* Let $B^x = \{B_t^x : t \geq 0\}$ be a one-dimensional standard Brownian motion, with $B_0^x = x$ a.s. Fix any $a \neq x$. Then the process W defined by

$$W_t = \begin{cases} B_t^x & \text{if } t < \tau_a \\ 2a - B_t^x & \text{if } t \geq \tau_a \end{cases} \tag{7.22}$$

is a standard one-dimensional Brownian motion starting at x.

Proof. For simplicity of notation, we will omit the superscript x of B^x and B_t^x, here. First note that by the strong Markov property (Theorem 7.7), the conditional distribution of the *after–*τ_a *process* $B_{\tau_a}^+ \equiv \{B_{\tau_a+t} : t \geq 0\}$, given the *pre–*$\tau_a$ σ*-field* $\mathcal{F}_{\tau_a}$ is the same as the distribution of $a + B^0$, say, P_a. Here $\mathcal{F}_{\tau_a} = \{A \in \mathcal{F} : A \cap [\tau_a \leq t] \in \mathcal{F}_t \ \forall\, t \geq 0\}$, and $\mathcal{F}_t := \sigma\{B_s, 0 \leq s \leq t\}$ $(t \geq 0)$. In particular, since the latter distribution is constant on Ω, $B_{\tau_a}^+$ is independent of $\mathcal{F}_{\tau_a}$. Now $B_{\tau_a}^+ - a$ and $a - B_{\tau_a}^+ = \{a - B_{\tau_a+t} : t \geq 0\}$ have the same distribution, namely, P_0. Therefore, $\{a + a - B_{\tau_a+t} : t \geq 0\} \equiv W_{\tau_a}^+$ is independent of $\mathcal{F}_{\tau_a}$ and has distribution P_a, the same holds for $B_{\tau_a}^+$. Since (a) W is the same function of $Y \equiv \{B_{t\wedge\tau_a} : t \geq 0\}$ and $W_{\tau_a}^+$ as B is of Y and $B_{\tau_a}^+$, and (b) $(Y, W_{\tau_a}^+)$ and $(Y, B_{\tau_a}^+)$ have the same distribution, it follows that W and B have the same distribution P_x (Also see Exercise 13). ■

For the following Corollaries, and elsewhere in this chapter, $\{B_t : t \geq 0\}$ denotes a standard Brownian motion, started at 0.

Corollary 7.11. The joint distribution of the *running maximum* $M_t := \max\{B_s : 0 \leq s \leq t\}$ and B_t is given by

$$\begin{aligned} P(M_t \geq a, B_t \leq y) &= P(B_t \geq 2a - y) \\ &= 1 - \int_{-\infty}^{(2a-y)/\sqrt{t}} (2\pi)^{-\frac{1}{2}} e^{-x^2/2} dx, \ \ (y \leq a). \end{aligned} \tag{7.23}$$

Proof. Note that it follows from Theorem 7.10, and the fact that both $B_{\tau_a}^+$ and $W_{\tau_a}^+$ are independent of τ_a (by the strong Markov property), that for $y \leq a$,

$$\begin{aligned} P(M_t \geq a, B_t \leq y) &\equiv P(\tau_a \leq t, 2a - W_t \leq y) = P(\tau_a \leq t, 2a - B_t \leq y) \\ &= P(\tau_a \leq t, B_t \geq 2a - y) = P(B_t \geq 2a - y), \end{aligned}$$

since $2a - y \geq a$, and $[B_t \geq 2a - y] \subset [\tau_a \leq t]$. ■

Corollary 7.12.

(i) The distribution of M_t is given by

$$P(M_t \geq a) = 2P(B_t \geq a) = \sqrt{\frac{2}{\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} dx \quad (a > 0). \tag{7.24}$$

(ii) The joint probability density function of (M_t, B_t) is

$$f(a, y) = \begin{cases} \left(\frac{2}{\pi}\right)^{1/2} t^{-3/2}(2a - y)e^{-(2a-y)^2/2t} & \text{for } a > 0,\ y < a \\ 0 & \text{otherwise.} \end{cases} \tag{7.25}$$

Proof.

(i) Apply 7.23 with $y = a$ to get $P(M_t \geq a) = P(M_t \geq a, B_t \leq a) + P(M_t \geq a, B_t > a) = P(B_t \geq a) + P(B_t > a)$.
(ii) Differentiate the integral in (7.23) with respect to y and a in succession (and change sign).

■

Corollary 7.13. For fixed $t > 0$, $R_t := M_t - B_t$ has the same distribution as $|B_t|$.

Proof. Since B_t has pdf $\varphi_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$, the distribution of $|B_t|$ has pdf $\varphi_t(x) + \varphi_t(-x) = 2\varphi_t(x), x \geq 0$. Computing the joint pdf of $(M_t - B_t, M_t)$ as a linear transformation of (M_t, B_t) with determinant one, one obtains (Exercise 11) $(\frac{2}{\pi})^{1/2} t^{-3/2}(x + y)e^{-(x+y)^2/2t} = -2\frac{d}{dy}\varphi_t(x + y)$. Integrating with respect to y therefore yields the marginal pdf of $M_t - B_t$ as $2\varphi_t(x), x \geq 0$. ■

Remark 7.7 ***(Lévy-Skorokhod Formula).*** In Chapter 19 of the present text it is shown that $|B_t|, t \geq 0$, is a Markov process starting at zero with transition probabilities $p(t; x, y) = \varphi_t(y + x) + \varphi_t(y - x), x, y \geq 0$, where φ_t is the mean zero, variance t Gaussian density. With a bit more work one shows (Theorem 19.3) that $M_t - B_t, t \geq 0$ is also a Markov process starting at zero, with the same transition probability densities as the *reflecting Brownian motion* $|B_t|, t \geq 0$. Since B and the reflection $|B|$ have the same zeroes, it will follow that *the distribution of the time of the last zero in* $[0, 1]$ *of Brownian motion* coincides with *the distribution of the location of the maximum value of Brownian motion on* $[0, 1]$. In Chapter 18 this arcsine distribution is computed.

In reference to the following corollary, recall that a real-valued random variable Y with distribution Q on $\mathbb{R}$ is said to have a *stable distribution* with *exponent* α and *centering constants* $c_n, n \geq 1$, if for every $n = 1, 2, \ldots$ $Q^{*n}((-\infty, n^{\frac{1}{\alpha}} z + c_n]) = Q((-\infty, z])\ \forall z \in \mathbb{R}$ or, alternatively, if $Y_1, Y_2, \ldots$ are i.i.d. with distribution Q, then $(Y_1 + \cdots + Y_n - c_n)/n^{\frac{1}{\alpha}}$ also has the same distribution Q; here Q^{*n} denotes the $n-$fold convolution of the probability Q. Familiar examples are the standard normal distribution ($\alpha = 2, c_n = 0, n \geq 1$) and the Cauchy distribution. ($\alpha = 1, c_n = 0, n \geq 1$).

Corollary 7.14 *(First Passage Time Process for Standard Brownian Motion).* (i) $\{\tau_a : a \geq 0\}$ is an increasing process with stationary independent increments, and (ii) the first passage time τ_a has the stable law distribution of exponent $\frac{1}{2}$, centering constants $c_n = 0$, concentrated on $[0, \infty)$ with probability density function

$$g_a(t) = a\left(\frac{1}{2\pi}\right)^{1/2} t^{-3/2}e^{-a^2/2t}, \qquad t > 0. \tag{7.26}$$

Proof. Let $\mathcal{F}_t := \sigma(B_s : s \leq t), t \geq 0$. (i) Let $0 < a < b$. Since $\tau_b = \tau_a + \overline{\tau}_b(B^+_{\tau_a})$ (See (7.21)) $\tau_b - \tau_a$ is a function of $B^+_{\tau_a}$ which is independent of $\mathcal{F}_{\tau_a}$. Note that the distribution of $\overline{\tau}_b(B_{\tau_a})$ is that of the first time Brownian motion B, starting at zero, reaches $b - a$, i.e., of $\overline{\tau}_{b-a}(B)$. Also, for all $a' < a$, one has $\tau_{a'} \leq \tau_a$ and therefore from the definition of stopping times, one may check $\mathcal{F}_{\tau_{a'}} \subset \mathcal{F}_{\tau_a}$ (Exercise 2). It follows that $\tau_b - \tau_a$ is independent of $\{\tau_{a'} : 0 \leq a' \leq a\}$. In particular, for any given $0 \leq a_1 < a_2 < \cdots < a_k$, the random variables $\tau_{a_1}, \tau_{a_2} - \tau_{a_1}, \tau_{a_3} - \tau_{a_2}, \ldots,$ $\tau_{a_k} - \tau_{a_{k-1}}$ are independent. (ii) Differentiate the right side of (7.24) with respect to t in order to get (7.26), using the identity $[\tau_a \leq t] = [M_t \geq a]$. The stable law property follows from the observations: (a) In view of (i) the pdf τ_{na} equals g_a^{*n}, and (b) the pdf of τ_a is the same as that of τ_{na}/n^2. ■

As an application one may obtain an otherwise quite challenging integral computation (see Exercise 19) of the Laplace transform of τ_a, namely, $\mathbb{E}e^{-\lambda\tau_a}$, $\lambda > 0$.

Proposition 7.15. $\mathbb{E}e^{-\lambda\tau_a} = e^{-|a|\sqrt{k\lambda}}$ for all $\lambda > 0$ for a constant $k > 0$.

Proof. Without loss of generality, take $a > 0$. In view of Corollary 7.14 it follows that $\tau_{a+b} = \tau_{a+b} - \tau_b + \tau_b = \tilde{\tau}_a + \tau_b$ with $\tilde{\tau}_a$ and τ_b independent first passage times to a and b, respectively. Thus, the Laplace transform $\mathbb{E}e^{-\lambda\tau_a}$ is log-linear in a. One may obtain $\sqrt{\lambda}$ up to a positive constant k from the Brownian motion scaling $\tau_a = c^{-2}\tau_{ca}, c > 0$. That is, $\tau_a = \inf\{t \geq 0 : B_t = a\} = \inf\{t \geq 0 : c^{-1}B_{c^2t} = a\} = \inf\{t \geq 0 : B_{c^2t} = ca\} = \frac{1}{c^2}\inf\{c^2t \geq 0 : B_{c^2t} = ca\} = c^{-2}\tau_{ca}$. Thus $\mathbb{E}e^{-\lambda\tau_a} = e^{-|a|\sqrt{k\lambda}}$, for a constant $k > 0$ to be determined. ■

Remark 7.8. The explicit evaluation of the integral $\int_0^\infty \frac{|a|}{\sqrt{2\pi}} t^{-\frac{3}{2}} e^{-\lambda t} e^{-\frac{a^2}{2t}} dt$ is a non-trivial exercise in calculus. A very simple approach can be achieved via martingale theory (Chapter 11). However, even the determination of $k = 2$ presents a calculus challenge (see Exercise 19). The Laplace transform will be used in the proof of Proposition 16.3 in Chapter 16; however, it does not require the determination of k, only the general form $e^{-|a|\sqrt{k\lambda}}$, $\lambda > 0$.

One may also explicitly obtain the distribution of the escape time $\tau_{-b,a} = \tau_{-b} \wedge \tau_a$, $a, b > 0$ using the strong Markov property as follows.

Lemma 1 ***(First Passage Decomposition).*** For arbitrary positive a, b and Borel set $C \subset \mathbb{R}$, one has

$$P(\tau_{-b,a} \le t, B_t \in C) = \int_{[0,t]\times\mathbb{R}} P\left(B^y_{t-u} \in C\right) \Gamma(dudy),$$

where Γ is the (joint) distribution of $(\tau_{-b,a}, B_{\tau_{-b,a}})$.

Proposition 7.16. The distribution function $G(t)$ of $\tau_{-b,a}$ is given by

$$\begin{aligned} G(t) = H_a(t) + H_b(t) - \{H_{2a+b}(t) + H_{a+2b}(t)\} + \{H_{3a+2b}(t) + H_{2a+3b}(t)\} \\ - \{H_{4a+3b}(t) + H_{3a+4b}(t)\} + \{H_{5a+4b}(t) + H_{4a+5b}(t)\} - + \dots , \end{aligned}$$

where $H_x(t) = 2(1 - \Phi(x/\sqrt{t})$, $\Phi(\cdot)$ being the distribution function of the standard normal distribution $\mathcal{N}(0, 1)$.

Proof. Since $B_{\tau_{-b,a}} = a$ or $-b$, with probability one, if one takes $C = [a, \infty) \cup (-\infty, -b]$ in the first passage decomposition (Lemma 1), then

$$\begin{aligned} &P(B_t \ge a \text{ or } \le -b) \\ &\quad \equiv P(\tau_{-b,a} \le t, B_t \ge a \text{ or } \le -b) \\ &\quad = \int_{[0,t]\times\{a,-b\}} P\left(B^y_{t-u} \ge a \text{ or } \le -b\right) \Gamma(dudy). \end{aligned} \tag{7.27}$$

But, at $y = a$, the last integrand equals

$$\begin{aligned} &P(B^a_{t-u} \ge a \text{ or } \le -b) \\ &\quad = P(B^a_{t-u} \ge a) + P(B^a_{t-u} \le -b) \\ &\quad = \frac{1}{2} + P(B_{t-u} \le -a - b) \\ &\quad = \frac{1}{2} + P(B_{t-u} \ge a + b). \end{aligned}$$

At $y = -b$, the integrand has the same value, since

$$\begin{aligned} &P\left(B^{-b}_{t-u} \ge a \text{ or } \le -b\right) \\ &\quad = P\left(B^{-b}_{t-u} \le -b\right) + P\left(B^{-b}_{t-u} \ge a\right) \\ &\quad = \frac{1}{2} + P\left(B_{t-u} \ge a + b\right). \end{aligned}$$

Hence one has

$$P\left(B_t \geq a\right)+P\left(B_t \geq b\right)=P\left(B_t \geq a \quad \text{or} \quad \leq -b\right)$$
$$=\int_{[0,t]}\left(\frac{1}{2}+P(B_{t-u} \geq a+b)\right) G(du), \quad (7.28)$$

where $G(du) = \int_{\{a,-b\}} \Lambda(dudy)$ is the distribution of $\tau_{a,-b}$. Next, by Corollary 7.12, recalling $\tau_x := \inf\{t \geq 0 : B_t = x\}$, one has

$$P(B_t \geq x)=\frac{1}{2} P(\tau_x \leq t), \qquad (x>0). \qquad (7.29)$$

Using this on both sides of (7.28), and denoting by $H_y(du)$ the distribution of τ_y (and by $H_y(t)$ its distribution function depending on the mathematical context), one gets

$$\frac{1}{2} H_a(t)+\frac{1}{2} H_b(t)=\frac{1}{2} G(t)+\frac{1}{2} G * H_{a+b}(t),$$

or,

$$G(t)=H_a(t)+H_b(t)-G * H_{a+b}(t). \qquad (7.30)$$

Iterating repeatedly, and remembering that $H_c * H_d = H_{c+d}$ (The process $\{\tau_y : y \geq 0\}$ has independent homogeneous increments), we arrive at the assertion. ∎

Let $m_t := \min\{B_s : 0 \leq s \leq t\}$ and $M_t := \max\{B_s : 0 \leq s \leq t\}$. Then the distribution of (m_t, M_t) readily follows.

Corollary 7.17 *(Joint Distribution of Maximum and Minimum of Brownian Motion)**.***

$$\begin{aligned} &P\left(M_t<a \text{ and } m_t>-b\right) \\ &\quad =1-H_a(t)-H_b(t)+\{H_{2a+b}(t)+H_{a+2b}(t)\}-\{H_{3a+2b}(t)+H_{2a+3b}(t)\} \\ &\qquad +\{H_{4a+3b}(t)+H_{3a+4b}(t)\}-\{H_{5a+4b}(t)+H_{4a+5b}(t)\}+-\ldots, \end{aligned} \qquad (7.31)$$

where $H_x(t) = 2(1 - \Phi(x/\sqrt{t})$.

Proof. One has

$$P\left(M_t<a \text{ and } m_t>-b\right)=1-P\left(\tau_{-b,a} \leq t\right)=1-G(t), \; (a>0, b>0).$$

∎

Just as with the random walk Example 1, the following illustrates a use of the strong Markov property in a similar calculation for Brownian motion.

Example 3 *(Boundary Value Distribution of Brownian Motion).* Let $\{Z_t := x + \sigma B_t : t \geq 0\}$ be a one-dimensional Brownian motion with zero drift and diffusion coefficient $\sigma^2 > 0$ started at $x \in [c, d]$ for $c < d$. The stopping time $\tau_c \wedge \tau_d$ denotes the first time to reach the "boundary" states $\{c, d\}$. Define

$$\psi(x) := P_x(Z_{\tau_c \wedge \tau_d} = c) \equiv P_x(\{Z_t : t \geq 0\} \text{ reaches } c \text{ before } d), \ (c \leq x \leq d). \tag{7.32}$$

Fix $x \in (c, d)$ and $h > 0$ such that $[x - h, x + h] \subset (c, d)$. Unlike the discrete parameter case there is no "first step" to consider. It will be convenient to consider $\tau = \tau_{x-h} \wedge \tau_{x+h}$, i.e., τ is the first time $\{Z_t : t \geq 0\}$ reaches $x - h$ or $x + h$. Then $P_x(\tau < \infty) = 1$, by the simple computation

$$\begin{aligned}
P_x(\tau > t) &\leq P_x(x - h < Z_t < x + h) \\
&= \frac{1}{(2\pi\sigma^2 t)^{1/2}} \int_{x-h}^{x+h} \exp\left\{-\frac{(y - x)^2}{2\sigma^2 t}\right\} dy \\
&= \frac{1}{(2\pi)^{1/2}} \int_{-h/\sigma\sqrt{t}}^{h/\sigma\sqrt{t}} \exp\left\{-\frac{z^2}{2}\right\} dz \to 0 \quad \text{as } t \to \infty.
\end{aligned} \tag{7.33}$$

Now,

$$\begin{aligned}
\psi(x) &= P_x(\{Z_t : t \geq 0\} \text{ reaches } c \text{ before } d) \\
&= P_x(\{(Z_\tau^+)_t : t \geq 0\} \text{ reaches } c \text{ before } d) \\
&= \mathbb{E}_x(P_x(\{(Z_\tau^+)_t : t \geq 0\} \text{ reaches } c \text{ before } d \mid \{Z_{t\wedge\tau} : t \geq 0\})).
\end{aligned} \tag{7.34}$$

The strong Markov property now gives that

$$\psi(x) = \mathbb{E}_x(\psi(Z_\tau)) \tag{7.35}$$

so that by symmetry of the normal distribution,

$$\begin{aligned}
\psi(x) &= \psi(x - h)P_x(Z_\tau = x - h) + \psi(x + h)P_x(Z_\tau = x + h) \\
&= \psi(x - h)\frac{1}{2} + \psi(x + h)\frac{1}{2}.
\end{aligned} \tag{7.36}$$

Rewriting (7.36) as $\frac{\psi(x+h)-\psi(x)}{2} - \frac{\psi(x)-\psi(x-h)}{2} = 0$, then dividing by h, and letting $h \downarrow 0$, one gets $\psi''(x) = 0$, i.e., ψ is linear in x. By (7.32), $\psi(x)$ satisfies the boundary conditions $\psi(c) = 1$, $\psi(d) = 0$. Therefore,

$$\psi(x) = \frac{d - x}{d - c}. \tag{7.37}$$

An alternative derivation using martingales is given in Chapter 13, Example 2. Now, by (7.33) and (7.37),

$$P_x(\{Z_t : t \geq 0\} \text{ reaches } d \text{ before } c) = 1 - \psi(x) = \frac{x - c}{d - c} \tag{7.38}$$

for $c \leq x \leq d$. It follows, on letting $d \uparrow \infty$ in (7.37) and $c \downarrow -\infty$ in (7.38) that

$$P_x(\tau_y < \infty) = 1 \qquad \text{for all } x, y. \tag{7.39}$$

Consider a Brownian motion $\{X_t^x \equiv x + \sigma B_t : t \geq 0\}$ with drift $\mu = 0$ and diffusion coefficient $\sigma^2 > 0$, starting at x. One has

$$\begin{aligned} P(\tau_c^x < \tau_d^x) &= P(\{X_t^x : t \geq 0\} \text{ reaches } c \text{ before } d)\ (c < x < d) \\ &= P(\{B_t^{\frac{x}{\sigma}} : t \geq 0\} \text{ reaches } \frac{c}{\sigma} \text{ before } \frac{d}{\sigma}),\ (c < x < d), \end{aligned} \tag{7.40}$$

where

$$\tau_y^x := \inf\{t \geq 0 : X_t^x = y\}. \tag{7.41}$$

Thus one has

Proposition 7.18 ***(****Boundary Distribution Under Zero Drift****).*** Let $c < x < d$. Then

$$P(\tau_c^x < \tau_d^x) = \frac{d - x}{d - c} \quad (c < x < d,\ \mu = 0),$$

$$P(\tau_d^x < \tau_c^x) = \frac{x - c}{d - c} \quad (c < x < d,\ \mu = 0).$$

Letting $d \to +\infty$ in the first result and $c \to -\infty$ in the second, one obtains the following.

Corollary 7.19 ***(****Pointwise Recurrence Under Zero Drift****).***

$$\begin{aligned} P(\tau_c^x < \infty) &= P(\{X_t^x : t \geq 0\} \text{ ever reaches } c) = 1 \quad (c < x, \mu = 0), \\ P(\tau_d^x < \infty) &= P(\{X_t^x : t \geq 0\} \text{ ever reaches } d) = 1 \quad (x < d, \mu = 0). \end{aligned}$$

Taken with the strong Markov property these relations imply that a (one-dimensional) Brownian B motion with zero drift is *(pointwise) recurrent* in the sense that $B_{t_n} = x, n = 1, 2, \dots,$ for an unbounded, increasing (random) sequence t_n, just as is a simple symmetric random walk.

It is much simpler to show that Brownian motion X with drift is *transient* in the sense that, with probability one, $|X_t| \to \infty$ as $t \to \infty$, for arbitrarily given starting state x. In fact, one may use the invariance under time-inversion for standard

Brownian motion starting at 0 to obtain the distribution of the time of the last visit to 0 by a Brownian motion with (nonzero) drift; see Proposition 16.6 of Chapter 16.

Proposition 7.20 *(Transience of One-Dimensional Brownian Motion with Drift).* Let $x \in \mathbb{R}$. With probability one,

$$\lim_{t\to\infty} x + \mu t + B_t = \begin{cases} +\infty & \text{if } \mu > 0 \\ -\infty & \text{if } \mu < 0. \end{cases}$$

Proof. This can be deduced by an appeal to a continuous time version of the strong law of large numbers, just as in (2.9), (2.10) of Chapter 2. The details are left to the reader (Exercise 16). ■

A proof of the arcsine law is given in Chapter 18, Corollary 18.2, as well.

In anticipation of Corollary 17.6 in Chapter 17, the calculation of the boundary value probabilities when the drift is a nonzero quantity μ can be made as a limit of the corresponding probabilities for asymmetric random walk. In particular, the following will be proven in Chapter 17.

Proposition 7.21. Consider the one-dimensional Brownian motion with nonzero drift μ.

$$P(\tau_c^x < \tau_d^x) = \frac{1 - \exp\{2(d-x)\mu/\sigma^2\}}{1 - \exp\{2(d-c)\mu/\sigma^2\}} \quad (c < x < d, \mu \neq 0),$$

$$P(\tau_d^x < \tau_c^x) = \frac{1 - \exp\{-2(x-c)\mu/\sigma^2\}}{1 - \exp\{-2(d-c)\mu/\sigma^2\}} \quad (c < x < d, \mu \neq 0).$$

$$P(\tau_c^x < \infty) = \exp\left\{-\frac{2(x-c)\mu}{\sigma^2}\right\}\ (c < x, \mu > 0);\ \ P(\tau_c^x < \infty) = 1\ (c < x, \mu < 0).$$

$$P(\tau_d^x < \infty) = 1\ (x < d, \mu > 0);\ \ P(\tau_d^x < \infty) = \exp\{2(d-x)\mu/\sigma^2\}\ (x < d, \mu < 0).$$

Remark 7.9. These distributions will also be computed by martingale methods in Chapter 13.

From Proposition 7.21 one also has (Exercise 15),

Corollary 7.22. For the one-dimensional Brownian motion X^0 with nonzero drift μ and diffusion coefficient $\sigma^2 > 0$, in the case $\mu > 0$, the extremal random variable $-m = -\inf_{t\geq 0} X_t^0 \geq 0, t \geq 0$, is exponentially distributed on $[0,\infty)$ with parameter $\frac{2\mu}{\sigma^2}$, and in the case $\mu < 0$, $M = \sup_{t\geq 0} X_t^0 \geq 0$ is exponentially distributed on $[0,\infty)$ with parameter $\frac{2|\mu|}{\sigma^2}$.

Of course these results also imply that a Brownian motion with a nonzero drift is transient.

As the following proof shows, the law of the iterated logarithm[5] (LIL) is yet another remarkable and powerful encapsulation of the basic structural properties of Brownian motion.

Theorem 7.23 *(Law of the Iterated Logarithm (LIL) for Brownian Motion)***.** Each of the following holds with probability one:

$$\overline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2t\log\log t}}=1, \qquad \underline{\lim}_{t\to\infty}\frac{B_t}{\sqrt{2t\log\log t}}=-1.$$

Proof. Let $\varphi(t) := \sqrt{2t\log\log t}, t > 0$. Let us first show that for any $0 < \delta < 1$, one has with probability one that

$$\overline{\lim}_{t\to\infty}\frac{B_t}{\varphi(t)} \le 1+\delta. \tag{7.42}$$

For arbitrary $\alpha > 1$, partition the time interval $[0,\infty)$ into subintervals of exponentially growing lengths $t_{n+1}-t_n$, where $t_n=\alpha^n$, and consider the event

$$E_n := \left[\max_{t_n\le t\le t_{n+1}}\frac{B_t}{(1+\delta)\varphi(t)} > 1\right].$$

Since $\varphi(t)$ is a non-decreasing function, one has, using Corollary 7.12, a scaling property, and Feller's tail probability estimates[6] for the normal distribution, that

$$\begin{aligned} P(E_n) &\le P\left(\max_{0\le t\le t_{n+1}} B_t > (1+\delta)\varphi(t_n)\right)\\ &= 2P\left(B_1 > \frac{(1+\delta)\varphi(t_n)}{\sqrt{t_{n+1}}}\right)\\ &\le \sqrt{\frac{2}{\pi}}\frac{\sqrt{t_{n+1}}}{(1+\delta)\varphi(t_n)}e^{-\frac{(1+\delta)^2\varphi^2(t_n)}{2t_{n+1}}} \le c\frac{1}{n^{(1+\delta)^2/\alpha}} \end{aligned} \tag{7.43}$$

for a constant $c>0$ and all $n^2 \ge (\log\alpha+1)^{-1}$. For a given $\delta>0$ one may select $1<\alpha<(1+\delta)^2$ to obtain $P(E_n\ i.o.)=0$ from the Borel–Cantelli lemma (Part I). Thus we have (7.42). Since $\delta > 0$ is arbitrary we have with probability one that

$$\overline{\lim}_{t\to\infty}\frac{B_t}{\varphi(t)} \le 1. \tag{7.44}$$

Next let us show that with probability one,

$$\overline{\lim}_{t\to\infty}\frac{B_t}{\varphi(t)} \ge 1. \tag{7.45}$$

[5]The LIL for Brownian motion was originally obtained in Khintchine (1933).

[6]See BCPT p.82.

For this consider the independent increments $B_{t_{n+1}} - B_{t_n}$, $n \geq 1$. For $\theta = \frac{t_{n+1}-t_n}{t_{n+1}} = \frac{\alpha-1}{\alpha} < 1$, again using Feller's tail probability estimate and Brownian scale change,

$$\begin{aligned} P\left(B_{t_{n+1}} - B_{t_n} > \theta\varphi(t_{n+1})\right) &= P\left(B_1 > \sqrt{\frac{\theta}{t_{n+1}}}\varphi(t_{n+1})\right) \\ &\geq \frac{c'}{\sqrt{2\theta \log\log t_{n+1}}} e^{-\theta \log\log t_{n+1}} \\ &\geq \frac{c}{\sqrt{\log n}} n^{-\theta} \end{aligned} \tag{7.46}$$

for suitable constants c, c' depending on α and for all $n^2 > (\log \alpha + 1)^{-1}$. It follows from the Borel–Cantelli Lemma (Part II) that with probability one,

$$B_{t_{n+1}} - B_{t_n} > \theta\varphi(t_{n+1})\ i.o. \tag{7.47}$$

Also, by (7.44) and replacing $\{B_t : t \geq 0\}$ by the standard Brownian motion $\{-B_t : t \geq 0\}$,

$$\underline{\lim}_{t\to\infty} \frac{B_t}{\varphi(t)} \geq -1,\ a.s. \tag{7.48}$$

Since $t_{n+1} = \alpha t_n > t_n$, we have

$$\frac{B_{t_{n+1}}}{\sqrt{2t_{n+1}\log\log t_{n+1}}} = \frac{B_{t_{n+1}} - B_{t_n}}{\sqrt{2t_{n+1}\log\log t_{n+1}}} + \frac{1}{\sqrt{\alpha}} \frac{B_{t_n}}{\sqrt{2t_n(\log\log t_n + \log\log\alpha)}}. \tag{7.49}$$

Now, using (7.47) and (7.48), it follows that with probability one,

$$\overline{\lim}_{n\to\infty} \frac{B_{t_{n+1}}}{\varphi(t_{n+1})} \geq \theta - \frac{1}{\sqrt{\alpha}} = \frac{\alpha-1}{\alpha} - \frac{1}{\sqrt{\alpha}}. \tag{7.50}$$

Since $\alpha > 1$ may be selected arbitrarily large, one has with probability one that

$$\overline{\lim}_{t\to\infty} \frac{B_t}{\varphi(t)} \geq \overline{\lim}_{n\to\infty} \frac{B_{t_{n+1}}}{\varphi(t_{n+1})} \geq 1. \tag{7.51}$$

This completes the computation of the limit superior. To get the limit inferior simply replace $\{B_t : t \geq 0\}$ by $\{-B_t : t \geq 0\}$. ■

Remark 7.10. An important functional generalization of Khinchine's LIL (Theorem 7.23) was obtained later by Strassen (1964), stated without proof as follows.

Theorem 7.24. Let $\{X_n : n \geq 1\}$ be an i.i.d. sequence with mean zero and variance one. Then , with probability one, the stochastic process on [0, 1] defined by

$$S_n(t) = \frac{S_i}{\sqrt{2n \log \log n}}, \ t = \frac{i}{n}(i = 1, \ldots, n), \ S_n(0) = 0,$$

and linearly interpolated for $t \in [i/n, (i+1)/n]$, $(i = 0, 1, \ldots, n)$, has its set of limit points (in the uniform topology for $C[0,1]$) given by the set of all absolutely continuous functions f on $[0,1]$ satisfying $f(0) = 0$, and such that $\int_{[0,1]} |f'(t)|^2 dt \leq 1$.

The ordinary LIL is obtained by taking $f(t) = t$, and $f(t) = -t$, respectively, or any absolutely continuous function f satisfying $f(0) = 0$ and $f(1) = \pm 1$. Since $f(1) = \int_{[0,1]} |f'(t)|^2 dt \leq 1$, it follows that the latter functions lead to the lim sup and lim inf, respectively, of $S_n(1)$, as $n \to \infty$.

As another illustrative application of the strong Markov property one may derive a Cantor-like structure of the random set of zeroes of Brownian motion as follows.

Proposition 7.25. With probability one, the set $\mathcal{Z} := \{t \geq 0 : B_t = 0\}$ of zeros of the sample path of a one-dimensional standard Brownian motion, starting at 0, is uncountable, closed, and unbounded. Moreover, $\mathcal{Z}$ a.s. has Lebesgue measure zero.

Proof. Let $A_0 = \{\omega \in \Omega : B_t(\omega) = 0$ for infinitely many t in every interval $[0, \varepsilon]$, $\varepsilon > 0\}$. The law of iterated logarithm (Theorem 7.23) may be applied to the Brownian motion $W_0 = 0$, $W_t = \sqrt{t} B_{\frac{1}{t}}$, $t > 0$, as $t \downarrow 0$ to obtain $P(A_0) = 1$. Since $t \to B_t(\omega)$ is continuous, $\mathcal{Z}(\omega)$ is closed. Applying the LIL as $t \uparrow \infty$, it follows $\mathcal{Z}(\omega)$ is unbounded a.s.

We will now show that, for $0 < c < d$, the probability is zero of the event $A(c,d)$, say, that B has a single zero in $[c,d]$. For this consider the stopping time $\tau := \inf\{t \geq c : B_t = 0\}$. By the strong Markov property, B_τ^+ is a standard Brownian motion, starting at zero. In particular, τ is a point of accumulation of zeros from the right (a.s.). Also, $P(B_d = 0) = 0$. This implies $P(A(c,d)) = 0$. Considering all pairs of rationals c, d with $c < d$, it follows that $\mathcal{Z}$ has no isolated point outside a set of probability zero. Alternatively, let $\mathcal{Z}_1(\omega)$ denote the set of zeros of a sample path ω of the path $t \to B_t(\omega)$ on $0 \leq t \leq 1$. suppose there is a subset $F \in \mathcal{F}$ such that $P(F) = \delta > 0$ and for each $\omega \in F$ there exists an isolated zero $z(\omega)$ of $\mathcal{Z}_1(\omega)$. Then there exists $\ell > 0$ such that the probability of the length of an open interval with center $z(\omega)$, free of other zeros, being larger than ℓ has a positive probability. This would imply if $(0,1)$ is divided into $[3/\ell]$ subintervals of equal length, then, with positive probability, one of these intervals contains isolated zeros of $\mathcal{Z}_1$. Finally, for each $T > 0$ let $H_T = \{(t,\omega) : 0 \leq t \leq T, B_t(\omega) = 0\} \subset [0,T] \times \Omega$. By Fubini's theorem, denoting the Lebesgue measure on $[0,\infty)$ by m, one has

$$\begin{aligned}
(m \times P)(H_T) &= \int_0^T \left\{ \int_\Omega \mathbf{1}_{\{\omega: B_t(\omega)=0\}} P(d\omega) \right\} dt \\
&= \int_0^T P(B_t = 0) dt = 0,
\end{aligned}$$

so that, $m(\{t \in [0, T] : B_t(\omega) = 0\}) = 0$ for P-almost all ω. ■

The following distributions can be obtained as a rather direct consequence of the Markov property for Brownian notion.

***Proposition 7.26** (Inverse Trigonometric Laws for First and Last Zeroes of Brownian Motion).* Let $S := \sup\{t \leq 1 : B_t = 0\}$, and $T :=:= \inf\{t \geq 1 : B_t = 0\}$. Then, S has the arcsine pdf $\frac{1}{\pi\sqrt{s(1-s)}}$, $0 < s < 1$, and T has pdf $\frac{1}{\pi t\sqrt{t-1}}$, $t > 1$.

Proof. Conditionally given $B_u, u \leq s \leq 1$, the event $[S \leq s]$ is equivalent to there being no zero in the time from s to 1 for the process restarting from B_s. So,

$$\begin{aligned}
&P(S \leq s)\\
&\quad= \mathbb{E}P(S \leq s|B_u, u \leq s)\\
&\quad= \mathbb{E}P_{B_s}(\tau_0 > 1 - s)\\
&\quad= \int_{-\infty}^{\infty} P_x(\tau_0 > 1 - s)P(B_s \in dx)\\
&\quad= \int_{-\infty}^{\infty} P_x(\tau_0 > 1 - s)g_s(x)dx\\
&\quad= \frac{1}{\pi}\int_0^s (1-t)^{-\frac{1}{2}}t^{-\frac{1}{2}}dt,
\end{aligned}$$

where $g_s(x)$ is the Gaussian density of B_s having mean zero and variance s, and τ_0 denotes the first passage time to 0 for standard Brownian motion started at x under P_x. By symmetry, the distribution of the hitting time at 0 starting from x coincides with the distribution of the hitting time of x starting from 0, i.e., with pdf $t \to \frac{|x|}{\sqrt{2\pi}}t^{-\frac{3}{2}}e^{-\frac{x^2}{2t}}$, $t > 0$. The details of the calculus are left as Exercise 12. Similarly, for the distribution of T one has

$$\begin{aligned}
&P(T > t)\\
&\quad= \mathbb{E}P(T > t|B_u, u \leq 1)\\
&\quad= \mathbb{E}P_{B_1}(\tau_0 > t - 1)\\
&\quad= \int_{-\infty}^{\infty} P_x(\tau_0 > t - 1)g_1(x)dx\\
&\quad= \int_t^{\infty} \frac{1}{\pi s\sqrt{s-1}}ds.
\end{aligned}$$

■

Remark 7.11. In anticipation of the previously cited Lévy-Skorokhod formula for the present context, one may show that $S^* = argmax B_t, 0 \leq t \leq 1$ is well-defined unique solution to $R_{S^*} = M_{S^*} - B_{S^*} = 0$. So, using the Lévy-Skorokhod formula, the zeroes of R_t are distributed as the zeroes of $|B_t|$ which, in turn are also the zeroes of B_t. Thus S^* is distributed as S, the last zero of B before $t = 1$, and therefore has the arcsine distribution.

The following general consequence of the Markov property can also be useful in the analysis of the (infinitesimal) fine scale structure of Brownian motion and may be viewed as a corollary to Proposition 7.6. As a consequence, for example, one sees that for any given function $\varphi(t)$, $t > 0$, the event

$$D_\varphi := [B_t < \varphi(t) \text{ for all sufficiently small } t] \tag{7.52}$$

will certainly occur or is certain not to occur, i.e., for almost every sample path $\omega \in C[0,\infty)$ there is an $\bar{t}(\omega)$ such that $B_t < \varphi(t)$ for all $t > \bar{t}$. Functions φ for which $P(D_\varphi) = 1$ are said to belong to the *upper class at the origin*;[7] see Exercise 20. Note that by a time-inversion this translates to behavior of Brownian motion at infinity as well; see Exercise 20 for the corresponding notion of *upper class at infinity*.

Proposition 7.27 ***(Blumenthal's Zero–One Law).*** With the notation of Proposition 7.6,

$$P(A) = 0 \text{ or } 1 \qquad \forall\, A \in \mathcal{F}_{0+}. \tag{7.53}$$

Proof. It follows from (the proof of) Proposition 7.6 that $\mathcal{F}_{s+}$ is independent of $\sigma\{B_{t+s} - B_s : t \geq 0\}$ $\forall\, s \geq 0$. Set $s = 0$ to conclude that $\mathcal{F}_{0+}$ is independent of $\sigma(B_t : t \geq 0) \supset \mathcal{F}_{0+}$. Thus $\mathcal{F}_{0+}$ is independent of $\mathcal{F}_{0+}$ so that $\forall\, A \in \mathcal{F}_{0+}$ one has $P(A) \equiv P(A \cap A) = P(A) \cdot P(A)$. ■

Exercises

1. Show that each $\tau_B^{(r)}$ is a stopping time.
2. (i) If τ_1, τ_2 are stopping times show that $\tau_1 \vee \tau_2$ and $\tau_1 \wedge \tau_2$ are stopping times.
   (ii) If $\tau_1 \leq \tau_2$ are stopping times show that $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.
3. In Example 1 prove that $P_x(\tau_x^{(1)} < \infty) = 1$. [*Hint*: Condition on $\sigma(S_0, S_1)$.]
4. Give a proof of (2.5), followed by a proof of Proposition 2.1, using the strong Markov property with stopping time $\tau = \tau_{x-1} \wedge \tau_{x+1}$.
5. Consider a random walk $\{S_n^x : n \geq 0\}$ on the integers starting at an integer x, and let $N_x = \sum_{n=0}^{\infty} \mathbf{1}[S_n^x = x]$ denote the number of returns to x counting $S_0^x = x$.

[7] Some authors will refer to $t^{-\frac{1}{2}}\varphi(t)$ as being upper class in such instances.

(i) Show that $P(N_x = k) = \rho^{k-1}(1-\rho), k = 1, 2, \ldots,$ where ρ is the probability of eventual return to x. [*Hint*: Use the strong Markov property to show N_x counts the number of i.i.d. return cycles to x.]
(ii) Show $\rho = 1$ if and only if $P(N_x = \infty) = 1$.
(iii) Show that $P(S_n^x = x\ i.o.) = 1$ if and only if $\sum_{n=0}^{\infty} P(S_n^x = x) \equiv \mathbb{E}N_x = \infty$.
(iv) Show that $P(S_n^x = x\ i.o.) = 0$ if and only if $\sum_{n=0}^{\infty} P(S_n^x = x) < \infty$.

6. Prove the following for simple asymmetric random walk:

(i) If $p < \frac{1}{2}$, $P_x(\tau_y^{(r)} < \infty) = (2p)^{r-1}(\frac{p}{q})^{y-x}$ if $y > x$, and $P_x(\tau_y^{(r)} < \infty) = (2p)^{r-1}$ if $y < x$, $(r \geq 1)$.
(ii) If $p > \frac{1}{2}$, $P_x(\tau_y^{(r)} < \infty) = (2q)^{r-1}(\frac{q}{p})^{x-y}$ if $y < x$, and $P_x(\tau_y^{(r)} < \infty) = (2q)^{r-1}$ if $y > x$, $(r \geq 1)$. [*Hint*: If $x \neq y$, $P_x(\tau_y^{(1)} < \infty)$ is given by Corollary 2.2. For $r > 1$, use induction based on $P_x(\tau_y^{(r)} < \infty) = P_x(\tau_y^{(r-1)} < \infty, (S^+_{\tau_y^{(r-1)}})_n = y$ for some $n \geq 1)$, and the strong Markov property with stopping time $\tau_y^{(r-1)}$.]

7. Complete the indicated details required in the proof of Corollary 7.5.
8. Verify the continuity of the function φ defined in (7.18) as the result of the Feller continuity condition.
9. Show that the strong Markov property holds for the compound Poisson process by verifying the conditions for Theorem 7.7. [*Hint*: For the Feller property note that all bounded functions on $\mathbb{Z}_+$ are continuous, bounded functions for the discrete topology.]
10. Suppose that X, Y, Z are three random variables with values in arbitrary measurable spaces $(S_i, \mathcal{S}_i), i = 1, 2, 3$, respectively. Assume that regular[8] conditional distributions exist. Show that $\sigma(Z)$ is conditionally independent of $\sigma(X)$ given $\sigma(Y)$ if and only if the conditional distribution of Z given $\sigma(Y)$ a.s. coincides with the conditional distribution of Z given $\sigma(X, Y)$.
11. For standard Brownian motion B starting at zero, fill in the details to show that $M_t - B_t$ is distributed as $|B_t|$.[*Hint*: Consider the linear transformation of (M_t, B_t) to $(M_t, M_t - B_t)$ and use Corollary 7.11.]
12. Complete the calculus details for the proof of Proposition 7.26.
13. In the notation of Theorem 7.10,

(i) Prove that $Y := \{B_{t\wedge\tau_a} : t \geq 0\}$ is $\mathcal{G}_{\tau_a}$–measurable,
(ii) Express $B \equiv \{B_t : t \geq 0\}$ explicitly as a measurable function of Y and $B^+_{\tau_a}$, and express W as the same function of Y and $W^+_{\tau_a}$. [*Hint*: (ii) Let $C^x = \{f \in C[0,\infty), f(0) = x\}$. Fix $a > 0$. On $C^0 \times C^a$ define φ by $\varphi(f, g)(t) = f(t)$ for $t \leq \overline{\tau}_a(f)$ and $\varphi(f, g)(t) = g(t - \overline{\tau}_a(f))$ for $t > \overline{\tau_a}(f)$. Then $B = \varphi(Y, B^+_{\tau_a})$, and $W = \varphi(Y, W^+_{\tau_a})$. Measurability of φ

[8]See BCPT p. 41 for explicit conditions.

follows from lower semicontinuity of $\overline{\mathbb{E}}\tau_a(f)$ as a function from $C[0,\infty)$ into $[0,\infty]$ and the relation $\{\varphi(f,g)(t) \leq x\} = \{\overline{\tau}_a(f) > t,\ f(t) \leq x\} \cup \{\overline{\tau}_a(f) \leq t,\ g(t-\overline{\tau}_a(f)) \leq x\}$.]

14. Derive the joint distribution of (m_t, B_t), where $m_t := \min\{B_s : 0 \leq s \leq t\}$, and $\{B_t : t \geq 0\}$ is standard Brownian motion with $B_0 = 0$. [*Hint*: Use Corollary 7.11 and the fact $-B$ is a one-dimensional standard Brownian motion.]
15. Show that the distribution of $\min_{t\geq 0} X_t^0$ is exponential if $\{X_t^0 : t \geq 0\}$ is Brownian motion starting at 0 with drift $\mu > 0$. Likewise, calculate the distribution of $\max_{t\geq 0} X_t^0$ when $\mu < 0$.
16. (i) Use the SLLN to show that the Brownian motion with nonzero drift is transient.
    (ii) Extend (i) to the k-dimensional Brownian motion with drift.

17. Let $X_t = X_0 + vt$, $t \geq 0$, where v is a nonrandom constant-rate parameter and X_0 is a random variable.

    (i) Calculate the conditional distribution of X_t, given $X_s = x$, for $s < t$.
    (ii) Show that all states are transient if $v \neq 0$ in the sense that $|X_t| \to \infty$ a.s. as $t \to \infty$.
    (iii) Calculate the distribution of X_t if the initial state is normally distributed with mean μ and variance σ^2.
    (iv) Repeat the above when v is a random variable, independent of the initial state X_0.

18. Let $\{X_t : t \geq 0\}$ be a Brownian motion starting at 0 with diffusion coefficient $\sigma^2 > 0$ and zero drift. Define $\{Y_t : t \geq 0\}$ by $Y_t = tX_{1/t}$ for $t > 0$ and $Y_0 = 0$. Recall that $\{Y_t : t \geq 0\}$ is distributed as Brownian motion starting at 0 by the time-inversion property.

    (i) Show that $\{X_t : t \geq 0\}$ has infinitely many zeroes in every neighborhood of $t = 0$ with probability 1.[*Hint*: Use the law of the iterated logarithm to show a.s. $\limsup_{t\to\infty} X_t = +\infty$ and $\liminf_{t\to\infty} X_t = -\infty$.]
    (ii) Show that the probability that $t \to X_t$ has a right-hand derivative at $t = 0$ is zero.

19. Use calculus, with the aid of integration software, to determine $k = 2$ in the formula $\mathbb{E}e^{-\lambda\tau_a} = e^{-|a|\sqrt{k\lambda}}$, $\lambda > 0$, for the Laplace transform of the first passage time in Proposition 7.15. [*Hint*: Consider $\mathbb{E}\tau_1 e^{-\tau_1} = -\frac{d}{d\lambda}\mathbb{E}e^{-\lambda\tau_1}|_{\lambda=1} = \frac{\sqrt{k}}{2}e^{-\sqrt{k}}$. On the other hand, the substitution $s = t^{\frac{1}{2}}$ yields

$$\begin{aligned}\mathbb{E}\tau_1 e^{-\tau_1} &= \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-t}t^{-\frac{1}{2}}e^{-\frac{1}{2t}}\,dt\\ &= \frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{s^2}}e^{-\frac{s^2}{2}}\,ds\end{aligned}$$

$$= \frac{2}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{4} (e^{-\sqrt{2}}(1 - \mathrm{erf}(\frac{1}{\sqrt{2s}} - s)) + e^{-\sqrt{2}}(-1 + \mathrm{erf}(\frac{1}{\sqrt{2s}} + s)))|_0^\infty$$
$$= \frac{2}{\sqrt{2\pi}} \frac{e^{-\sqrt{2}}}{2} \sqrt{\pi}.$$

where this last line used Matlab integration, and $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$ is the *error function*.]

20. (*Kolmogorov's Test for Upper Class Functions*) Suppose φ is a non-negative increasing continuous function such that $t^{-\frac{1}{2}}\varphi(t)$ is a decreasing function.

 (i) Show[9] that for φ to belong to the upper class at the origin (7.52), it is sufficient that

$$\int_0^1 t^{-\frac{3}{2}} \varphi(t) \exp(-\frac{\varphi^2(t)}{2t}) dt < \infty.$$

 [*Hint*: The similarity of the integrand with a first passage time pdf is not accidental. The general idea for the proof is that convergence of $\int_0^1 t^{-\frac{3}{2}} \varphi(t) \exp(-\frac{\varphi^2(t)}{2t}) dt < \infty$ yields a (nonrandom) sequence $\tau_1 > \tau_2 > \dots$, decreasing to zero, such that $\sum_{k=1}^\infty P(\max_{\tau_k \le t \le \tau_{k-1}} B_t > \varphi(\tau_{k-1})) < \infty$. Thus, by the Borel–Cantelli lemma Part I, there is a (random) K such that $B_t < \varphi(t)$ for all $t < \tau_K$. Complete the following steps to provide the details: (a) For a decreasing sequence $\tau_k, k \ge 1$, to be determined, show that $\frac{\max_{\tau_k \le t \le \tau_{k-1}} B_t}{\varphi(t)} \le \frac{\max_{\tau_k \le t \le \tau_{k-1}} B_t}{\varphi(\tau_k)} \le \frac{\max_{0 \le t \le \tau_{k-1}} B_t}{\varphi(\tau_k)}$. (b) Apply the (reflection principle) bound (7.24) on the running maximum, together with scaling and Feller's tail probability estimate for the normal distribution to show $P(\frac{\max_{\tau_k \le t \le \tau_{k-1}} B_t}{\varphi(t)} > 1) \le P(M_{\tau_{k-1}} > \varphi(\tau_k)) \le 2\frac{1}{\sqrt{2\pi}} \frac{\sqrt{\tau_{k-1}}}{\varphi(\tau_k)} e^{-\frac{\varphi^2(\tau_k)}{2\tau_{k-1}}} \le \sqrt{\frac{2}{\pi}} \frac{\sqrt{\frac{\tau_{k-1}}{\tau_k}}}{\tau_{k-1}^{-\frac{1}{2}} \varphi(\tau_{k-1})} e^{-\frac{\frac{\tau_k}{\tau_{k-1}} (\tau_{k-1}^{-\frac{1}{2}} \varphi(\tau_{k-1}))^2}{2}}$. (c) Define $\tau_{k+1} = \frac{\tau_k \varphi^2(\tau_k)}{\tau_k + \varphi^2(\tau_k)}, k \ge 1$. Show that τ_k decreases to 0 as $k \to \infty$, $\frac{\tau_{k-1}}{\tau_k} < 2$, and using the simple algebraic inequality $\frac{a}{a+b} \ge \frac{a-b}{a}, a, b > 0$, $\frac{\tau_k}{\tau_{k-1}} > 1 - \frac{\tau_{k-1}}{\varphi^2(\tau_{k-1})}$. (d) Use part (c) in part (b), together with the monotonicity hypotheses, to show $P(\frac{\max_{\tau_k \le t \le \tau_{k-1}} B_t}{\varphi(t)} > 1) \le \sqrt{\frac{2}{\pi}} \frac{\sqrt{2}}{\tau_{k-1}^{-\frac{1}{2}} \varphi(\tau_{k-1})} e^{\frac{1}{2}} e^{-\frac{(\tau_{k-1}^{-\frac{1}{2}} \varphi(\tau_{k-1}))^2}{2}}$. (e) Finally, again using monotonicities and the definition of τ_k, check that

[9] Although this theorem was first announced by Paul Lévy and attributed to A.N. Kolmogorov, a proof did not appear until (Sirao and Nisida, 1952).

$$\int_{0^+}^{\tau_1} \tau^{-\frac{3}{2}}\varphi(\tau)e^{-\frac{\varphi^2(\tau)}{2\tau}}d\tau = \sum_{k=1}^{\infty}\int_{\tau_{k+1}}^{\tau_k}\frac{1}{\tau}\tau^{-\frac{1}{2}}\varphi(\tau)e^{-\frac{(\tau^{-\frac{1}{2}}\varphi(\tau))^2}{2}}d\tau$$
$$\geq \sum_k \frac{\tau_k-\tau_{k+1}}{\tau_k}\tau_k^{-\frac{1}{2}}\varphi(\tau_k)e^{-\frac{\varphi^2(\tau_k)}{2\tau_k}}$$
$$\geq \sum_k \frac{\sqrt{\tau_k}}{\frac{1}{(\tau_1^{-\frac{1}{2}}\varphi(\tau_1))^2+1}}\frac{1}{\varphi^2(\tau_k)}\varphi(\tau_k)e^{-\frac{\varphi^2(\tau_k)}{2\tau_k}}.$$

Invoke Borel–Cantelli I to complete the proof.]

(ii) Show[10] that if φ belongs to the upper class at the origin then $B_t < t\varphi(\frac{1}{t})$ for all sufficiently large t, i.e., $\varphi(\frac{1}{t}$ is an upper class function at infinity for Brownian motion. [*Hint*: Use time-inversion invariance of Brownian motion.]

(iii) Show that $\varphi(t) = t^\theta$ is an upperclass function at the origin for $0 < \theta < 1/2$.

[10] Hölder continuity of Brownian paths does not extend to exponent $\alpha = 1/2$. An important result of Lévy (1937) shows that the modulus of continuity is only slightly worse, and at resolution δ is given by $\psi(\delta) := \sqrt{2\delta\log(\frac{1}{\delta})}$ in the sense that with probability one

$$\overline{\lim}_{\delta\downarrow 0}\frac{1}{\psi(\delta)}\max_{0\leq t-s\leq\delta}|B_t - B_s| = 1. \tag{7.54}$$

In particular, for every $\epsilon > 0$, the function $\psi(t)(1+\epsilon)$ belongs to the upper class.

Chapter 8
Coupling Methods for Markov Chains and the Renewal Theorem for Lattice Distributions

The coupling method is a powerful tool of stochastic analysis that has enjoyed many successes since its original introduction by Doeblin (1938) to prove convergence to a unique invariant probability for finite state Markov chains. In fact it was applied in Chapter 5 to obtain an error bound in the Poisson approximation to the binomial distribution, i.e., the law of rare events. The convergence to steady state for a class of finite state Markov chains together with a proof of a related powerful result, the renewal theorem, is presented. In the latter one seeks to find how much time a general random walk on the integers with increasing paths, i.e., having non-negative integer-valued displacements, spends in an interval of length m, say $(n, n+m]$. Renewal theory computes the precise amount asymptotically as $n \to \infty$. This chapter is devoted to a cornerstone theorem in the case of integer-valued renewal times, while the much more general theory is provided as a special topic in Chapter 25.

The notion of *coupling* was introduced by Doeblin as a method to prove convergence to a unique invariant probability for irreducible aperiodic finite state Markov chains.

Definition 8.1. A Markov chain on a countable state space S and transition probabilities $p_{ij}, i, j \in S$ is *irreducible* if for every $i, j \in S$ there is a positive integer n such that $p_{ij}^{(n)} > 0$, where $p^{(n)}$ denotes the n-step transition probability matrix. If the greatest common divisor (g.c.d.) d of the set $\{n \geq 1 : p_{ii}^{(n)} > 0\}$ is one for all $i \in S$ then the Markov chain is said to be *aperiodic*. If on the other hand $d > 1$ for all $i \in S$, then the chain is said to be periodic with *period* d.

The long-time behavior of Markov chains is a subject of great interest to theory and applications. Before consideration of Doeblin's ideas, let us record some of the

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_8

most basic properties of an irreducible, aperiodic Markov chain $X = \{X_n : n \geq 0\}$ on a denumerable state space S with one-step transition probabilities $p_{ij}, i, j \in S$.

Definition 8.2. A state $j \in S$ is said to be *recurrent* if $P_j(X_n = j \ i.o.) = 1$, and *transient* if $P_j(X_n = j \ i.o.) = 0$.

The successive *return times* to the state j are defined by

$$\tau_j^{(0)} = 0, \quad \tau_j^{(1)} := \inf\{n > 0 : X_n = j\}, \quad \tau_j^{(r)} = \inf\{n > \tau_j^{(r-i)} : X_n = j\}, \tag{8.1}$$

for $r = 1, 2, \ldots$, with the convention that $\tau_j^{(r)} = \infty$ if there is no $n > \tau_j^{(r-1)}$ for which $X_n = j$. Write

$$\rho_{ij} = P_i(X_n = j \text{ for some } n \geq 1) = P_i(\tau_j^{(1)} < \infty). \tag{8.2}$$

Using the strong Markov property for discrete parameter Markov chains (Theorem 7.2), it follows that

$$\begin{aligned} P_i(\tau_j^{(r)} < \infty) &= P_i(\tau_j^{(r-1)} < \infty \text{ and } X_{\tau_j^{(r-1)}+n} = j \text{ for some } n \geq 1) \\ &= \mathbb{E}_i(\mathbf{1}_{[\tau_j^{(r-1)} < \infty]} P_{X_{\tau_j^{(r-1)}}}(X_n = j \text{ for some } n \geq 1)) \\ &= \mathbb{E}_i(\mathbf{1}_{[\tau_j^{(r-1)} < \infty]})\rho_{jj} = P_i(\tau_j^{(r-1)} < \infty)\rho_{jj}. \end{aligned} \tag{8.3}$$

Therefore, by iteration,

$$P_i(\tau_j^{(r)} < \infty) = P_i(\tau_j^{(1)} < \infty)\rho_{jj}^{r-1} = \rho_{ij}\rho_{jj}^{r-1} \qquad (r = 2, 3, \ldots). \tag{8.4}$$

In particular, with $i = j$,

$$P_j(\tau_j^{(r)} < \infty) = \rho_{jj}^r \qquad (r = 1, 2, 3, \ldots). \tag{8.5}$$

Now

$$P_j(X_n = j \text{ i.o.}) = P_j(\cap_{r=1}^{\infty}[\tau_j^{(r)} < \infty]) = \lim_{r\to\infty} P_j(\tau_j^{(r)} < \infty) = \begin{cases} 1 & \text{if } \rho_{jj} = 1. \\ 0 & \text{if } \rho_{jj} < 1. \end{cases} \tag{8.6}$$

Further, write $N(j) \equiv \sum_{n=0}^{\infty} \mathbf{1}_{[X_n=j]}$ for the *number of visits to the state* j by the Markov chain $\{X_n\}_{n\geq 0}$, and denote its expected value by

$$G(i, j) = \mathbb{E}_i N(j) = \sum_{n=0}^{\infty} p_{ij}^{(n)}. \tag{8.7}$$

$G(i, j)$ is also referred to as the (discrete parameter) *Green's function* of the Markov chain. Also, if $i \in S$ and $p_{ij}^{(n)} > 0$ for some $n \geq 1$, denoted $i \to j$, then we say that j is *accessible* from i. Now using (8.4)

$$\mathbb{E}_i N(j) = \sum_{r=0}^{\infty} P_i(N(j) > r) = \delta_{ij} + \sum_{r=0}^{\infty} P_i(\tau_j^{(r+1)} < \infty) = \delta_{ij} + \rho_{ij} \sum_{r=0}^{\infty} \rho_{jj}^r, \tag{8.8}$$

where δ_{ij} is 1 or 0 according as $i = j$ or $i \neq j$. Thus,

$$G(i, j) = \begin{cases} \delta_{ij} & \text{if } i \not\to j, \text{ i.e., } \rho_{ij} = 0, \\ \delta_{ij} + \rho_{ij}/(1 - \rho_{jj}) & \text{if } i \to j \text{ and } \rho_{jj} < 1, \\ \infty & \text{if } i \to j \text{ and } \rho_{jj} = 1. \end{cases} \tag{8.9}$$

This calculation provides two useful characterizations of recurrence. One is in terms of the long-run expected number of returns and the other in terms of the probability of eventual return.

Theorem 8.1.

a. Every state is either recurrent or transient. A state j is recurrent iff $\rho_{jj} = 1$ iff $G(j, j) = \infty$, and transient iff $\rho_{jj} < 1$ iff $G(j, j) \equiv (1 - \rho_{jj})^{-1} < \infty$. If j is transient $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for all i.

b. If i is recurrent and $p_{ij}^{(n)} > 0$ for some $n \geq 1$, denoted $i \to j$, then j is recurrent, and $\rho_{ij} = \rho_{ji} = 1$. In particular, if for every $i, j \in S$, $i \to j$ and $j \to i$, then either every state is recurrent, or every state is transient.

c. Let i be recurrent, and $S(i) := \{j \in S : i \to j\}$. Let $\bar{\pi}$ be a probability distribution on $S(i)$. Then

$$P_{\bar{\pi}}(X_n \text{ visits every state in } S(i) \text{ i.o.}) = 1. \tag{8.10}$$

Proof. Part (a) follows from (8.6), (8.7), (8.9). For part (b), suppose i is recurrent and $i \to j (j \neq i)$. Let A_r denote the event that the Markov chain visits j between the r-th and $(r + 1)$st visits to state i. Then under P_i, $A_r (r \geq 0)$ are independent events and have the same probability θ, say. Now $\theta > 0$. For if $\theta = 0$, then $P_i(X_n = j \text{ for some } n \geq 1) = P_i(\bigcup_{r \geq 0} A_r) = 0$, contradicting $i \to j$. It now follows from the second half of the Borel–Cantelli Lemma that $P_i(A_r \text{ i.o.}) = 1$. This implies $G(i, j) = \infty$ and hence, by (8.9), $\rho_{jj} = 1$. Hence j is recurrent. Also, $\rho_{ij} \geq P_i(A_r \text{ i.o.}) = 1$. By the same argument, $\rho_{ji} = 1$. Note that $G(j, j) = 1 + \rho_{jj}/(1 - \rho_{jj}) = 1/(1 - \rho_{jj})$ for transient states $j \in S$.

To prove part (c) use part (b) to get for arbitrary $i \in S(j)$,

$$P_{\bar{\pi}}(X_n \text{ visits } i \text{ i.o.}) \sum_{k \in S(j)} \bar{\pi}_k P_k(X_n \text{ visits } i \text{ i.o.}) \sum_{k \in S(j)} \bar{\pi}_k = 1. \tag{8.11}$$

Hence

$$P_{\bar{\pi}}\left(\bigcap_{i\in S(j)}[X_n \text{ visits } i \text{ i.o.}]\right)=1. \tag{8.12}$$

■

Consider a Markov chain on a (countable) state space S having a transition probability matrix $\mathbf{p}$ such that there is a probability $\boldsymbol{\pi}$ on S such that

$$\pi_j=\sum_{j\in S}\pi_i p_{ij},\quad \forall j\in S. \tag{8.13}$$

Then $\boldsymbol{\pi}$ is referred to as an *invariant probability* for the transition probabilities $\mathbf{p}$ and/or for the Markov chain X.

Note that if $\boldsymbol{\pi}$ is an invariant probability for $\mathbf{p}$, then then $\pi_j=\sum_{i\in S}\pi_i p_{ij}^{(n)},\ n=1,2,\dots$. Moreover, if X has initial distribution $\boldsymbol{\pi}$, then $(X_0,X_1,\dots)$ is a stationary process as defined in Definition 1.2, (Exercise 2).

Corollary 8.2. Suppose X has an invariant probability $\boldsymbol{\pi}$ on S, such that $\pi_j>0$ for all j. Then X is recurrent.

Proof. Assume $\pi_j>0$ but j is transient. This immediately leads to a contradiction since $\mathbb{E}_i N(j)=\sum_n p_{ij}^{(n)}<\infty$ implies $p_{ij}^{(n)}\to 0$ as $n\to\infty$, and therefore $0<\pi_j=\sum_{i\in S}\pi_i p_{ij}^{(n)}\to 0$ as $n\to\infty$. ■

Remark 8.1. If $\boldsymbol{\pi}$ is an invariant probability for an irreducible $\mathbf{p}$, then $\pi_j>0\,\forall j$. For if $\pi_j>0$ for some j, the relation $\pi_{j'}=\sum_{i\in S}\pi_i p_{ij'}^{(n)}\geq \pi_j p_{jj'}^{(n)}\,\forall n$ implies $\pi_{j'}>0$.

Construct a probability space $(\Omega,\mathcal{F},P)$ on which are defined two Markov chains $\{X_n^{(1)}:n\geq 0\}$ and $\{X_n^{(2)}:n\geq 0\}$ with initial distributions μ and γ, respectively. Define

$$T:=\inf\{n\geq 0: X_n^{(1)}=X_n^{(2)}\}, \tag{8.14}$$

and

$$Y_n^{(1)}:=\begin{cases}X_n^{(1)} & \text{if } T>n,\\ X_n^{(2)} & \text{if } T\leq n.\end{cases} \tag{8.15}$$

Since T is a *stopping time* for the Markov process on $S\times S$ defined by $\{\mathbf{X}_n:=(X_n^{(1)},X_n^{(2)}):n\geq 0\}$, (w.r.t the filtration $\mathcal{F}_n:=\sigma\{\mathbf{X}_j:0\leq j\leq n\},\ n\geq 0$), it follows that the *coupled process* defined by $\{Y_n^{(1)}:n\geq 0\}$ has the same distribution as $\{X_n^{(1)}:n\geq 0\}$ (Exercise 1).

Definition 8.3. Let Z_1, Z_2 be two random maps defined on probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{S})$. A coupling of Z_1 and Z_2, or a coupling of their respective distributions P_{Z_1} and P_{Z_2}, is any bivariate random map $(X^{(1)}, X^{(2)})$ with values in $(S \times S, \mathcal{S} \otimes \mathcal{S})$ whose marginals coincide with P_{Z_1} and P_{Z_2}, respectively.

In particular, the pair $\{(X_n^{(1)}, X_n^{(2)}) : n \geq 0\}$ constructed above defines a coupling of (the distributions of) Markov processes with transition probabilities $\mathbf{p}$ having respective initial distributions μ and γ, i.e., a coupling of the distributions P_μ, P_γ. This coupling is used to define the coupled process $\{Y_n^{(1)} : n \geq 0\}$ having distribution P_μ. Other couplings are illustrated in the exercises.

Lemma 1 *(A Coupling Lemma).*

$$|P_\mu(X_n \in B) - \gamma(B)| \leq P(T > n), n \geq 0. \tag{8.16}$$

Proof. Since $Y_n^{(1)} = X_n^{(2)}$ for $n \geq T$ one has, for all $B \subset S$,

$$\begin{aligned}
|P(X_n^{(1)} \in B) - P(X_n^{(2)} \in B)| &= |P(Y_n^{(1)} \in B) - P(X_n^{(2)} \in B)| \\
&\leq |P(Y_n^{(1)} \in B, n \geq T) - P(X_n^{(2)} \in B, n \geq T)| \\
&\quad + |P(Y_n^{(1)} \in B, T > n) - P(X_n^{(2)} \in B, T > n)| \\
&= |P(Y_n^{(1)} \in B, T > n) - P(X_n^{(2)} \in B, T > n)| \\
&\leq P(T > n).
\end{aligned}$$

■

Definition 8.4. If $T < \infty$ a.s. one says that the *coupling is successful*,

It follows from the coupling lemma that if the Markov chain has a unique invariant probability π, then the distribution of $X_n^{(1)}$ converges in total variation distance to π, as $n \to \infty$ if the coupling can be shown to be successful. In particular, if $\mu = \delta_i$ (and $T < \infty$ a.s.) one has

$$\sum_{j \in S} |p_{ij}^{(n)} - \pi_j| \longrightarrow 0 \quad \text{as } n \to \infty. \tag{8.17}$$

Indeed, one can prove a stronger result, if $T < \infty$ a.s. To state this let $Q^{(n)}$ denote the distribution (on $(S^\infty, \mathcal{S}^{\otimes\infty})$) of the Markov process $(X_n^{(1)})^+ := \{X_{n+m}^{(1)} : m \geq 0\}$ and Q_π that of $X^{(2)} = \{X_n^{(2)} : n \geq 0\}$. Here $\mathcal{S}$ is the class of all subsets of S and $\mathcal{S}^{\otimes\infty}$ is the product $\sigma-$field on the set S^∞ of all sequences in S. Then one has

$$\sup_{A \in \mathcal{S}^{\otimes\infty}} |Q^{(n)}(A) - Q_\pi(A)| \longrightarrow 0 \quad \text{as} \quad n \to \infty. \tag{8.18}$$

To prove this simply replace $X_n^{(1)}$, $X_n^{(2)}$ in (8.17) by the after–n processes $(X_n^{(1)})^+$, $(X_n^{(2)})^+$, respectively, and replace $B \in \mathcal{S}$ by $A \in \mathcal{S}^{\otimes\infty}$. The following result makes use of a successful coupling. We use the same notation as above.

***Theorem 8.3** (Convergence of Irreducible Aperiodic Markov Chains).* Let $\mathbf{p}$ be an irreducible aperiodic transition probability matrix on a countable state space S. If $\mathbf{p}$ admits an invariant probability $\boldsymbol{\pi}$ then (8.18) holds for the distribution $Q^{(n)}$ of the after–n process, whatever the initial distribution μ. In particular (8.17) holds for every i, and the invariant probability is unique.

First we need a couple of simple but useful lemmas.

Lemma 2. Let $\mathbf{p}$ be an irreducible aperiodic transition probability matrix on a countable state space S.

a. Then, for each pair (i, j) there exists an integer $\nu(i, j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq \nu(i, j)$.
b. If S is finite there exists ν_0 such that $p_{ij}^{(n)} > 0 \; \forall\, i, j$, if $n \geq \nu_0$.

Proof.

a. Let $B_{ij} = \{\nu \geq 1 : p_{ij}^{(\nu)} > 0\}$. For each j, B_{jj} is closed under addition, since $p_{jj}^{(\nu_1+\nu_2)} \geq p_{jj}^{(\nu_1)} p_{jj}^{(\nu_2)}$. By hypothesis, the greatest common divisor (g.c.d.) of B_{jj} is 1. We now argue that, if B is a set of positive integers closed under addition, and has g.c.d. 1 then the smallest subgroup G of $\mathbb{Z}$ (a *group* under addition) that contains B is $\mathbb{Z}$. Note that G equals $\{u - v : u, v \in B\}$. If G does not equal $\mathbb{Z}$, then $1 \notin G$, so that $G = \{rn : n \in \mathbb{Z}\}$ for some $r > 1$. But, since $B \subset G$, this would imply that the g.c.d. of $B \geq r$, a contradiction.

We have shown that $1 \in G$, i.e., there exists an integer $b \geq 1$ such that $b + 1, b$ both belong to B_{jj}. Let $\nu_j = (2b + 1)^2$. If $n \geq \nu_j$, one may write $n = q(2b + 1) + r$, where r and q are integers, $0 \leq r < 2b + 1$, and $q \geq 2b + 1$. Then $n = q\{b+b+1\}+r\{b+1-b\} = (q-r)b+(q+r)(b+1) \in B$. Thus $b_{jj}^{(n)} > 0$ for all $n \geq \nu_j$. Find $k \equiv k_{ij}$ such that $p_{ij}^{(k)} > 0$ then $p_{ij}^{(n+k)} \geq p_{ij}^{(k)} p_{jj}^{(n)} > 0$ for all $n \geq \nu_j$. Now take $\nu(i, j) = k_{ij} + \nu_j$.

b. If S is finite, let $\nu_0 = \max\{\nu_j + k_{ij} : i, j \in S\}$. Then, for all i, j, one has $p_{ij}^{(n)} > 0$ provided $\nu \geq \nu_0$.

■

Lemma 3. Let $\mathbf{p} = ((p_{ij}))$ be an irreducible aperiodic transition probability matrix on a (countable) state space S, which admits an invariant probability $\boldsymbol{\pi}$. Then the Markov chain $\mathbf{X} = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ on $S \times S$ formed by two independent Markov chains $\mathbf{X}^{(i)} = \{X_n^{(i)} : n \geq 0\}$, $i = 1, 2$, each having transition probability $\mathbf{p}$, is irreducible, aperiodic and has an invariant probability $\boldsymbol{\pi} \times \boldsymbol{\pi}$. In particular, $\mathbf{X}$ is recurrent.

Proof. The transition probabilities of $\mathbf{X}$ are given by

$$\begin{aligned} r((i,i');(j,j')) &:= P\left(X_{n+1}^{(1)} = j, X_{n+1}^{(2)} = j' \mid X_n^{(1)} = i, X_n^{(2)} = i'\right) \\ &= p_{ij}\,p_{i'j'} \qquad ((i,i'),(j,j') \in S \times S). \end{aligned}$$

It is simple to check that $\boldsymbol{\pi} \times \boldsymbol{\pi}$ is an invariant probability for $\mathbf{X}$. Also, by Lemma 2, there exist integers $\nu(i,j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq \nu(i,j)$. Since for all $n \geq \max\{\nu(i,j), \nu(i',j')\} \equiv \gamma(i,i';j,j')$, say, the n-step transition probability $r^{(n)}((i,i');(j,j'))$ is at least $p_{ij}^{(n)} p_{i',j'}^{(n)} > 0$, the Markov chain $\mathbf{X}$ is irreducible and aperiodic. By Corollary 8.2 and Remark 8.1, $\mathbf{X}$ is recurrent. ■

Proof of Theorem 8.3. Using the notation of Lemma 3, let $\mathbf{X} = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ denote the Markov chain on $S \times S$, with $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ independent each with transition probability matrix $\mathbf{p}$, but $\mathbf{X}^{(1)}$, having an (arbitrary) initial distribution μ and $\mathbf{X}^{(2)}$ having the invariant initial distribution $\boldsymbol{\pi}$. By recurrence of $\mathbf{X} \equiv (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$, the first passage time $\tau_{(j,j)}$ of $\mathbf{X}$ to (j,j) is finite a.s., for every $j \in S$. Therefore, defining T as in (8.14)

$$T \leq \tau_{(j,j)} < \infty \text{ a.s.} \tag{8.19}$$

As argued earlier, (8.17), (8.18) follows from (8.19). ■

We now turn to another important application of coupling, namely for a proof of the renewal theorem for lattice random variables. Consider a sequence $\{Y_n : n \geq 1\}$ of i.i.d. positive integer-valued random variables, and Y_0 a non-negative integer-valued random variable independent of $\{Y_n : n \geq 1\}$. The partial sum process $S_0 = Y_0$, $S_k = Y_0 + \cdots + Y_k$ $(k \geq 1)$ defines the so-called *delayed renewal process* $\{N_n : n \geq 1\}$:

$$N_n := \inf\{k \geq 0 : S_k \geq n\} \qquad (n = 0, 1, \ldots), \tag{8.20}$$

with Y_0 as the *delay* and its distribution as the *delay distribution*. In the case $Y_0 \equiv 0$, $\{N_n : n \geq 0\}$ is simply referred to as a *renewal process*. This nomenclature is motivated by classical renewal theory in which components subject to failure (e.g., light bulbs) are instantly replaced upon failure, and $Y_1, Y_2, \ldots,$ represent the random durations or *lifetimes* of the successive replacements. The delay random variable Y_0 represents the length of time remaining in the life of the initial component with respect to some specified time origin. For a special context, again consider a Markov chain $\{X_n : n \geq 0\}$ on a (countable) state space S, and fix a state y such that the first passage time $\tau_y \equiv \tau_y^{(0)}$ to state y is finite a.s., as are the successive return times $\tau_y^{(k)}$ to y $(k \geq 1)$. By the strong Markov property, $Y_0 := \tau_y^{(0)}$, $Y_k := \tau_y^{(k)} - \tau_y^{(k-1)}$ $(k \geq 1)$, are independent, with $\{Y_k : k \geq 1\}$ i.i.d. The renewal process (8.20) may now be defined with $S_k = \tau_y^{(k)}$ $(k \geq 1)$, the kth *renewal* occurring at time $\tau_y^{(k)}$ if $\tau_y^{(0)} = 0$, i.e., y is the initial state.

It may be helpful to think of the partial sum process $\{S_k : k \geq 0\}$ to be a *point process* $0 \leq S_0 < S_1 < S_2 < \cdots$ realized as a randomly located increasing sequence of *renewal times* or *renewal epochs* on $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$. By the definition (8.20) the processes $\{S_k : k \geq 0\}$ and $\{N_n : n \geq 0\}$ may be thought of as (approximate) *inverse functions* of each other on $\mathbb{Z}_+$. Another related process of interest is the *residual (remaining) life process* defined by

$$R_n := S_{N_n} - n = \inf\{S_k - n : k \text{ such that } S_k - n \geq 0\}. \tag{8.21}$$

Remark 8.2. The indices n for which $R_n = 0$ are precisely the renewal times, including the delay. In general $S_{N_n - 1} < n \leq S_{N_n}$ and when the renewal process is viewed at time n, R_n is the residual time until the next renewal; also see Exercises 11, 12.

Proposition 8.4. Let f be the probability mass function of Y_1. Then the following hold:

a. $\{R_n : n \geq 0\}$ is a Markov chain on $S = \mathbb{Z}_+$ with the transition probability matrix $\mathbf{p} = ((p_{i,j}))$ given by

$$p_{0,j} = f(j+1) \quad \text{for } j \geq 0, \quad p_{j,j-1} = 1 \quad \text{for } j \geq 1. \tag{8.22}$$

b. If $\mu \equiv \mathbb{E}Y_1 < \infty$, then there exists a unique invariant probability $\boldsymbol{\pi} = \{\pi_j : j \geq 0\}$ for $\{R_n : n \geq 0\}$ with probability mass function given by the tail probability distribution function for f:

$$\pi_j = \sum_{i=j+1}^{\infty} f(i)/\mu \qquad (j = 0, 1, 2, \cdots). \tag{8.23}$$

c. If $\mathbb{E}Y_1 = \infty$, then $\overline{\pi}_j = \sum_{i=j+1}^{\infty} f(i)$ $(j \geq 0)$ provides an invariant measure $\overline{\boldsymbol{\pi}} = \{\overline{\pi}_j : j \geq 0\}$ for $\mathbf{p}$, which is unique up to a multiplicative constant.

Proof.

a. Although the result is clear informally (Exercise 7), here is a formal argument. Observe that $\{N_n : n \geq 0\}$ are $\{\mathcal{F}_k : k \geq 0\}$–stopping times where $\mathcal{F}_k = \sigma\{Y_j : 0 \leq j \leq k\}$. We will first show that $V_n := S_{N_n}$ $(n \geq 0)$ has the (inhomogeneous) Markov property. For this note that if $S_{N_n} > n$ then $N_{n+1} = N_n$ and $S_{N_{n+1}} = S_{N_n}$, and if $S_{N_n} = n$ then $N_{n+1} = N_n + 1$ and $S_{N_{n+1}} = S_{N_n} + Y_{N_n+1}$. Hence

$$\begin{aligned} S_{N_{n+1}} &= S_{N_n} \mathbf{1}_{[S_{N_n} > n]} + (S_{N_n} + Y_{N_n+1}) \mathbf{1}_{[S_{N_n} = n]} \\ &= S_{N_n} \mathbf{1}_{[S_{N_n} > n]} + (S^+_{N_n})_1 \mathbf{1}_{[S_{N_n} = n]}, \end{aligned} \tag{8.24}$$

where $S^+_{N_n}$ is the after–N_n process $\{(S^+_{N_n})_k := S_{N_n+k} : k = 0, 1, 2, \ldots\}$. It follows from (8.24) and the strong Markov property that the conditional distribution of $V_{n+1} \equiv S_{N_{n+1}}$, given the pre–$N_n$ σ–field $\mathcal{G}_n \equiv \mathcal{F}_{N_n}$, depends only on $V_n = S_{N_n}$. Since V_n is $\mathcal{G}_n$–measurable, it follows that $\{V_n : n \geq 0\}$ has the Markov property, and that its time-dependent transition probabilities are

$$q_n(n, n+j) \equiv P(V_{n+1} = n+j \mid V_n = n) = P(Y_1 = j) = f(j), \quad j \geq 1,$$
$$q_n(m, m) = 1, \qquad m > n. \tag{8.25}$$

Since $R_n = V_n - n$ $(n \geq 0)$, $\{R_n : n \geq 0\}$ has the Markov property, and its transition probabilities are $P(R_{n+1} = j \mid R_n = 0) \equiv P(V_{n+1} = n+1+j \mid V_n = n) = f(j+1)$ $(j \geq 0)$, $P(R_{n+1} = j-1 \mid R_n = j) \equiv P(V_{n+1} = n+j \mid V_n = n+j) = 1$ $(j \geq 1)$. Thus $\{R_n : n \geq 0\}$ is a time-homogeneous Markov process on $S = \mathbb{Z}_+$ with transition probabilities given by (8.22).

b. Assume $\mu \equiv \mathbb{E}Y_1 < \infty$. If $\boldsymbol{\pi} = \{\pi_j : j \geq 0\}$ is an invariant probability for $\mathbf{p}$, then one must have

$$\pi_0 = \sum_{j=0}^{\infty} \pi_j p_{j,0} = \pi_0 p_{0,0} + \pi_1 p_{1,0} = \pi_0 f(1) + \pi_1,$$

$$\pi_i = \sum_{j=0}^{\infty} \pi_j p_{j,i} = \pi_0 p_{0,i} + \pi_{i+1} p_{i+1,i} = \pi_0 f(i+1) + \pi_{i+1} \quad (i \geq 1).$$

Thus

$$\pi_i - \pi_{i+1} = \pi_0 f(i+1) \qquad (i \geq 0). \tag{8.26}$$

Summing (8.26) over $i = 0, 1, \ldots, j-1$, one gets $\pi_0 - \pi_j = \pi_0 \sum_{i=1}^{j} f(i)$, or, $\pi_j = \pi_0 \sum_{i=j+1}^{\infty} f(i)$ $(j \geq 0)$. Summing over j one finally obtains $\pi_0 = 1/\mu$, since $\sum_{j=0}^{\infty}(1 - F(j)) = \mu$, with $F(j) = \sum_{i=0}^{j} f(i)$.

c. If $\overline{\boldsymbol{\pi}} = \{\overline{\pi}_j : j \geq 0\}$ is an invariant measure for $\mathbf{p}$, then it satisfies (8.26), with $\overline{\pi}_i$ replacing π_i $(i \geq 0)$. Hence, by the computation above, $\overline{\pi}_j = \overline{\pi}_0(1 - F(j))$ $(j \geq 0)$. One may choose $\overline{\pi}_0 > 0$ arbitrarily, for example, $\overline{\pi}_0 = 1$.

■

Thus the residual lifetime process $\{R_n : n \geq 0\}$ is a stationary Markov process if and only if (i) $\mu \equiv \mathbb{E}Y_1 < \infty$ and (ii) the delay distribution is given by $\boldsymbol{\pi}$ in (8.23).

Proposition 8.4 plays a crucial role in providing a *successful coupling* for a proof of the renewal theorem below. Define the *lattice span* d of the probability mass function (pmf) f on the set $\mathbb{N}$ of natural numbers as the greatest common divisor (g.c.d) of $\{j \geq 1 : f(j) > 0\}$.

Theorem 8.5 *(Erdös-Feller-Pollard Renewal Theorem).* Let the common pmf f of Y_j on $\mathbb{N}$ $(j \geq 1)$ have span 1 and $\mu \equiv EY_j < \infty$. Then whatever the (delay) distribution of Y_0 one has for every positive integer m,

$$\lim_{n\to\infty} \mathbb{E}(N_{n+m} - N_n) = \frac{m}{\mu}\,. \tag{8.27}$$

Proof. On a common probability space $(\Omega, \mathcal{F}, P)$ construct two independent sequences of random variables $\{Y_0, Y_1, Y_2, \cdots\}$ and $\{\widetilde{Y}_0, \widetilde{Y}_1, \widetilde{Y}_2, \cdots\}$ such that (i) Y_k $(k \geq 1)$ are i.i.d. with common pmf f and the same holds for $\widetilde{Y}_k$ $(k \geq 1)$, (ii) Y_0 is independent of $\{Y_k : k \geq 1\}$ and has an arbitrary delay distribution, while $\widetilde{Y}_0$ is independent of $\{\widetilde{Y}_k : k \geq 1\}$ and has the equilibrium delay distribution $\boldsymbol{\pi}$ of (8.23). Let $\{S_k : k \geq 0\}$, $\{\widetilde{S}_k : k \geq 0\}$ be the partial sum processes, and $\{N_n : n \geq 0\}$, $\{\widetilde{N}_n n \geq 0\}$ the renewal processes, corresponding to $\{Y_k : k \geq 0\}$ and $\{\widetilde{Y}_k : k \geq 0\}$, respectively. The residual lifetime processes $\{R_n := S_{N_n} - n : n \geq 0\}$ and $\{\widetilde{R}_n := \widetilde{S}_{\widetilde{N}_n} - n : n \geq 0\}$ are independent Markov chains on $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ each with transition probabilities given by (8.22). Since the span of f is one, it is simple to check that these are aperiodic and irreducible (see the proof of Lemma 2 and Exercise 8). Hence by Lemma 3, preceding the proof of Theorem 8.3, the Markov chain $\{(R_n, \widetilde{R}_n) : n \geq 0\}$ on $\mathbb{Z}_+ \times \mathbb{Z}_+$ is recurrent, so that

$$T := \inf\{n \geq 0 : (R_n, \widetilde{R}_n) = (0, 0)\} < \infty \;\text{ a.s.} \tag{8.28}$$

Define

$$R'_n := \begin{cases} R_n & \text{if } T > n, \\ \widetilde{R}_n & \text{if } T \leq n. \end{cases} \tag{8.29}$$

Then $\{R'_n : n \geq 0\}$ has the same distribution as $\{R_n : n \geq 0\}$. Note also that

$$N_{n+m} - N_n = \sum_{j=n+1}^{n+m} \mathbf{1}_{[R_j=0]}, \quad \widetilde{N}_{n+m} - \widetilde{N}_n = \sum_{j=n+1}^{n+m} \mathbf{1}_{[\widetilde{R}_j=0]}, \tag{8.30}$$

$$\mathbb{E}(\widetilde{N}_{n+m} - \widetilde{N}_n) = \sum_{j=n+1}^{n+m} P(\widetilde{R}_j = 0) = m\pi_0 = m/\mu. \tag{8.31}$$

Now

$$\begin{aligned} &\mathbb{E}(N_{n+m} - N_n) \\ &\quad = \mathbb{E}(N_{n+m} - N_n)\mathbf{1}_{[T>n]} + \mathbb{E}(\widetilde{N}_{n+m} - \widetilde{N}_n)\mathbf{1}_{[T\leq n]} \\ &\quad = \mathbb{E}(N_{n+m} - N_n)\mathbf{1}_{[T>n]} - \mathbb{E}(\widetilde{N}_{n+m} - \widetilde{N}_n)\mathbf{1}_{[T>n]} + \mathbb{E}(\widetilde{N}_{n+m} - \widetilde{N}_n). \end{aligned}$$

Since the first two terms on the right side of the last equality are each bounded by $mP(T > n) \to 0$ as $n \to \infty$, the proof is complete. ■

Corollary 8.6. If f has a lattice span $d > 1$ and $\mu \equiv EY_1 < \infty$ then, whatever the delay distribution on the lattice $\{kd : k = 1, 2, \ldots\}$,

$$\lim_{n\to\infty} \mathbb{E}(N_{nd+md} - N_{nd}) = \frac{md}{\mu} \qquad (m = 1, 2, \cdots). \tag{8.32}$$

Proof. Consider the renewal process for the sequence of lifetimes $\{Y_k/d : k = 1, 2, \cdots\}$, and apply Theorem 8.5, noting that $EY_1/d = \mu/d$. ■

For another perspective on renewal theory, by conditioning on Y_1 and noting that $S_n - Y_1, n \geq 1$ is also an ordinary renewal process with the same inter-arrival distribution f, it follows that if $S_n = \sum_{j=0}^{n} Y_j, n = 0, 1, \ldots$ is an ordinary ($Y_0 = 0$) renewal process, then the *renewal measure* defined by

$$u(k) = \sum_{n=0}^{\infty} P(S_n = k) \equiv \mathbb{E}\sum_{n=0}^{\infty} \mathbf{1}[S_n = k], \quad k = 0, 1, \ldots, \tag{8.33}$$

solves the equation

$$u(k) = \delta_0(k) + \sum_{j=0}^{k} f(j)u(k-j), k = 0, 1, 2, \ldots, \tag{8.34}$$

where $f(j) = P(Y_1 = j), j = 0, 1, 2, \ldots$ (Exercise 6). The equation (8.34) is a special case of the so-called *renewal equation*

$$v(k) = g(k) + \sum_{j=0}^{k} f(j)v(k-j), \quad k = 0, 1, 2, \ldots, \tag{8.35}$$

where $g = \{g(k) : k = 0, 1, \ldots\} \in \ell_\infty$, and $f = \{f(j) : j = 0, 1, \ldots\}$ is a probability mass function.

Exercises

1. (a) Suppose $\mathbf{p}$ is an irreducible periodic transition probability matrix on a countable state space S (of period $d > 1$). Then the Markov chain $\{\mathbf{X}_n := (X_n^{(1)}, X^{(2)}) : n \geq 0\}$ with $\{X_n^{(i)} : n \geq 0\}, i = 1, 2$, independent Markov chains each with transition probability $\mathbf{p}$, then $\{\mathbf{X}_n : n \geq 0\}$ is not irreducible, and has d equivalence classes.
   (b) Example:

$$\mathbf{p} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad (d = 2)$$

2. (a) Show that if $\boldsymbol{\pi}$ is an invariant probability for $\mathbf{p}$, then $\pi_j = \sum_{i \in S} \pi_i p_{ij}^{(n)}, \quad \forall n = 1, 2, \ldots$.
   (b) Show that if X has initial distribution $\boldsymbol{\pi}$ then $(X_0, X_1, \ldots)$ is a stationary process as defined in Definition 1.2.
3. Show that the product probability measure $P \times Q$ always provides a trivial coupling of two distributions P,Q on $(S, \mathcal{S})$.
4. (*Monotone Coupling*) Let X_1 and X_2 be Bernoulli $0-1$-valued random variables with $P(X_i = 1) = p_i, i = 1, 2$ with $p_1 < p_2$. Let U be a uniformly distributed random variable on [0, 1]. Let $Y_i = \mathbf{1}_{[U \le p_i]}, i = 1, 2$. Compute the distribution of (Y_1, Y_2) and check that (Y_1, Y_2) is a coupling of (X_1, X_2) such that $Y_1 \le Y_2$.
5. (*Maximal Coupling*)[1] Suppose that μ_1, μ_2 are probability measures defined on the power set of a finite set S. Let $\mathcal{C}$ denote the set of all couplings (X_1, X_2) of μ_1, μ_2. Show that $||\mu_1 - \mu_2||_{TV} = \inf_{(X_1, X_2) \in \mathcal{C}} P(X_1 \neq X_2)$, where $|| \cdot ||_{TV}$ denotes the total variation norm.[2][*Hint*: Show that the infimum is achieved by the coupling (X_1^*, X_2^*) defined as follows: Let $S_1 = \{x \in S : \mu_1(x) > \mu_2(x)\}$, $S_2 = S_1^c$, $p_i^* = \sum_{x \in S_i} |\mu_1 - \mu_2|$, $p^* = p_1^* + p_2^*$. With probability p^*, choose a value $X_1^* = X_2^* = x$ from the distribution $\frac{1}{p^*}\mu_1(x) \wedge \mu_2(x)$ or, with probability $1 - p^*$ choose a value X_1^* from the distribution $\frac{1}{1-p^*}(\mu_1(x) - \mu_2(x)), x \in S_1$, and independently choose a value X_2^* from the distribution $\frac{1}{1-p^*}(\mu_2(x) - \mu_1(x)), x \in S_2$. Check that $(X_1^*, X_2^*) \in \mathcal{C}$ and $P(X_1^* \neq X_2^*) = 1 - p^* = ||\mu_1 - \mu_2||_{TV}$.]
6. Verify the solution to the equation (8.34).
7. Give an informal argument to prove Proposition 8.4(a). [*Hint*: Given $R_n = 0$, $R_{n+1} = j$ means that $Y_{N_{n+1}} = j + 1$ given $S_{N_n} = n$. The other part of (a) is obvious.]
8. Prove that the residual renewal process $\{R_n : n = 0, 1, 2 \ldots\}$, i.e., its transition probability, is aperiodic and irreducible if the probability mass function f has unit span.
9. Let $X_1, X_2, \ldots$ be an i.i.d. Bernoulli $0 - 1$ sequence with $p = P(X_1 = 1)$. Define $S_0 = 0$, $S_0 = \inf\{n > 0 : (X_n, X_{n+1}) = (1, 1)\}$, $S_1 = \inf\{n > S_0 : (X_n, X_{n+1}) = (1, 1)\}, \ldots, S_m = \inf\{n > S_{m-1} : (X_n, X_{n+1}) = (1, 1)\}, m \ge 1$.

[1] A more general version of this result for probabilities on Polish spaces is given in the monograph Lindvall (1992). This provides a proof of the maximality of the coupling used for the Poisson approximation in Chapter 5.

[2] BCPT p. 136.

(a) Show that $S_0, S_1, \ldots$ is a delayed renewal process.
(b) Show that $\mathbb{E}(S_m - S_{m-1}) = 4, m \geq 1$, and $\mathbb{E}S_0 = 5$.[*Hint*: For the latter write $\mathbb{E}S_0 = \mathbb{E}S_0\mathbf{1}_{[X_1=0]} + \mathbb{E}S_0\mathbf{1}_{[X_1=1,X_2=0]} + \mathbb{E}S_0\mathbf{1}_{[X_1=1,X_2=1]}$ and condition each term accordingly.]
(c) Generalize this example to an arbitrary fixed pattern $(\theta_1, \ldots, \theta_k) \in \{0, 1\}^k$.

10. (*Generalized Fibonacci Sequences*) Fix a positive integer $k \geq 2$. The generalized Fibonacci sequence is defined[3] by $x_n = 0, n \leq 0$, $x_n = 1, 1 \leq n < k$, $x_n = \sum_{i=1}^{k} x_{n-i}, n \geq k$. The standard Fibonacci sequence is obtained in the case $k = 2$.

(a) Show that there is a unique number q in $(0, 1)$ such that $\sum_{j=1}^{k} q^j = 1$.
(b) Define a probability distribution $f(j) = q^j, j = 1, \ldots, k$, $f(j) = 0, j \geq k + 1$, on the positive integers for the i.i.d. times between renewals $Y_1, Y_2 \ldots$. Let $S_n = \sum_{j=1}^{n} Y_j$ be the corresponding ordinary renewal process, and $N(n) = \sup\{j \geq 0 : S_j \leq n\}$.

(i) Show that $P(S_{N(n)} = n) = q^n x_{n+1}$. [*Hint*: Note that $P(S_{N(0)} = 0) = 1$, and for $n \geq 1$, $P(S_{N(n)} = n) = \sum_{(j_1,\ldots,j_m)\in\{1,\ldots,k\}^m, \sum_{i=1}^{m} j_i = n-1, m\geq 1} P(Y_1 = j_1, \ldots, Y_m = j_m)$. Check that for each $n \geq 1$, $c_n = |\cup_{m=1}^{\infty} \{(j_1, \ldots, j_m) \in \{1, \ldots, k\}^m : \sum_{i=1}^{k} j_i = n-1\}|$ obeys the generalized Fibonacci recursion, i.e., $x_n = c_n \forall n$.]
(ii) Show that $u(n) = P(N(n) = n)$ satisfies the renewal equation $u(n) = \sum_{j=1}^{n} u(n - j)f(j)$, and $q^{n-1}x_n \to \frac{(1-q)^2}{kq^{k+2}-(k+1)q^{k+1}+q}$ as $n \to \infty$.

(c) Compute the rate q for the standard Fibonacci sequence, i.e., $k = 2$.

11. Consider that $S_{N_n-1} < n \leq S_{N_n}$. For an ordinary renewal process, the *age of a renewal* , is $A_n = n - S_{N_n-1}, n > S_1 \equiv Y_1$, and the *lifespan* is $L_n = R_n + A_n = S_{N_n} - S_{N_n-1}, n > S_1 \equiv Y_1$,, where R_n is the *residual (remaining) life* defined by (8.21), i.e., $R_n := S_{N_n} - n, n = 1, 2, \ldots$. Fix positive integers[4] $r, a, \ell \geq 1$. Assume that the span of the renewal distribution is $h = 1$ with mean $0 < \mu < \infty$.

(a) Show that $P(R_n = r), n \geq 1$ satisfies a renewal equation with $g(n) = f(n + r), n \geq 1$, and $\lim_{n\to\infty} P(R_n = r) = \frac{\sum_{k=r+1}^{\infty} f(k)}{\mu}$.
(b) Show that $P(A_n = a), n \geq 1$ satisfies a renewal equation, $g(n) = f(n)\mathbf{1}_{[a,\infty]}(n)$ and $\lim_{n\to\infty} P(A_n = a) = \frac{1}{\mu}\sum_{n=a}^{\infty} f(n)$.
(c) Show that $P(L_n = \ell), n \geq 1$ satisfies a renewal equation with $g(n) = f(\ell)\mathbf{1}_{\{1,\ldots,\ell\}}(n), n = 0, 1, \ldots$, and $\lim_{n\to\infty} P(L_n = \ell) = \frac{\ell f(\ell)}{\mu}$.
(d) Show

[3]This was introduced by Miles (1960). The analysis via renewal theory was inspired by Christensen (2012).

[4]The formulae can be a bit different for integer renewal times than for continuously distributed renewals.

$$P(A_n = a, L_n = \ell)$$
$$= \begin{cases} f(\ell) & \text{if } a = n, \ell > n \\ P(S_m = n - a \text{ for some } m) f(\ell) & \text{if } a < n, \ell > n \\ 0 & \text{otherwise.} \end{cases}$$

[*Hint*: If $a = n, \ell > n$ then $[A_n = n]$ occurs if and only $Y_1 = k$ for some $k > n$, and, in particular also for $[L_n = \ell]$ to occur, $k = \ell$. If $a < n, \ell > n$, then $A_n = a$ if and only if there is a renewal at $n - a$, an event with probability $P(S_m = n - a$ for some $m)$ independently of the next renewal time.]

12. (*Waiting Time/Inspection Paradox*) The waiting time paradox, or inspection paradox, refers to the counterintuitive experience of longer average waits for arrivals (renewals) relative to arbitrarily fixed times; e.g., occurrences of defectives on an assembly line. To examine how it may happen in the case of integer renewal times with unit span, assume $\mathbb{E}Y_1^2 = \sigma^2 + \mu^2 < \infty$, and show that in steady state[5] (a) $\mathbb{E}R_n = \frac{1}{2}((\frac{\sigma^2}{\mu^2} + 1)\mu - 1)$, where $\frac{\sigma^2}{\mu^2}$ is the squared coefficient of variation[6] of Y_1, and (b) $\frac{1}{2}((\frac{\sigma^2}{\mu^2} + 1)\mu - 1) > \mu$ if and only if $\sigma^2 > \mu^2 + \mu$. [Intuitively, for a system in equilibrium if there is a lot of variability in production then to achieve the mean μ many intervals will be short to compensate for the long intervals, so it is more likely the inspection is made during a long window between arrivals than a short one].

[5] More generally one may use (i) of the previous exercise to compute $\lim_{n\to\infty} \mathbb{E}R_n$.

[6] The precise form of the mean residual time for integer renewal times differs a bit from that of arrivals having a density, see Chapter 25, Exercise 4.

Chapter 9
Bienaymé–Galton–Watson Simple Branching Process and Extinction

The Bienaymé–Galton–Watson simple branching process is defined by the successive numbers X_n of progeny at the n-th generation, $n = 0, 1, 2, \ldots,$ recursively and independently generated according to a given offspring distribution, starting from a non-negative integer number of initial X_0 progenitors. The state zero, referred to as extinction, is an absorbing state for the process. In this chapter a celebrated formula for the probability of extinction is given as a fixed point of the moment generating function of the offspring distribution. The mean μ of the offspring distribution is observed to play a characteristic role in the determination of the behavior of the generation sizes X_n as $n \to \infty$. The critical case in which $\mu = 1$ is analyzed under a finite second moment condition to determine the precise asymptotic nature of the survival probability, both unconditionally and conditionally on survival, in a theorem referred to as the Kolmogorov–Yaglom–Kesten–Ney–Spitzer theorem.

The (discrete parameter) *Bienaymé–Galton–Watson simple branching* process $\{X_n : n = 0, 1, 2, \ldots\}$ describes the successive sizes of generations resulting from an initial number of X_0 individuals as follows. Let $\{L_i^{(n)} : i, n = 1, 2, \ldots\}$ be a family of i.i.d. random variables with distribution $f(k) = P(L_i^{(n)} = k)$, $k = 0, 1, 2, \ldots$, referred to as the *offspring distribution*. If the initial generation is comprised of $X_0 \geq 1$ individuals, then the first generation has size $X_1 = \sum_{i=1}^{X_0} L_i^{(1)}$, while if $X_0 = 0$, then one defines $X_1 = 0$. This random replication rule is extended to successive generations by the stochastic recursion for the number of $n + 1$-st generation offspring as:

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_9

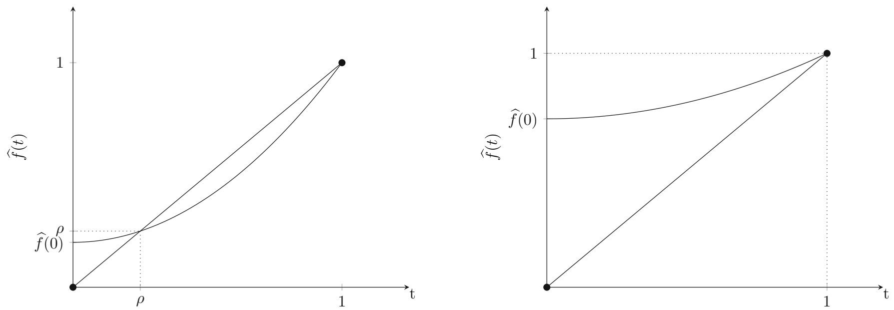


Fig. 9.1 Extinction Probability for Subcritical and Supercritical Branching

$$X_{n+1} = \begin{cases} \sum_{i=1}^{X_n} L_i^{(n+1)} & \text{if} \quad X_n \geq 1 \\ 0 & \text{if} \quad X_n = 0 \end{cases}, \qquad n = 0, 1, 2, \ldots. \tag{9.1}$$

The state space is $S = \{0, 1, 2, \ldots\}$ and the boundary state $i = 0$ is *absorbing* (permanent *extinction*). Write ρ_{i0} for the probability that extinction eventually occurs given $X_0 = i$. Also write $\rho = \rho_{10}$. Then $\rho_{i0} = \rho^i$ since each of the i sequences of generations arising from the i initial individuals has the same chance ρ of extinction, and the i sequences evolving independently must all be extinct in order that there may be eventual extinction, given $X_0 = i$.

If $f(0) = 0$, then $\rho = \rho_{10} = 0$ and extinction is impossible. If $f(0) = 1$, then $\rho_{10} \geq P(X_1 = 0 \mid X_0 = 1) = 1$ and extinction is certain (no matter what X_0 is). To avoid these and other trivialities we assume, unless otherwise specified,

$$0 < f(0) < 1. \tag{9.2}$$

As will be seen the extinction probability depends on the mean number of offspring as depicted in Figure 9.1. The case in which $\mu < 1$ is referred to as *subcritical*, $\mu = 1$ as *critical*, and $\mu > 1$ as *supercritical*.

Theorem 9.1. Assume $0 < f(0) < 1$. For the Bienaymé–Galton–Watson simple branching process with $\mu = \sum_{k=1}^{\infty} kf(k) < \infty$, $X_0 = i$, the probability $P(\lim_{n\to\infty} X_n = 0)$ of eventual extinction is ρ^i and $\rho = \rho_{10}$ is the smallest fixed point of $\hat{f}(t) = \sum_{k=0}^{\infty} t^k f(k)$, $0 \leq t \leq 1$. Moreover $\rho = 1$ iff $\mu \leq 1$.

Proof. First let us consider the case $f(0) + f(1) < 1$. Introduce the *probability generating function* of f:

$$\hat{f}(z) = \sum_{j=0}^{\infty} f(j)z^j = f(0) + \sum_{j=1}^{\infty} f(j)z^j \quad (|z| \leq 1). \tag{9.3}$$

Since a power series can be differentiated term by term within its radius of convergence, one has

$$\hat{f}'(z) \equiv \frac{d}{dz}\hat{f}(z) = \sum_{j=1}^{\infty} jf(j)z^{j-1} \quad (|z| < 1). \tag{9.4}$$

If the *mean* μ *of the number of particles* generated by a single particle is finite, i.e., if

$$\mu \equiv \sum_{j=1}^{\infty} jf(j) < \infty, \tag{9.5}$$

then (9.4) holds even for the left-hand derivative at $z = 1$, i.e.,

$$\mu = \hat{f}'(1). \tag{9.6}$$

Since $\hat{f}'(z) > 0$ for $0 < z < 1$, $\hat{f}$ is strictly increasing. Also, since $\hat{f}''(z)$ (which exists and is finite for $0 \le z < 1$) satisfies

$$\hat{f}''(z) \equiv \frac{d^2}{dz^2}\hat{f}(z) = \sum_{j=2}^{\infty} j(j-1)f(j)z^{j-2} > 0 \qquad \text{for } 0 < z < 1, \tag{9.7}$$

the function $\hat{f}$ is *strictly convex* on [0, 1]. In other words, the line segment joining any two points on the curve $y = \hat{f}(z)$ lies strictly above the curve (except at the two points joined). Because $\hat{f}(0) = f(0) > 0$ and $\hat{f}(1) = \sum_{j=0}^{\infty} f(j) = 1$, the possible graph of $\hat{f}$ is as depicted in Figure 9.1, accordingly.

The maximum of $\hat{f}'(z)$ is μ, which is attained at $z = 1$. Hence, in the case $\mu > 1$, the graph of $y = \hat{f}(z)$ must lie below that of $y = z$ near $z = 1$ and, because $\hat{f}(0) = f(0) > 0$, must cross the line $y = z$ at a point z_0, $0 < z_0 < 1$. Since the slope of the curve $y = \hat{f}(z)$ continuously increases as z increases in (0, 1), z_0 is the unique solution of the equation $z = \hat{f}(z)$ that is smaller than 1.

In case $\mu \le 1$, $y = \hat{f}(z)$ must lie strictly above the line $y = z$, except at $z = 1$. For if it meets the line $y = z$ at a point $z_0 < 1$, then it must go under the line in the immediate vicinity to the right of z_0, since its slope falls below that of the line (i.e., unity). In order to reach the height $\hat{f}(1) = 1$ (also reached by the line at the same value $z = 1$) its slope then must exceed 1 somewhere in $(z_0, 1]$; this is impossible since $\hat{f}'(z) \le \hat{f}'(1) = \mu \le 1$ for all z in [0, 1]. Thus, the only solution of the equation $z = \hat{f}(z)$ is $z = 1$.

Now observe

$$\rho = \rho_{10} = \sum_{j=0}^{\infty} P(X_1 = j \mid X_0 = 1)\rho_{j0} = \sum_{j=0}^{\infty} f(j)\rho^j = \hat{f}(\rho), \tag{9.8}$$

thus *if* $\mu \le 1$, *then* $\rho = 1$ *and extinction is certain*. On the other hand, suppose $\mu > 1$. Then ρ is either z_0 or 1. We shall now show that $\rho = z_0$ (< 1). For this, consider the quantities

$$q_n := P(X_n = 0 \mid X_0 = 1) \quad (n = 1, 2, \ldots). \tag{9.9}$$

That is, q_n is the probability that the sequence of generations originating from a single particle is extinct at time n. As n increases, $q_n \uparrow \rho$; for clearly, $[X_n = 0] \subset [X_m = 0]$ for all $m \ge n$, so that $q_n \le q_m$. Also

$$[\lim_{n\to\infty} X_n = 0] = \bigcup_{n=0}^{\infty} [X_n = 0] = [\text{extinction occurs}].$$

Now, by independence of the generations originating from different particles,

$$P(X_n = 0 \mid X_0 = j) = q_n^j \quad (j = 0, 1, 2, \ldots),$$

$$\begin{aligned}
q_{n+1} &= P(X_{n+1} = 0 \mid X_0 = 1) = P(X_1 = 0 \mid X_0 = 1) \\
&\quad + \sum_{j=1}^{\infty} P(X_1 = j, X_{n+1} = 0 \mid X_0 = 1) \\
&= f(0) + \sum_{j=1}^{\infty} P(X_1 = j \mid X_0 = 1) P(X_{n+1} = 0 \mid X_0 = 1, X_1 = j) \\
&= f(0) + \sum_{j=1}^{\infty} f(j) q_n^j = \hat{f}(q_n) \quad (n = 1, 2, \ldots).
\end{aligned} \tag{9.10}$$

Since $q_1 = f(0) = \hat{f}(0) < \hat{f}(z_0) = z_0$ (recall that $\hat{f}(z)$ is strictly increasing in z for $0 < z < 1$), one has using (9.10) with $n = 1$, $q_2 = \hat{f}(q_1) < \hat{f}(z_0) = z_0$, and so on. Hence, $q_n < z_0$ for all n. Therefore, $\rho = \lim_{n\to\infty} q_n \le z_0$. This proves $\rho = z_0$. If $f(0) + f(1) = 1$ and $0 < f(0) < 1$, then $\hat{f}''(z) = 0$ for all z, and the graph of $\hat{f}(z)$ is the line segment joining $(0, f(0))$ and $(1, 1)$. Hence, $\rho = 1$ in this case. ■

Let us now compute the average size of the nth generation. One has

$$\begin{aligned}
\mathbb{E}(X_{n+1} \mid X_0 = 1) = E\left(\sum_{i=1}^{X_n} L_i^{(n+1)} \,\middle|\, X_0 = 1\right) \\
= \mathbb{E}\left[\mathbb{E}\left(\sum_{i=1}^{X_n} L_i^{(n+1)} \,\middle|\, X_n\right) \,\middle|\, X_0 = 1\right] \\
= \mathbb{E}[\mu X_n \mid X_0 = 1] = \mu \mathbb{E}(X_n \mid X_0 = 1).
\end{aligned} \tag{9.11}$$

Continuing in this manner, one obtains

$$\begin{aligned}\mathbb{E}(X_{n+1} \mid X_0 = 1) &= \mu\mathbb{E}(X_n \mid X_0 = 1) = \mu^2\mathbb{E}(X_{n-1} \mid X_0 = 1)\\ &= \cdots = \mu^n\mathbb{E}(X_1 \mid X_0 = 1) = \mu^{n+1}.\end{aligned} \tag{9.12}$$

It follows that

$$\mathbb{E}(X_n \mid X_0 = j) = j\mu^n. \tag{9.13}$$

Thus, in the subcritical case ($\mu < 1$), the expected size of the population at time n decreases to zero exponentially fast as $n \to \infty$. In particular, therefore,

$$P(X_n > 0 \mid X_0 = 1) = \sum_{j=1}^{\infty} P(X_n = j \mid X_0 = 1) \le \mathbb{E}(X_n \mid X_0 = 1) = \mu^n \tag{9.14}$$

goes to zero exponentially fast if $\mu < 1$. In the critical case ($\mu = 1$), the expected size at time n does not depend on n (i.e., it is the same as the initial size).

Finally in the supercritical case ($\mu > 1$), the expected size of the population increases to infinity exponentially fast. These processes are examined further in Chapter 14. However, we will conclude this chapter with a classic result on the decay of processes in the critical case. In particular we prove the following

Theorem 9.2 *(Kolmogorov–Yaglom–Kesten–Ney–Spitzer).* If $\mu = 1$ and $\sigma^2 := \sum_{k=1}^{\infty} k(k-1)f(k) < \infty$, then

a. $\lim_{n\to\infty} nP(X_n > 0) = \frac{2}{\sigma^2}$

b. $\lim_{n\to\infty} P(\frac{X_n}{n} > x | X_n > 0) = e^{-\frac{2x}{\sigma^2}} \quad x \ge 0.$

The proof is given below following a preparatory example and comparison lemma.

Notice that if $\mu = 1$, then the *variance* $\sigma^2 = \sum_{k=1}^{\infty}(k-\mu)^2 f(k)$ of the offspring distribution coincides with the second factorial moment $\gamma^2 = \sum_{k=1}^{\infty} k(k-1)f(k)$.

Remark 9.1. In this generality Theorem 9.2 is often referred to as the Kesten-Ney-Spitzer theorem. However, due to earlier versions of these results which were originally proved[1] under stronger (finite third moment) assumptions, the first result is often called *Kolmogorov's probability decay rate* and the latter[2] is *Yaglom's exponential law*.

Let $\hat{g}(s) = \sum_{k=0}^{\infty} s^k g(k) \equiv \mathbb{E}s^{X_1}$ be the *generating function* of X_1, $0 \le s \le 1$. Denote the generating function of X_n by $\hat{g}^{(n)}$. Note that, due to the independence

[1] Kolmogorov (1938).

[2] Yaglom (1947).

of X_{n-1} and the i.i.d. sequence $\{L_i^{(n)} : i = 1, 2, \ldots\}$, for $n > 1$, the generating function of $X_n = \sum_{i=1}^{X_{n-1}} L_i^{(n)}$ is given by

$$
\begin{aligned}
\mathbb{E}s^{X_n} &= \mathbb{E}\mathbb{E}(s^{\sum_{i=1}^{X_{n-1}} L_i^{(n)}}|X_{n-1}) \\
&= \mathbb{E}\hat{g}(s)^{X_{n-1}} = \hat{g}(\hat{g}^{(n-1)}(s)).
\end{aligned}
\tag{9.15}
$$

That is, one has the useful property that $\hat{g}^{(n)}$ is the n-fold composition of $\hat{g}$.

The proof of Theorem 9.2 will be based on a comparison of the general case to the following explicit example.

Example 1 *(Geometric Offspring Distribution: Explicit Calculations)***.** In the important special case of (critical) geometric offspring distribution $g(0) = p = 1 - q, g(k) = q^2 p^{k-1}, k = 1, 2 \ldots,$ this result may be proven by an explicit computation. Note that $\sum_{k=1}^{\infty} kg(k) = 1$, and $\sigma^2 = \sum_{k=2}^{\infty} k(k-1)g(k) = 2p/q$. This particular form of the critical "geometric distribution" is selected to afford some flexibility with the size of second factorial moment σ^2 which will be exploited in the proof of Theorem 9.2 below. In the case of this example one has $\hat{g}(s) = \sum_{k=0}^{\infty} g(k)s^k = \frac{p-(p-q)s}{1-ps}$ and the generating function of X_n is defined by the composite function $\hat{g}^{o(n)}(s)$ inductively given by $\hat{g}^{o(1)}(s) := \hat{g}(s)$, $\hat{g}^{o(n+1)}(s) := \hat{g}(\hat{g}^{o(n)}(s))$. One then has

$$
\hat{g}^{o(n)}(s) = \frac{np - (np - q)s}{q + np - nps}, \tag{9.16}
$$

as can be easily verified by mathematical induction (Exercise 5). Therefore, taking $s = 0$,

$$
nP(X_n > 0) = n(1 - \hat{g}^{o(n)}(0)) = \frac{nq}{q + np} \to \frac{2}{\sigma^2} \quad \text{as } n \to \infty. \tag{9.17}
$$

Similarly, taking $s = e^{it/n}$ the conditional distribution of $\frac{1}{n}X_n$ given $X_n > 0$ has characteristic function

$$
\mathbb{E}(e^{itX_n/n}|X_n > 0) = \frac{\hat{g}^{o(n)}(e^{it/n}) - \hat{g}^{o(n)}(0)}{1 - \hat{g}^{o(n)}(0)} \quad (= \frac{qe^{it/n}}{q + np(1 - e^{it/n})}). \tag{9.18}
$$

In particular, using L'Hospital's rule,

$$
\lim_{n\to\infty} \mathbb{E}(e^{itX_n/n}|X_n > 0) = \frac{1}{1 - it\frac{\sigma^2}{2}} \quad t \in \mathbb{R}. \tag{9.19}
$$

Hence, by the continuity theorem, the conditional distribution converges weakly to the exponential distribution.

To proceed with the general case it is sufficient to establish that the equality in (9.18) for each n and all $s \in [0, 1)$ can much more generally be replaced by a limit as $n \to \infty$ which is uniform in s. This will be accomplished with the aid of the following lemma.[3]

Lemma 1 *(Spitzer's Comparison Lemma).* Suppose that f and g are mean one probability mass functions on the non-negative integers having probability generating functions $\hat{f}(s) = \sum_{k=0}^{\infty} s^j f(j)$ and $\hat{g}(s) = \sum_{j=0}^{\infty} s^j g(j)$, respectively. Assume that

$$\hat{f}'(1^-) = \hat{g}'(1^-) = 1, \quad \hat{f}''(1^-) < \hat{g}''(1^-) < \infty.$$

Then there exist integers k and m such that

$$\hat{f}^{o(n+k)}(s) \le \hat{g}^{o(n+m)}(s) \quad \text{for all} \quad s \in [0, 1], \quad n = 0, 1, 2, \ldots.$$

Proof. For simplicity, we prove the lemma under the stronger assumption that the radius of convergence of $\hat{f}$ is larger than one here, and provide steps for the proof under the state hypothesis in Exercise 2. One may then expand $\hat{f}(s)$ in a Taylor series up to second (or third) derivatives around $s = 1$ to conclude from the assumptions on derivatives at 1^-, that for s close to one, say on $[s_0, 1]$, $\hat{f}(s) < \hat{g}(s)$, since $\hat{f}(1) = \hat{g}(1)$, $\hat{f}'(1) = \hat{g}'(1)$, and $\hat{f}''(1) < \hat{g}''(1)$. Using monotonicity, one has $\hat{f}^{o(n)}(s) \le \hat{g}^{o(n)}(s)$ on $[s_0, 1]$ for all $n \ge 1$, (see Exercise 2(iv)). Since for $s \in [0, 1]$ one has both $\hat{f}^{o(n)}(s) \to 1$ and $\hat{g}^{o(n)}(s) \to 1$ as $n \to \infty$, there is a non-negative integer k such that $s_0 < \hat{f}^{o(k)}(0)$, and then one may select $m > k$ such that $\hat{f}^{o(k)}(s_0) \le \hat{g}^{o(m)}(0)$. Thus, for $0 \le s \le s_0$ one also has

$$s_0 \le \hat{f}^{o(k)}(0) \le \hat{f}^{o(k)}(s) \le \hat{f}^{o(k)}(s_0) \le \hat{g}^{o(m)}(0) \le \hat{g}^{o(m)}(s).$$

So the assertion holds on $0 \le s \le s_0$ as well. This proves the lemma under the stronger assumption made at the outset of the proof. For the proof under the stated hypothesis we refer to Exercise 2. ■

Proposition 9.3. Assume $f(1) < 1$. Let $\hat{f}(s) = \sum_{k=1}^{\infty} f(k)s^k$. If $\sum_{k=1}^{\infty} kf(k) = 1$ and $\sigma^2 = \sum_{k=2}^{\infty} k(k-1)f(k) < \infty$, then

$$\frac{1}{n}\{\frac{1}{1 - \hat{f}^{o(n)}(s)} - \frac{1}{1-s}\} \to \frac{\sigma^2}{2} \quad \text{uniformly in } s \text{ as } n \to \infty.$$

Proof. Take $g(0) = p = 1 - q, \quad g(j) = q^2 p^{j-1}, j = 1, \ldots.$ Then $\hat{g}(s) = \frac{p-(2p-1)s}{1-ps}$ satisfies the hypothesis of Spitzer's comparison lemma for $p \in (0, 1)$ selected such that $\hat{g}''(1^-) = \frac{2p}{1-p} = (1+\epsilon)\sigma^2 > \sigma^2$. Thus one has $k < m$ such that

[3] This result, often attributed to Frank Spitzer, appears in Kesten et al. (1966).

$$\hat{f}^{o(n+k)}(s) \leq \hat{g}^{o(n+m)}(s), \quad s \in [0,1], \quad n \geq 0.$$

Therefore

$$\begin{aligned}(1-\hat{f}^{o(n+k)}(s))^{-1} - (1-s)^{-1} &\leq (1-\hat{g}^{o(n+m)}(s))^{-1} - (1-s)^{-1} \\ &= (n+m)p/q = (n+m)(1+\epsilon)\frac{\sigma^2}{2}.\end{aligned}$$

On the other hand choosing p such that $\frac{2p}{1-p} = (1-\epsilon)\sigma^2 < \sigma^2$, one has for some $k' < m'$ that

$$(k'+n)(1-\epsilon)\frac{\sigma^2}{2} \leq (1-\hat{f}^{o(m'+n)}(s))^{-1} - (1-s)^{-1}, \quad \forall n \geq 1.$$

Since $\epsilon > 0$ is arbitrary the asserted limit follows. ■

Proof of Kolmogorov–Yaglom–Kesten–Ney–Spitzer Theorem. Part (a) now follows since $P(X_n > 0) = 1 - \hat{f}^{o(n)}(0)$. The part (b) also follows essentially just as in the case of the geometrically distributed offspring since, using Laplace transforms in place of characteristic functions,

$$\begin{aligned}\mathbb{E}\big(e^{-tX_n/n}|X_n > 0\big) &= \frac{\hat{f}^{(n)}(e^{-t/n}) - \hat{f}^{(n)}(0)}{1-\hat{f}^{(n)}(0)} \\ &= 1 - \frac{\big[n\big(1-\hat{f}^{(n)}(0)\big)\big]^{-1}}{\big[n\big(1-\hat{f}^{(n)}(e^{-t/n})\big)\big]^{-1}}, \ t \geq 0.\end{aligned}$$

Letting $n \to \infty$ and using the uniformity of convergence in the preceding proposition, one calculates from this using L'Hospital's rule that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\big(e^{-tX_n/n}|X_n > 0\big) &= 1 - \lim_{n\to\infty} \frac{\sigma^2/2}{\sigma^2/2 + \frac{1}{n(1-e^{-t/n})}} \\ &= 1 - \frac{\sigma^2/2}{\sigma^2/2 + 1/t} = \frac{1}{1+\frac{t\sigma^2}{2}}.\end{aligned}$$

But this is the Laplace transform of the asserted exponential distribution, and the result follows by the continuity theorem[4] for Laplace transforms.

[4] Feller (1971), p. 431.

Exercises

1. Let $\{X_n : n = 0, 1, \dots\}$ be the simple branching process on $S = \{0, 1, 2, \dots\}$ with offspring distribution f such that $\mu = \sum_{k=0}^{\infty} kf(k) \le 1$. Show that all nonzero states are transient and 0 is an absorbing state and, therefore, recurrent.
2. (*Spitzer's comparison lemma*) Complete the steps below to prove Spitzer's comparison lemma under the stated hypothesis.
   (i) Show $\lim_{s\uparrow 1} \frac{\hat{f}(s)-s}{(1-s)^2} = \frac{1}{2}\hat{f}''(1^-)$.
   (ii) Show that $\frac{1}{2}\hat{f}''(1^-) - \frac{\hat{f}(s)-s}{(1-s)^2}$ can be expressed as the obviously non-negative, albeit quite clever, series

$$\frac{1}{2}\hat{f}''(1^-) - \frac{\hat{f}(s)-s}{(1-s)^2} = \sum_{k=3}^{\infty} f(k) \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} (1-s^i) \ge 0 (0 \le s < 1).$$

   Note that the left side considers the difference between $\hat{f}(s)$ and the sum of the first three terms in a formal Taylor expansion about $s = 1$ divided by $(1-s)^2$. However, such an expansion is not generally permitted under the assumed moment conditions. So the right side $r(s) := \sum_{k=3}^{\infty} f(k) \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} (1-s^i) \ge 0$ is a quite clever series derived from the (scaled) difference between $\hat{f}(s)$ and $\hat{f}(1) + \hat{f}'(1^-)(s-1) + \frac{1}{2}\hat{f}''(1^-)(s-1)^2$. [*Hint*: Use $(1-s)\sum_{j=1}^{k-1} s^j = s - s^k$, noting $\mathbb{E}L = 1, \mathbb{E}(L-1)(L-2) = \mathbb{E}L(L-1) = \frac{1}{2}\hat{f}''(1^-)$, and rather extensive algebraic simplifications.]
   (iii) Show that $\hat{f}(s) = s + \frac{\hat{f}''(1^-)}{2}(1-s)^2 - (1-s)^2 r(s)$, $r(s) \ge 0$, $0 \le s < 1$. Similarly for $\hat{g}(s)$. [*Hint*: $r(s) \downarrow 0$ as $s \uparrow 1$.]
   (iv) Show that there is an $s_0 \in [0, 1)$ such that $s_0 \le s \le \hat{f}(s) \le \hat{g}(s)$ for $s_0 \le s \le 1$, and hence, using monotonicity and induction, $\hat{f}^{o(n)}(s) \le \hat{g}^{o(n)}(s)$ on $[s_0, 1]$ for all $n \ge 1$.
   (v) Show that the proof of Spitzer's comparison lemma can be completed from here by the same arguments in the proof given there.
3. Under the conditions of Theorem 9.2 show that $\mathbb{E}(X_n|X_n > 0) \sim \frac{\sigma^2}{2}n$ as $n \to \infty$. [*Hint*: $\mathbb{E}(X_n|X_n > 0)P(X_n > 0) = 1$.]
4. Suppose that $\mu = \sum_{k=0}^{\infty} kf(k) < \infty$ and $\sigma^2 = \sum_{k=0}^{\infty}(k-\mu)^2 f(k) < \infty$. Show that given $X_0 = 1$,

$$\operatorname{Var} X_n = \begin{cases} \sigma^2\mu^{n-1}(\mu^n - 1)/(\mu - 1) & \text{if } \mu \ne 1 \\ n\sigma^2 & \text{if } \mu = 1 \end{cases}$$

5. (a) Provide the induction argument to prove (9.16).

(b) Suppose that the offspring distribution is given by the (critical) geometric distribution $g(0) = p = 1 - q,\ g(k) = q^2 p^{k-1}$, $k = 1, 2 \ldots$ where $0 < p < 1$. Check that

(i) $\hat{g}(s) := \sum_{k=0}^{\infty} s^k g(k) = \dfrac{p - (p - q)s}{q + p(1 - s)}$

(ii) $\hat{g}^{o(n)}(s) = \dfrac{np + (q - np)s}{q + np(1 - s)}$ where the composite function $\hat{g}^{o(n)}(s)$ is inductively defined $\hat{g}^{o(1)}(s) := \hat{g}(s)$, $\hat{g}^{o(n+1)}(s) := \hat{g}(\hat{g}^{o(n)}(s))$.

Chapter 10
Martingales: Definitions and Examples

Martingale theory is a cornerstone to stochastic analysis and is included in this book from that perspective. This chapter introduces the theory with examples and their basic properties. For some readers this chapter may serve as a review.

Statistical independence is long recognized for its dominant role in classic probability, especially the limit theorems, and is the single notion that has distinguished probability theory as a distinct mathematical discipline set apart from real analysis and measure theory. Martingale theory, pioneered by Doob (1953), has taken this paradigm to the next level, as a more general form of statistical dependence under which much of classic probability is subsumed. Much of the classical theory of sums of i.i.d. random variables such as laws of large numbers and central limit theorems may be viewed more generally as consequences of martingale structure.

If, for example, $Z_1, Z_2, \dots$ is a sequence of independent integrable mean zero random variables on $(\Omega, \mathcal{F}, P)$, then the sequence $X_n = Z_1 + \cdots + Z_n$, $n \geq 1$, enjoys the seemingly simple but quite more general property,

$$\mathbb{E}(X_{n+1}|\sigma(X_1, \dots, X_n)) = X_n, \quad n \geq 1, \tag{10.1}$$

having far-reaching consequences.

As usual, $(\Omega, \mathcal{F}, P)$ will continue to denote the underlying probability space on which all random variables introduced in this chapter are defined, unless stated otherwise.

Let T be a *linearly ordered* parameter set, e.g., $T = \mathbb{Z}_+ \equiv \{0, 1, 2, \dots\}$, $\{1, 2, \dots, N\}$, $[0, \infty)$, or $T = [0, 1]$. The parameter sets $\mathbb{Z}_+$ and $\mathbb{N} = \{1, 2, \dots\}$

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_10

are interchangeable by a unit shift. A family of σ-fields $\{\mathcal{F}_t : t \in T\}$ is said to be a *filtration* if (i) $\mathcal{F}_t \subset \mathcal{F} \forall t$, and (ii) $\mathcal{F}_s \subset \mathcal{F}_t \ \forall s < t$. A family of random variables $\{X_t : t \in T\}$, with values in a measurable space $(S, \mathcal{S})$, is said to be $\{\mathcal{F}_t : t \in T\}$-*adapted* if X_t is $\mathcal{F}_t$-measurable $\forall t \in T$.

Definition 10.1. A family $\{X_t : t \in T\}$ of integrable real-valued random variables is a $\{\mathcal{F}_t : t \in T\}$-*submartingale* with respect to a filtration $\{\mathcal{F}_t : t \in T\}$ if it is $\{\mathcal{F}_t : t \in T\}$-adapted and

$$\mathbb{E}(X_t \mid \mathcal{F}_s) \geq X_s \text{ a.s.} \quad \forall s < t. \tag{10.2}$$

If (a.s.) equality holds in (10.2) $\forall s < t$, then $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$-*martingale*. If the inequality in (10.2) is reversed, then $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$-*supermartingle*.

By taking successive conditional expectations, in the case $T = \mathbb{Z}_+$ or $\mathbb{N}$ the requirement (10.2) is equivalent to

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \geq X_n \text{ a.s.} \quad \forall n, \tag{10.3}$$

with equality, if $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale. In particular, taking expectations one sees that submartingales have monotonically non-decreasing expected values while martingales have constant expected values.

Convention. *Unless otherwise stated we will always assume that T has a smallest element.*

A $\{\mathcal{F}_t : t \in T\}$-martingale (submartingale) is always a $\{\mathcal{F}_t^{\mathbf{X}} : t \in T\}$-martingale (respectively, submartingale) where $\mathcal{F}_t^{\mathbf{X}} := \sigma\{X_s : s \leq t\}$, but often the process $\{X_t : t \in T\}$ arises in a broader context with its martingale property with respect to a larger filtration. Some of the examples in this chapter illustrate this, as does Proposition 10.1. Larger filtrations than $\{\mathcal{F}_t^{\mathbf{X}} : t \in T\}$ may also arise, e.g., by letting $\mathcal{F}_t = \mathcal{F}_t^{\mathbf{X}} \vee \mathcal{A} \equiv \sigma(\mathcal{G}_t \cup \mathcal{A})$, where $\mathcal{A}$ ($\subset \mathcal{F}$) is a σ-field independent of $\sigma(\cup_t \mathcal{F}_t^{\mathbf{X}})$.

In our statements the prefix $\{\mathcal{F}_t : t \in T\}$-attached to the term martingale (or submartingale) is sometimes suppressed when the particular filtration is either clear from the context or its specification is not essential to the discussion.

If $\{Z_n : n \in \mathbb{Z}_+\}$ is a sequence of integrable $\{\mathcal{F}_n : n \in \mathbb{Z}_+\}$-adapted random variables such that $\mathbb{E}|Z_0| < \infty$ and $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = 0 \ \forall n \geq 0$, then $\{Z_n : n \in \mathbb{Z}_+\}$ is said to be a $\{\mathcal{F}_n : n \in \mathbb{Z}_+\}$-*martingale difference sequence*. In this case $X_n := Z_0 + Z_1 + \cdots + Z_n$ $(n \geq 0)$ is easily seen to be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale. Conversely, given a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale $\{X_n\}_{n=1}^{\infty}$, one obtains the $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale difference sequence $Z_0 = X_0$, $Z_n := X_n - X_{n-1}$ $(n \geq 1)$. One may similarly define a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale (or supermartingale-) *difference sequence* $\{\mathcal{F}_n\}_{n=1}^{\infty}$ by requiring that $\{Z_n\}_{n=1}^{\infty}$ be $\{\mathcal{F}_n\}_{n=1}^{\infty}$-adapted, $\mathbb{E}|Z_0| < \infty$, $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) \geq 0$ (respectively, ≤ 0) $\forall n \geq 0$.

In gambling language, a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale $\{X_n\}_{n=1}^{\infty}$ may be thought to represent a gambler's fluctuating fortunes, as time progresses, in a *fair game*. The gambler designs the n-th play based on past information (embodied in $\mathcal{F}_{n-1}$) and receives a payoff $Z_n = X_n - X_{n-1}$. But no matter what strategy is followed, $\mathbb{E}(Z_n \mid \mathcal{F}_{n-1}) = 0$, Similarly, a submartingale represents fortunes in a game *favorable* to the gambler, while a supermartingale corresponds to an *unfair* game. In pedestrian terms the martingale property conveys weather forecasts for which given the historical weather record, the expected weather tomorrow is the weather you see today! The following examples are easily verified using familiar properties of conditional expectation and, therefore, left as exercises.

Remark 10.1. Let $\{X_n : n \geq 0\}$ be a square-integrable martingale, and $\{Z_n = X_n - X_{n-1} : n \geq 1\}$ the corresponding sequence of martingale differences. The martingale property implies (and is equivalent to) the *lack of correlation*: $\mathbb{E}Z_nY_{n-1} = 0$, where Y_{n-1} is any $\mathcal{F}_{n-1}$-measurable square-integrable random variable, whereas *independence* of Z_n and $\mathcal{F}_{n-1}$ means $\mathbb{E}f(Z_n)Y_{n-1} = \mathbb{E}f(Z_n) \cdot \mathbb{E}Y_{n-1}$ for any f such that $\mathbb{E}f^2(Z_n) < \infty$, i.e. $f(Z_n)$ and Y_{n-1} are uncorrelated.

Example 1 (*Random Walks, Processes with Independent Increments*)**.** Let $\{Z_n : n \in \mathbb{Z}_+\}$ be a sequence of independent random variables satisfying $\mathbb{E}|Z_0| < \infty$ and $\mathbb{E}Z_n = 0\ \forall n \geq 1$. Then the random walk $S_n = Z_0 + Z_1 + \cdots + Z_n$ $(n \geq 0)$ is a $\{\mathcal{G}_n\}_{n=1}^{\infty}$-martingale where $\mathcal{G}_n = \sigma\{S_0, S_1, \ldots, S_n\}$, or equivalently $\mathcal{G}_n = \sigma\{Z_0, Z_1, \ldots, Z_n\}$, $(n \geq 0)$. If $\mathbb{E}Z_n \geq 0\ \forall n \geq 1$, then $\{S_n\}_{n=0}^{\infty}$ is a $\{\mathcal{G}_n\}_{n=0}^{\infty}$-submartingale. In continuous time, one may similarly consider a process $\{X_t : t \in [0, \infty)\}$ with independent increments. If $\mathbb{E}|X_0| < \infty$ and $\mathbb{E}(X_t - X_s) = 0\ \forall s < t$, then $\{X_t : t \in T\}$ is a $\{\mathcal{G}_t : t \in T\}$-martingale with $\mathcal{G}_t := \sigma\{X_s : s \leq t\}$. If $\mathbb{E}(X_t - X_s) \geq 0\ \forall s < t$, then $\{X_t : t \in T\}$ is a $\{\mathcal{G}_t : t \in T\}$-submartingale. For such a process $\{X_t : t \in T\}$ the mean-adjusted process $Y_t := X_t - \mathbb{E}X_t$ $(t \geq 0)$ is a $\{\mathcal{G}_t : t \in T\}$-martingale. Compound Poisson processes and Brownian motion are important examples.

Example 2 (*Conditional Expectations*)**.** Let X be an integrable random variable and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ an arbitrary filtration. Then the sequence

$$X_n := \mathbb{E}(X \mid \mathcal{F}_n) \qquad (n \geq 1) \tag{10.4}$$

is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale.

Example 3 (*Independent Products*)**.** Let $\{Y_n : n \geq 1\}$ be a sequence of independent random variables such that $\mathbb{E}|Y_1| < \infty$ and $\mathbb{E}Y_n = 1$ for $n \geq 2$. Write $X_n = Y_1Y_2 \ldots Y_n$, $\mathcal{F}_n = \sigma\{Y_1, \ldots, Y_n\}$. Then $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale.

Example 4 (*Critical Branching Processes*)**.** Let $X_0, X_1, X_2, \ldots$ denote the successive generations of a *Bienaymé–Galton–Watson Simple Branching* process with mean-one offspring distribution, i.e.,

$$X_{n+1} = \begin{cases} \sum_{i=1}^{X_n} L_i^{(n+1)} & \text{if } X_n \geq 1 \\ 0 \quad \text{if } X_n = 0, \end{cases} \tag{10.5}$$

for $n = 0, 1, \ldots$, where $\{L_i^{(n)} : i, n \geq 1\}$ is a collection of i.i.d. mean-one non-negative integer-valued random variables, independent of X_0, then $\{X_n\}_{n=1}^{\infty}$ is a martingale. Similarly in the supercritical and subcritical cases defined by $\mathbb{E}L^{(n)} \geq 1$ and $\mathbb{E}L^{(n)} \leq 1$, respectively, these are examples of supermartingale and submartingale structure.

Example 5 (*Likelihood Ratios from Statistics*). Let $U_1, U_2, \ldots,$ be i.i.d. S-valued observations from a distribution $f(u;\theta)\nu(du)$, where ν is a measure on a measurable space $(S, \mathcal{S})$. Here θ is a parameter. Under a certain hypothesis H^0, θ has value θ_0. Under an alternative hypothesis H^1, θ has value θ_1. The *likelihood ratios*

$$X_n := \prod_{j=1}^{n} f(U_j;\theta_1)/f(U_j;\theta_0) \quad (n \geq 1)$$

are a.s. finite under H^0 and form a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale with $\mathcal{F}_n := \sigma\{U_1, U_2, \ldots, U_n\}$. This is a special case of Example 3 with $Y_j = f(U_j;\theta_1)/f(U_j;\theta_0)$ $(j \geq 1)$, since, under H^0,

$$\mathbb{E}Y_j = \int \frac{f(u;\theta_1)}{f(u;\theta_0)} f(u;\theta_0)\nu(du) = \int f(u;\theta_1)\nu(du) = 1.$$

Example 6 (*Exponential Martingales*). Let $\{Z_n\}_{n=1}^{\infty}$ be an $\{\mathcal{F}_n\}_{n=1}^{\infty}$-adapted sequence. Let $S_n = Z_1 + \cdots + Z_n$. Assume that for some $\xi \neq 0$, $\mathbb{E}\exp\{\xi Z_n\} < \infty$ for all n. Define

$$\varphi_n(\xi) := \mathbb{E}(\exp\{\xi Z_n\}|\mathcal{F}_{n-1}), \quad X_n := \exp\{\xi S_n\} \Big/ \prod_{k=1}^{n} \varphi_k(\xi).$$

Assume X_n, are integrable for all n. Then, noting $\prod_{k=1}^{n+1} \varphi_k(\xi)$ is $\mathcal{F}_n$-measurable, $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale. The integrability conditions are satisfied if, in addition to the finiteness of $\mathbb{E}\exp\{\xi Z_n\}$ for all n, either (i) Z_n is bounded for every n, or (ii) $Z_n, n \geq 1$, are independent.

Instead of real ξ one may take imaginary $i\xi$ and let $\psi_n(\xi) = \mathbb{E}(\exp\{i\xi Z_n\} \mid \mathcal{F}_{n-1})$. Then $X_n := \exp\{i\xi S_n\}/\prod_{k=1}^{n} \psi_k(\xi)$ is a *complex martingale*, i.e., the real and imaginary parts of X_n are both martingales, provided $\mathbb{E}|X_n| < \infty \; \forall\, n$.

The following proposition provides further important examples of submartingales which arise under transformations of martingales.

Proposition 10.1. (a) If $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale and $\varphi(X_n)$ is a convex and integrable function of X_n, then $\{\varphi(X_n)\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale. (b)If $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, $\varphi(X_n)$ a convex increasing and integrable function of X_n, then $\{\varphi(X_n)\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale.

Proof. (a) Assume that φ is convex and differentiable on an interval I. Then

$$\varphi(v) \geq \varphi(u) + (v - u)\varphi'(u) \quad (u, v \in I), \tag{10.6}$$

i.e., the tangent line through any point $(u, \varphi(u))$ on the graph of φ lies below the graph of φ. Letting $v = X_{n+1}$ and $u = \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$, (10.6) becomes

$$\varphi(X_{n+1}) \geq \varphi(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) + (X_{n+1} - \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \cdot \varphi'(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)). \tag{10.7}$$

Now take conditional expectations of both sides, given $\mathcal{F}_n$, to get

$$\mathbb{E}(\varphi(X_{n+1}) \mid \mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) = \varphi(X_n). \tag{10.8}$$

If φ is not differentiable everywhere on I, then one may replace $\varphi'(u)$ in (10.6) by the left-hand derivative of φ at u.

(b) If $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, then instead of the last equality in (10.8) one gets an inequality, $\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) \geq \varphi(X_n)$, since $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ and φ is increasing. ■

As an immediate consequence of the proposition one gets

Corollary 10.2. Suppose $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$-submartingale.

a. Then, for every real c, $\{Y_t := \max(X_t, c)\}$ is a$\{\mathcal{F}_t : t \in T\}$-submartingale. In particular, $\{X_t^+ := \max(X_t, 0)\}$ is a $\{\mathcal{F}_t : t \in T\}$-submartingale.
b. If $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$-martingale, then $\{|X_t - c|\}$ is a $\{\mathcal{F}_t : t \in T\}$-submartingale, for every $c \in \mathbb{R}$.

Proposition 10.3 *(Discrete Parameter Doob–Meyer Decomposition).* Let $\{X_n\}_{n=0}^{\infty}$ be a $\{\mathcal{F}_n\}_{n=0}^{\infty}$-submartingale. Then $X_n \equiv M_n + A_n$ $(n \geq 0)$ is the sum of a $\{\mathcal{F}_n\}_{n=0}^{\infty}$-martingale $\{M_n\}_{n=0}^{\infty}$ and a non-decreasing process $\{A_n\}_{n=0}^{\infty}$, with $A_0 = 0$, which is predictable, i.e., A_n is $\mathcal{F}_{n-1}$-measurable $\forall\, n$. Moreover such a decomposition of $\{X_n\}_{n=0}^{\infty}$ is *unique*.

Proof. Write $Z_n := \mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1})$ $(n \geq 1)$, $A_n = \sum_{m=1}^{n} Z_m$, $A_0 = 0$. Then the submartingale property makes $\{A_n\}_{n=0}^{\infty}$ non-decreasing and the conditional expectations make it predictable. Now observe that $M_n := X_n - A_n$ $(n \geq 0)$ is a $\{\mathcal{F}_n\}_{n=0}^{\infty}$-martingale since for each $n \geq 1$,

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n = Z_n + X_{n-1} - Z_n - A_{n-1} = M_{n-1}.$$

If $X_n = M_n + A_n = M_n' + A_n'$ are two such decompositions, then $M_n - M_n'$ $(= A_n' - A_n)$ is $\mathcal{F}_{n-1}$-measurable, implying $M_n - M_n' = \mathbb{E}[M_n - M_n' \mid \mathcal{F}_{n-1}] = M_{n-1} - M_{n-1}'$, which may be inductively iterated to $M_n - M_n' = M_0 - M_0' = 0$. ■

Definition 10.2. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a martingale with $\mathbb{E}X_n^2 < \infty$. Then the predictable part A_n in the Doob–Meyer decomposition of the submartingale $\{X_n^2\}$ is called the *quadratic variation* of $\{X_n\}_{n=1}^{\infty}$ and denoted $A_n = \langle X_n \rangle$.

Example 7 (*Random Walk Revisited*)**.** Let $S_n = Z_1 + \cdots + Z_n$ $(n \geq 1)$ with Z_j's independent mean zero random variables, $\mathbb{E}Z_j^2 = \sigma_j^2 < \infty$. Then S_n is a martingale and $S_n^2 - \sum_{j=1}^n \sigma_j^2$ is a martingale. Since $\sum_{j=1}^n \sigma_j^2$ is non-decreasing non-negative process, this is the quadratic variation of S_n. The special form of $\langle S_n \rangle = \sum_{j=1}^n \sigma_j^2, n \geq 1$, in this case is due to the independence of the summands.

The next result is useful in proving L^1-convergence and in other contexts. To state it define a martingale or a submartingale $\{X_t : t \in T\}$ to be *closed on the right*, or *right-closed*, if T has a maximum, i.e., a largest element b. In particular, taking $Y = X_b \in L^1$, one has $X_t = \mathbb{E}(Y|\mathcal{F}_t), t \in T$, if $\{X_t : t \in T\}$ is such a right-closed martingale.

Proposition 10.4.

a. If $\{X_t : t \in T\}$ is a right-closed $\{\mathcal{F}_t : t \in T\}$-submartingale, then $\{\max(X_t, c) : t \in T\}$ is uniformly integrable for every $c \in \mathbb{R}$.
b. A right-closed $\{\mathcal{F}_t : t \in T\}$-martingale is uniformly integrable.

Proof. (a) Since $\max(X_t, c) = c + \max(X_t - c, 0) = c + (X_t - c)^+$, it is enough to prove the uniform integrability of the submartingale $\{Y_t := (X_t - c)^+\}$. Let t_R denote the largest element of T. Then the inequalities

$$\int_{[Y_t > \lambda]} Y_t dP \leq \int_{[Y_t > \lambda]} Y_{t_R} dP, \quad P(Y_t > \lambda) \leq \frac{\mathbb{E}(Y_t)}{\lambda} \leq \frac{\mathbb{E}(Y_{t_R})}{\lambda} \tag{10.9}$$

prove that $\{Y_t : t \in T\}$ is uniformly integrable.

(b) If $\{X_t : t \in T\}$ is a martingale, $\{|X_t| : t \in T\}$ is a non-negative right-closed submartingale, and (a) applies with $c = 0$. ■

It may be shown that the criterion for uniform integrability given above is optimal (i.e., essentially *necessary*) in the sense that if $\{X_t : t \in T\}$ is a uniformly integrable martingale (or, submartingale), then it is *closeable*; that is, there exists an integrable Y which may be used as the last element (to the right) of the martingale (submartingale); see Example 2.

Proposition 10.5. If a $\{\mathcal{F}_n\}$-martingale $\{X_n\}$ converges in L^1 to a random variable X, then the martingale is closed by X.

Proof. Fix $\epsilon > 0$ and m. Let $n_\epsilon \geq m$ such that $||X_n - X||_1 < \epsilon$ for all $n \geq n_\epsilon$. Then $||X_m - \mathbb{E}(X|\mathcal{F}_m)||_1 = ||\mathbb{E}(X_{n_\epsilon}|\mathcal{F}_m) - \mathbb{E}(X|\mathcal{F}_m)||_1 \leq ||X_{n_\epsilon} - X|| < \epsilon$. ■

Remark 10.2. It follows from Proposition 10.5 and Theorem 12.2 of a forthcoming chapter that a uniformly integrable martingale converges almost surely and in L^1 to a random variable X. Hence a martingale is closeable if and only if it is uniformly integrable.

Among the most remarkable consequences of Doob's martingale theory is the control over extremes in terms of moments as illustrated by the following.

***Theorem 10.6** (Doob's Maximal Inequality).* Let $\{X_1, X_2, \ldots, X_n\}$ be an $\{\mathcal{F}_k : 1 \leq k \leq n\}$-martingale, or a non-negative submartingale, and $\mathbb{E}|X_n|^p < \infty$ for some $p \geq 1$. Then, for all $\lambda > 0$, $M_n := \max\{|X_1|, \ldots, |X_n|\}$ satisfies

$$P(M_n \geq \lambda) \leq \frac{1}{\lambda^p} \int_{[M_n > \lambda]} |X_n|^p dP \leq \frac{1}{\lambda^p} \mathbb{E}|X_n|^p. \tag{10.10}$$

Proof. Let $A_1 = [|X_1| \geq \lambda]$, $A_k = [|X_1| < \lambda, \ldots, |X_{k-1}| < \lambda, \ |X_k| \geq \lambda]$ $(2 \leq k \leq n)$. Then $A_k \in \mathcal{F}_k$ and $[A_k : 1 \leq k \leq n]$ is a (disjoint) partition of $[M_n \geq \lambda]$. Therefore,

$$\begin{aligned} P(M_n \geq \lambda) &= \sum_{k=1}^{n} P(A_k) \leq \sum_{k=1}^{n} \frac{1}{\lambda^p} \mathbb{E}(\mathbf{1}_{A_k} |X_k|^p) \leq \sum_{k=1}^{n} \frac{1}{\lambda^p} E(\mathbf{1}_{A_k} |X_n|^p) \\ &= \frac{1}{\lambda^p} \int_{[M_n \geq \lambda]} |X_n|^p dP \leq \frac{\mathbb{E}|X_n|^p}{\lambda^p}. \end{aligned}$$

■

One may note that Kolmogorov's classic maximal inequality for mean zero random walks having square-integrable increments is an easy consequence of Doob's maximal inequality, (Exercise 7).

Corollary 10.7. Let $\{X_1, X_2, \ldots, X_n\}$ be an $\{\mathcal{F}_k : 1 \leq k \leq n\}$-martingale or nonnegative submartingale such that $\mathbb{E}|X_n|^p < \infty$ for some $p \geq 2$, and $M_n = \max\{|X_1|, \ldots, |X_n|\}$. Then $\mathbb{E}M_n^p \leq p^q \mathbb{E}|X_n|^p$, where q is the conjugate to p, $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. A standard application of the Fubini–Tonelli theorem provides the second moment formula

$$\mathbb{E}M_n^p = p \int_0^\infty x^{p-1} P(M_n > x) dx.$$

Noting that $p - 1 \geq 1$ to first apply the Doob maximal inequality (10.10), one then makes another application of the Fubini–Tonelli theorem, and finally the Hölder inequality, noting $pq - q = p$ for the conjugacy $\frac{1}{p} + \frac{1}{q} = 1$, to obtain

$$\begin{aligned} \mathbb{E}M_n^p &\leq p \int_0^\infty \mathbb{E}\left(|X_n|^{p-1} \mathbf{1}_{[M_n \geq x]} \right) dx = p\mathbb{E}\left(|X_n|^{p-1} M_n \right) \\ &\leq p(\mathbb{E}|X_n|^{(p-1)q})^{\frac{1}{q}} (\mathbb{E}M_n^p)^{\frac{1}{p}}. \end{aligned}$$

Divide both sides by $(\mathbb{E}|M_n|^p)^{\frac{1}{p}}$ and use monotonicity of $x \to x^{\frac{1}{q}}, x \geq 0$, to complete the proof. ■

A so-called L^p-maximal inequality ($p > 1$), or maximal moment inequality, was also obtained by Doob with a smaller constant $q^p \le p^q$ when $p \ge 2$, using a somewhat more clever estimation than in the above proof as follows.

Theorem 10.8 *(Doob's L^p-Maximal Inequality).* Let $\{X_1, X_2, \dots, X_n\}$ be an $\{\mathcal{F}_k : 1 \le k \le n\}$-martingale, or a non-negative submartingale, and let $M_n = \max\{|X_1|, \dots, |X_n|\}$. Then

1. $\mathbb{E}M_n \le \frac{e}{e-1}\big(1 + \mathbb{E}|X_n|\log^+|X_n|\big)$.
2. If $\mathbb{E}|X_n|^p < \infty$ for some $p > 1$, then $\mathbb{E}M_n^p \le q^p\mathbb{E}|X_n|^p$, where q is the conjugate exponent defined by $\frac{1}{q} + \frac{1}{p} = 1$, i.e., $q = \frac{p}{p-1}$.

Proof. For any non-decreasing function F_1 on $[0, \infty)$ with $F_1(0) = 0$, one may define a corresponding Lebesgue–Stieltjes measure $\mu_1(dy)$. Use the integration by parts formula[1] for a Lebesgue–Stieltjes integral to get

$$\begin{aligned}\mathbb{E}F_1(M_n) &= \int_{[0,\infty)} P(M_n \ge y)F_1(dy) \\ &\le \int_{[0,\infty)} \Big[\frac{1}{y}\int_{[M_n \ge y]} |X_n| dP\Big] F_1(dy) \\ &= \int_\Omega |X_n| \Big(\int_{[0,M_n]} \frac{1}{y} F_1(dy)\Big) dP, \end{aligned} \tag{10.11}$$

where the inequality follows from Theorem 10.6 (with $p = 1$). For the first part, consider the function $F_1(y) = y\mathbf{1}_{[1,\infty)}(y)$. Then $y - 1 \le F_1(y)$, and one gets

$$\begin{aligned}\mathbb{E}(M_n - 1) \le \mathbb{E}F_1(M_n) &\le \int_\Omega |X_n| \Big(\int_{[1,\max\{1,M_n\}]} \frac{1}{y} dy\Big) dP \\ &= \int_\Omega |X_n| \log(\max\{1, M_n\}) dP \\ &= \int_{[M_n \ge 1]} |X_n| \log M_n dP. \end{aligned} \tag{10.12}$$

Now use the inequality (proved in the remark below)

$$a \log b \le a \log^+ a + \frac{b}{e}, \quad a, b \ge 0, \tag{10.13}$$

to further arrive at

$$\mathbb{E}M_n \le 1 + \mathbb{E}|X_n| \log^+ |X_n| + \frac{\mathbb{E}M_n}{e}. \tag{10.14}$$

[1] See BCPT, Proposition 1.4, p. 10.

This establishes the inequality for the case $p = 1$. For $p > 1$ take $F_1(y) = y^p$. Then

$$\begin{aligned}
\mathbb{E}M_n^p &\le \mathbb{E}\big(|X_n| \int_{[0,M_n]} p y^{p-2} dy\big) \\
&= \mathbb{E}\big(|X_n| \frac{p}{p-1} M_n^{p-1}\big) \\
&\le \frac{p}{p-1} (\mathbb{E}|X_n|^p)^{\frac{1}{p}} (\mathbb{E}M_n^{(p-1)q})^{\frac{1}{q}} \\
&= q (\mathbb{E}|X_n|^p)^{\frac{1}{p}} (\mathbb{E}M_n^p)^{\frac{1}{q}}. \qquad (10.15)
\end{aligned}$$

The bound for $p > 1$ now follows by dividing by $(\mathbb{E}M_n^p)^{\frac{1}{q}}$ and a little algebra. ■

Remark 10.3. To prove the inequality (10.13) it is sufficient to consider the case $1 < a < b$, since it obviously holds otherwise. In this case it may be expressed as

$$\log b \le \log a + \frac{b}{ae},$$

or

$$\log \frac{b}{a} \le \frac{b}{ae}.$$

But this follows from the fact that $f(x) = \frac{\log x}{x}$, $x > 1$, has a maximum value $\frac{1}{e}$.

Corollary 10.9. Let $\{X_t : t \in [0,T]\}$ be a right-continuous non-negative $\{\mathcal{F}_t\}$-submartingale with $\mathbb{E}|X_T|^p < \infty$ for some $p \ge 1$. Then $M_T := \sup\{X_s : 0 \le s \le T\}$ is $\mathcal{F}_T$-measurable and, for all $\lambda > 0$,

$$P(M_T > \lambda) \le \frac{1}{\lambda^p} \int_{[M_T > \lambda]} X_T^p dP \le \frac{1}{\lambda^p} \mathbb{E}X_T^p. \qquad (10.16)$$

Proof. Consider the non-negative submartingale $\{X_0, \ldots, X_{Ti2^{-n}}, \ldots, X_T\}$, for each $n = 1, 2, \ldots$, and let $M_n := \max\{X_{iT2^{-n}} : 0 \le i \le 2^n\}$. For $\lambda > 0$, $[M_n > \lambda] \uparrow [M_T > \lambda]$ as $n \uparrow \infty$. In particular, M_T is $\mathcal{F}_T$-measurable. By Theorem 10.6,

$$P(M_n > \lambda) \le \frac{1}{\lambda^p} \int_{[M_n > \lambda]} X_T^p dP \le \frac{1}{\lambda^p} \mathbb{E}X_T^p.$$

Letting $n \uparrow \infty$, (10.16) is obtained using Proposition 10.4. ■

The final notion to be introduced in this chapter is that of a martingale reversed in time. First consider a finite filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_m$ and a $\{\mathcal{F}_n : n = 1, 2, \ldots, m\}$ martingale $\{X_n : n = 1, 2, \ldots, m\}$. Then denote the reversed sequence by $\mathcal{R}_1 := \mathcal{F}_m$, $\mathcal{R}_2 := \mathcal{F}_{m-1}, \ldots, \mathcal{R}_m = \mathcal{F}_1$, and $Y_1 := X_m$, $Y_2 :=$

$X_{m-1}, \dots, Y_m := X_1$. Then $\{Y_n : n = 1, 2, \dots, m\}$ is called a $\{\mathcal{R}_n : 1 \le n \le m\}$-*reverse martingale*, in the sense that (i) $\{\mathcal{R}_n\}_{n=1}^{\infty}$ is a *decreasing* sequence of σ-fields, (ii) Y_n is $\mathcal{R}_n$-measurable $\forall\, n$, and (iii) $\mathbb{E}(Y_n \mid \mathcal{F}_{n+1}) = Y_{n+1}\ \forall n$. More generally, $\{Y_n\}_{n \ge 1}$ is called a $\{\mathcal{R}_n\}_{n\ge1}$-*reverse martingale* if (i)–(iii) hold for a finite or an infinite sequence of indices $\{1, 2, \dots, m\}$ or $\mathbb{N} = \{1, 2, \dots\}$. In particular an infinite sequence $\{Y_n\}_{n=1}^{\infty}$ is a reverse martingale if and only if the reversal in time of every finite segment of it (for example, reversal of the random variables $Y_j, Y_{j+1}, \dots, Y_{j+k}$ and σ-fields $\mathcal{R}_j, \mathcal{R}_{j+1}, \dots, \mathcal{R}_{j+k}$, say) is a martingale.

Example 8 (*Conditional Expectations Revisited***).** Let X be integrable and $\{\mathcal{R}_n : n \in \mathbb{N}\}$ a decreasing sequence of sub σ-fields of $\mathcal{F}$. Then $Y_n := \mathbb{E}(X \mid \mathcal{R}_n)$ $(n \in \mathbb{N})$ is a $\{\mathcal{R}_n : n \in \mathbb{N}\}$-reverse martingale.

Exercises

1. Let f be a real-valued step function on $[0, \infty)$ (into $\mathbb{R}$) : $f(t) = f(t_i)$ for $t_i < t \le t_{i+1}$ $(i = 0, 1, 2, \dots)$ where $0 = t_0 < t_1 < t_2 < \cdots$. Let $\{B_t : t \ge 0\}$ be a standard Brownian motion and define the *stochastic integral* $X_t \equiv \int_0^t f(s)dB_s = \sum_{i=0}^{m-1} f(t_i)(B_{t_{i+1}} - B_{t_i}) + f(t_m)(B_t - B_{t_m})$ for $t \in (t_m, t_{m+1}]$.
   (a) Show that $\{X_t : 0 \le t < \infty\}$ is a $\{\mathcal{F}_t : 0 \le t < \infty\}$-martingale with $\mathcal{F}_t := \sigma\{B_s : 0 \le s \le t\}$ $(t \ge 0)$, and
   (b) the $\{\mathcal{F}_t : 0 \le t < \infty\}$-submartingale $\{X_t^2 : t \ge 0\}$ has the Doob–Meyer decomposition: $X_t^2 = M_t + A_t$, where $A_t = \int_0^t f^2(s)ds = \sum_{i=0}^{m-1} f^2(t_i)(t_{i+1} - t_i) + f^2(t_m)(t - t_m)$ for $t \in (t_m, t_{m+1}]$ $A_0 = 0$, and $M_t = X_t^2 - A_t$ $(t \ge 0)$ is a $\{\mathcal{F}_t : 0 \le t < \infty\}$-martingale.
   (c) Extend (a), (b) to the case of a (random) *nonanticipative* locally bounded *step function*, i.e., $f(t_i)$ is $\mathcal{F}_{t_i}$-measurable $\forall\, i$ and $|f(t)| \le C_T$ for $0 \le t \le T$, and C_T are nonrandom constants for all $0 < t < \infty$.

2. (*Exponential Martingale*) Let $\{B_t : t \ge 0\}$ be a standard one-dimensional Brownian motion starting at zero, $\mathcal{F}_t := \sigma\{B_s : 0 \le s \le t\}$.
   (a) Show that for each pair $x, \xi \in \mathbb{R}$, $X_t := \exp\{x + \xi B_t - (\xi^2/2)t\}$ $(t \ge 0)$ is a $\{\mathcal{F}_t : 0 \le t < \infty\}$-martingale.
   (b) For the stochastic integrals in Exercise 1, prove that $X_t := \exp\{x + \int_0^t f(s)dB_s - \frac{1}{2}\int_0^t f^2(s)ds\}$ $(t \ge 0)$ is a $\{\mathcal{F}_t : 0 \le t < \infty\}$-martingale.

3. (*Likelihood Ratios*) Let P^0, P^1 be two probability measures on $(\Omega, \mathcal{F})$ and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ a filtration. Let f_n^0, f_n^1 be the densities of P^0, P^1, respectively, on $\mathcal{F}_n$ with respect to a measure ν_n (e.g., ν_n is the restriction of $P^0 + P^1$ on $\mathcal{F}_n$). Show that $\{f_n^1/f_n^0 : n \ge 1\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale under P^0.

   (For example, P^i may be the distribution of a stochastic process $\{Y_n : n \ge 1\}$ under the hypothesis H^i $(i = 0, 1)$ and $\mathcal{F}_n := \sigma\{Y_1, \dots, Y_n\}, n \ge 1$.)

4. (a) Show that the sum of two $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingales (or submartingales) is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale (resp., submartingales),
   (b) Give an example to show that the sum of a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale and a $\{\mathcal{G}_n\}_{n=1}^{\infty}$-martingale, with $\mathcal{G}_n \subset \mathcal{F}_n \ \forall n$, is not, in general, a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale. [*Hint*: A $\{\mathcal{G}_n\}_{n=1}^{\infty}$-martingale is not necessarily a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale].
5. (*Exchangeable Martingale Differences*) Let $\{Q_\theta : \theta \in \Phi\}$ be a family of probability measures on $(\mathbb{R}^1, \mathcal{B}^1)$ such that $\int x\, Q_\theta(dx) = 0$ for each θ. Suppose Φ is a Borel subset of $\mathbb{R}^k$, and π a probability measure on $(\Phi, \mathcal{B}(\Phi))$. Consider an experiment in which a value of θ is chosen at random according to $\pi(d\theta)$ and then, conditionally given θ, an *i.i.d.* sequence $\{Z_{\theta,n}\}$ is chosen having common distribution Q_θ.

   (a) Show that $\{Z_{\theta,n} : n \geq 1\}$ is an *exchangeable sequence*, i.e., its distribution is invariant under any permutation of the indices n.
   (b) Show that $\{X_n := \sum_1^n Z_{\theta,j}, n \geq 1\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale, where $\mathcal{F}_n = \sigma\{\theta, Z_{\theta,1}, \ldots, Z_{\theta,n})$.

6. Let $\{X_n\}_{n=1}^{\infty}$ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-supermartingale. Prove that if φ is concave and increasing on the range of X_n ($n \geq 1$), and $\varphi(X_n)$ is integrable, then $\{\varphi(X_n)\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-supermartingale.
7. Suppose that $S_n = X_1 + \cdots + X_n, n \geq 1$, where $X_1, X_2, \ldots$ are i.i.d. with mean zero and finite variance. Use Doob's maximal inequality to prove *Kolmogorov's maximal inequality* $P(\max_{1\leq j\leq n} |S_j| \geq \lambda) \leq \frac{\mathbb{E}|S_n|^2}{\lambda^2}, \lambda > 0$.

Chapter 11
Optional Stopping of (Sub)Martingales

The development of martingale theory is continued for discrete time martingales with a focus on the use of stopping times in their analysis. An application to the ruin problem in insurance is included as an application.

In this chapter we consider discrete parameter processes. Extensions to the continuous parameter martingales and submartingales are given in Chapter 13. Since every martingale is a submartingale and versions of some theorems apply to both, we sometimes refer to (sub)martingales to include both, possibly with equalities replaced by inequalities.

Recall that a random variable τ with values in $\mathbb{Z}_+ \cup \{\infty\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time with respect to a filtration $\{\mathcal{F}_n\}_{n=1}^{\infty}$ if $[\tau \le n] \in \mathcal{F}_n$ for all $n \in \mathbb{Z}_+$. Since $\mathcal{F}_n$ are increasing, τ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time if and only if $[\tau = n] \in \mathcal{F}_n$ for all $n \in \mathbb{Z}_+$. Informally, τ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time if whether or not to stop at time n depends only on the past and present information as embodied in $\mathcal{F}_n$.

Events depending only on times up to a stopping time τ comprise the so-called *pre-τ σ-field*, defined more precisely as follows.

Definition 11.1. Let τ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time. The *pre-τ σ-field* $\mathcal{F}_\tau$ comprises all $F \in \mathcal{F}$ satisfying $F \cap [\tau \le n] \in \mathcal{F}_n$ for every $n = 0, 1, 2, \dots$.

It is clear from this definition, by taking $F = [\tau \le m]$ $(m \in \mathbb{Z}_+)$, that τ is $\mathcal{F}_\tau$-measurable. To get an intuitive feeling for $\mathcal{F}_\tau$, consider a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ with values in some measurable space $(S, \mathcal{S})$, and let τ be $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time with $\mathcal{F}_n := \sigma\{X_0, X_1, \dots, X_n\}$. One may check in this case that $\mathcal{F}_\tau$ is generated by the *stopped process* $\{X_{\tau\wedge n} : n = 0, 1, 2, \cdots\}$, which is

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_11

the process $\{X_n\}_{n=1}^{\infty}$ observed up to time τ. In other words (Exercise 1),

$$\mathcal{F}_\tau = \sigma\{X_{\tau\wedge n} : n = 0, 1, 2, \cdots\}. \tag{11.1}$$

More generally, for a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-adapted process $\{X_n\}_{n=1}^{\infty}$ (with values in some measurable space $(S, \mathcal{S})$), one has, for every $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time τ, the relation (Exercise 1)

$$\mathcal{F}_\tau \supset \sigma\{X_{\tau\wedge n} : n = 0, 1, 2, \ldots\}. \tag{11.2}$$

Also, if $\tau_1 \leq \tau_2$ are two $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times, it is simple to check from Definition 11.1 that (Exercise 2)

$$\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}. \tag{11.3}$$

In the case $\tau < \infty$ a.s., we will often write X_τ for $X_\tau \mathbf{1}_{[\tau<\infty]}$, which is easily seen to be $\mathcal{F}_\tau$-measurable (Exercise 2).

***Theorem 11.1** (Optional Stopping Theorem).* Let $\tau_1 \leq \tau_2$ be two a.s. finite $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times and $\{X_n\}_{n=1}^{\infty}$ a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale. Assume that (i) $\mathbb{E}|X_{\tau_i}| < \infty$ $(i = 1, 2)$, and (ii) $\lim_{m\to\infty} \mathbb{E}|X_m \mathbf{1}_{[\tau_2>m]}| = 0$. Then

$$\mathbb{E}\left(X_{\tau_2} \mid \mathcal{F}_{\tau_1}\right) \geq X_{\tau_1} \quad \text{a.s.,} \quad \text{and} \quad \mathbb{E}X_{\tau_2} \geq EX_{\tau_1}. \tag{11.4}$$

In the case $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale, the inequalities in (11.4) become equalities.

We first prove a simple lemma.

Lemma 1. Let $\tau_1 \leq \tau_2$ be two a.s. finite $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times and $\{X_n\}_{n=1}^{\infty}$ a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale. Then conditions (i) and (ii) are equivalent to

$$X_{\tau_i\wedge m} \longrightarrow X_{\tau_i} \quad \text{in } L^1 \quad \text{as } m \to \infty \ (i = 1, 2). \tag{11.5}$$

Proof. Assume (i), (ii). Then $\mathbb{E}|X_{\tau_2\wedge m} - X_{\tau_2}| = \mathbb{E}|(X_m - X_{\tau_2})\mathbf{1}_{[\tau_2>m]}| \leq \mathbb{E}|X_m \mathbf{1}_{[\tau_2>m]}| + \mathbb{E}|X_{\tau_2}\mathbf{1}_{[\tau_2>m]}| \to 0$ as $m \to \infty$, by (i) and (ii). Also, $\mathbb{E}|X_{\tau_1\wedge m} - X_{\tau_1}| = \mathbb{E}|(X_m - X_{\tau_1})\mathbf{1}_{[\tau_1>m]}| \leq \mathbb{E}|X_m\mathbf{1}_{[\tau_1>m]}| + \mathbb{E}|X_{\tau_1}\mathbf{1}_{[\tau_1>m]}| \leq \mathbb{E}|X_m \mathbf{1}_{[\tau_2>m]} + \mathbb{E}|X_{\tau_1}\mathbf{1}_{[\tau_1>m]}| \to 0$ as $m \to \infty$. Conversely, suppose (11.5) holds, which obviously implies (i). As to (ii),

$$\mathbb{E}|X_m \mathbf{1}_{[\tau_2>m]}| = \mathbb{E}|(X_{\tau_2\wedge m} - X_{\tau_2})\mathbf{1}_{[\tau_2>m]} + X_{\tau_2}\mathbf{1}_{[\tau_2>m]}| \longrightarrow 0$$

by (11.5) and (i). ■

Proof of Theorem 11.1. First assume τ_1 and τ_2 are bounded a.s., i.e., $\tau_2 \leq m$ a.s. for some integer m. Fix $F \in \mathcal{F}_{\tau_1}$ and an integer j, $0 \leq j \leq m$. Then, writing

$Z_j = X_j - X_{j-1}$ ($j \geq 1$) and $X_r = X_j + Z_{j+1} + \cdots + Z_r$ for $j > r$, one has

$$\begin{aligned}\mathbb{E}X_{\tau_2}\mathbf{1}_{F\cap[\tau_1=j]} &= \mathbb{E}X_j\mathbf{1}_{[\tau_2=j]}\mathbf{1}_{F\cap[\tau_1=j]} \\ &\quad + \sum_{r=j+1}^{m}\mathbb{E}(X_j + Z_{j+1} + \cdots + Z_r)\mathbf{1}_{[\tau_2=r]}\mathbf{1}_{F\cap[\tau_1=j]} \qquad (11.6)\\ &= \mathbb{E}\left(X_j\mathbf{1}_{[\tau_2\geq j]} + Z_{j+1}\mathbf{1}_{[\tau_2\geq j+1]} + \cdots + Z_m\mathbf{1}_{[\tau_2\geq m]}\right)\mathbf{1}_{F\cap[\tau_1=j]}.\end{aligned}$$

Since $[\tau_2 \geq r] = [\tau_2 \leq r-1]^c \in \mathcal{F}_{r-1}$, and $\mathbb{E}(Z_r|\mathcal{F}_{r-1}) \geq 0$ for a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, we get

$$\mathbb{E}Z_r\mathbf{1}_{[\tau_2\geq r]}\mathbf{1}_{F\cap[\tau_1=j]} = \mathbb{E}\mathbf{1}_{[\tau_2\geq r]}\mathbf{1}_{F\cap[\tau_1=j]}\mathbb{E}(Z_r|\mathcal{F}_{r-1}) \geq 0. \qquad (11.7)$$

Using (11.7) in (11.6), and noting that $[\tau_2 \geq r] \cap [\tau_1 = j] = [\tau_1 = j]$ a.s., one obtains

$$\begin{aligned}\mathbb{E}X_{\tau_2}\mathbf{1}_{F\cap[\tau_1=j]} \geq \mathbb{E}X_j\mathbf{1}_{[\tau_2\geq j]}\mathbf{1}_{F\cap[\tau_1=j]} &= \mathbb{E}X_{\tau_1}\mathbf{1}_{[\tau_2\geq j]}\mathbf{1}_{F\cap[\tau_1=j]}\\ &= \mathbb{E}X_{\tau_1}\mathbf{1}_{F\cap[\tau_1=j]}. \qquad (11.8)\end{aligned}$$

Now sum over j to get the desired result

$$\mathbb{E}X_{\tau_2}\mathbf{1}_F \geq \mathbb{E}X_{\tau_1}\mathbf{1}_F \quad \text{for all } F \in \mathcal{F}_{\tau_1}. \qquad (11.9)$$

For a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale $\{X_n\}_{n=1}^{\infty}$, $\mathbb{E}(Z_r|\mathcal{F}_{r-1}) = 0$ a.s. for every r, so that one has equalities in (11.7)–(11.9).

For the general case, apply (11.9) to the stopping times $\tau_i \wedge m$ ($i = 1, 2$), and note that if $F \in \mathcal{F}_{\tau_1}$, then $F \cap [\tau_1 \leq m] \in \mathcal{F}_{\tau_1\wedge m}$ (see Exercise 2(b)) to get

$$\mathbb{E}X_{\tau_2\wedge m}\mathbf{1}_{F\cap[\tau_1\leq m]} \geq \mathbb{E}X_{\tau_1\wedge m}\mathbf{1}_{F\cap[\tau_1\leq m]} \quad (F \in \mathcal{F}_{\tau_1}). \qquad (11.10)$$

By the lemma, it now follows, letting $m \to \infty$ in (11.10), that $\mathbb{E}X_{\tau_2}\mathbf{1}_F \geq \mathbb{E}X_{\tau_1}\mathbf{1}_F$ ($F \in \mathcal{F}_{\tau_1}$), with equality if $\{X_n\}_{n=1}^{\infty}$ is a martingale. ■

Corollary 11.2. Let $\{X_n\}_{n=1}^{\infty}$ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, and suppose $\tau_1 \leq \tau_2$ are two $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times. If τ_2 is bounded a.s., then the conclusions of Theorem 11.1 hold.

The technical conditions (i) and (ii) of Theorem 11.1 can also be verified under the hypothesis of the following version of the result.

Corollary 11.3. Let $\{X_n\}_{n=1}^{\infty}$ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, and suppose $\tau_1 \leq \tau_2$ are two a.s. finite $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times. Assume that the sequence $Z_n := X_n - X_{n-1}$ satisfies the two conditions: (a) there exist constants c_n such that $\mathbb{E}(|Z_n| \mid \mathcal{F}_{n-1}) \leq c_n$ a.s. on $[\tau_2 \geq n]$ ($n = 1, 2, \ldots$) and (b) $\mathbb{E}(c_1 + c_2 + \cdots + c_{\tau_2})\mathbf{1}_{[\tau_2\geq 1]} < \infty$. Then the conclusions of Theorem 11.1 hold.

Proof. Assumption (i) of Theorem 11.1 follows from the relations

$$\mathbb{E}|X_{\tau_2}| \leq \mathbb{E}|X_0| + \mathbb{E}\left(\sum_{r=1}^{\tau_2} |Z_r| \mathbf{1}_{[\tau_2 \geq 1]}\right),$$

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{r=1}^{\tau_2} |Z_r| \mathbf{1}_{[\tau_2 \geq 1]}\right)\right] &= \sum_{n=1}^{\infty} \mathbb{E}\left[\sum_{r=1}^{n} |Z_r| \mathbf{1}_{[\tau_2 = n]}\right] \\ &= \sum_{r=1}^{\infty} \mathbb{E}|Z_r| \mathbf{1}_{[\tau_2 \geq r]} = \sum_{r=1}^{\infty} \mathbb{E}\left[\mathbf{1}_{[\tau_2 \geq r]} \mathbb{E}(|Z_r| \mid \mathcal{F}_{r-1})\right] \\ &\leq \sum_{r=1}^{\infty} \mathbb{E}\left[\mathbf{1}_{[\tau_2 \geq r]} c_r\right] = \mathbb{E}\left(c_1 + c_2 + \cdots + c_{\tau_2}\right) \mathbf{1}_{[\tau_2 \geq 1]} < \infty,\end{aligned}$$

where we have used the fact $[\tau_2 \leq r] = [\tau_2 \leq r-1]^c \in \mathcal{F}_{r-1}$.

To verify assumption (ii) of Theorem 11.1, note that for all $k < n$,

$$\begin{aligned}\mathbb{E}|X_m| \mathbf{1}_{[\tau_2 \geq m]} &\leq \mathbb{E}\left(|X_0| + |Z_1| + \cdots + |Z_k|\right) \mathbf{1}_{[\tau_2 \geq m]} + \sum_{r=k+1}^{m} \mathbb{E}|Z_r| \mathbf{1}_{[\tau_2 \geq r]} \\ &\leq \mathbb{E}\left(|X_0| + |Z_1| + \cdots + |Z_k|\right) \mathbf{1}_{[\tau_2 \geq m]} + \mathbb{E}\left(\sum_{r=k+1}^{m} c_r \mathbf{1}_{[\tau_2 \geq r]}\right).\end{aligned}$$

Hence, for each k, noting that in the second sum $[\tau_2 \geq r]$ implies $[\tau_2 \geq k+1]$,

$$\limsup_{m \to \infty} \mathbb{E}|X_m| \mathbf{1}_{[\tau_2 \geq m]} \leq \mathbb{E} \sum_{r=k+1}^{\infty} c_r \mathbf{1}_{[\tau_2 \geq r]} \leq \mathbb{E}(c_1 + \cdots + c_{\tau_2}) \mathbf{1}_{[\tau_2 \geq k+1]},$$

which goes to zero as $k \to \infty$. ■

It may be noted that if $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-supermartingale, then $\{-X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale, and vice versa. Hence Theorem 11.1 and Corollaries 11.2 and 11.3 apply to supermartingales with the inequality in (11.4) reversed.

The following proposition and its corollary are often useful for verifying the hypothesis of Theorem 11.1 in examples.

Proposition 11.4. Let $\{Z_n : n \in \mathbb{N}\}$ be real-valued random variables such that for some $\varepsilon > 0$, $\delta > 0$, one has

$$P(Z_{n+1} > \varepsilon \mid \mathcal{G}_n) \geq \delta, \ \text{a.s.} \quad \forall\, n = 0, 1, 2, \ldots$$

or

$$P(Z_{n+1} < -\varepsilon \mid \mathcal{G}_n) \geq \delta \text{ a.s.} \quad \forall\, n = 0, 1, 2, \ldots, \tag{11.11}$$

where $\mathcal{G}_n = \sigma\{Z_1, \ldots, Z_n\}$ $(n \geq 1)$, $\mathcal{G}_0 = \{\emptyset, \Omega\}$. Let $S_n^x = x + Z_1 + \cdots + Z_n$ $(n \geq 1)$, $S_0^x = x$, and let $a < x < b$. Let τ be the first escape time of $\{S_n^x\}_{n=0}^\infty$ from (a, b), i.e., $\tau = \inf\{n \geq 1 : S_n^x \in (a, b)^c\}$. Then $\tau < \infty$ a.s., and

$$\sup_{\{x : a < x < b\}} \mathbb{E}e^{\tau z} < \infty \quad \text{for } |z| < \frac{1}{n_0}\left(\log \frac{1}{1 - \delta_0}\right), \tag{11.12}$$

where, writing $[y]$ for the integer part of y,

$$n_0 = \left[\frac{b - a}{\varepsilon}\right] + 1 \qquad \delta_0 = \delta^{n_0}. \tag{11.13}$$

Proof. Suppose the first relation in the proposition holds. Clearly, if $Z_j > \varepsilon \forall\, j = 1, 2, \ldots, n_0$, then $S_{n_0}^x > b$, so that $\tau \leq n_0$. Therefore, $P(\tau \leq n_0) \geq P(Z_1 > \varepsilon, \ldots, Z_{n_0} > \varepsilon) \geq \delta^{n_0}$, by taking successive conditional expectations (given $\mathcal{G}_{n_0 - 1}, \mathcal{G}_{n_0 - 2}, \ldots, \mathcal{G}_0$, in that order). Hence $P(\tau > n_0) \leq 1 - \delta^{n_0} = 1 - \delta_0$. For every integer $k \geq 2$, $P(\tau > kn_0) = P(\tau > (k-1)n_0, \tau > kn_0) = \mathbb{E}[\mathbf{1}_{[\tau > (k-1)n_0]} P(\tau > kn_0 | \mathcal{G}_{(k-1)n_0})] \leq (1 - \delta_0) P(\tau > (k-1)n_0)$, since, on the set $[\tau > (k-1)n_0]$, $P(\tau \leq kn_0 | \mathcal{G}_{(k-1)n_0}) \geq P(Z_{(k-1)n_0 + 1} > \varepsilon, \ldots, Z_{kn_0} > \varepsilon | \mathcal{G}_{(k-1)n_0}) \geq \delta^{n_0} = \delta_0$. Hence, by induction, $P(\tau > kn_0) \leq (1 - \delta_0)^k$. Hence $P(\tau = \infty) = 0$ and, for all $z > 0$,

$$\begin{aligned}
\mathbb{E}e^{z\tau} &= \sum_{r=1}^{\infty} e^{zr} P(\tau = r) \leq \sum_{k=1}^{\infty} e^{zkn_0} \sum_{r=(k-1)n_0+1}^{kn_0} P(\tau = r) \\
&\leq \sum_{k=1}^{\infty} e^{zkn_0} P(\tau > (k-1)n_0) \leq \sum_{k=1}^{\infty} e^{zkn_0} (1 - \delta_0)^{k-1} \\
&= e^{zn_0}(1 - (1 - \delta_0)e^{zn_0})^{-1} \quad \text{if } e^{zn_0}(1 - \delta_0) < 1.
\end{aligned}$$

An entirely analogous argument holds if the second relation in the proposition holds. ■

Corollary 11.5 *(Stein's Lemma).* Let $\{Z_n : n = 1, 2, \cdots\}$ be an i.i.d. sequence such that $P(Z_1 = 0) < 1$. Let $S_n^n = x + Z_1 + \cdots + Z_n$ $(n \geq 1)$, $S_0^x = x$, and $a < x < b$. Then the first escape time τ of the random walk from the interval (a, b) has a finite moment generating function in a neighborhood of 0. In particular, there is a constant $c_b > 0$ such that $P(S_n < b) \leq e^{-c_b n}$.

Proof. Finiteness of the moment generating function is an immediate consequence. The remaining assertion follows from the bound $P(S_n < b) \le P(\tau > n)$ and the estimates given in the proof of Proposition 11.4. ■

Example 1 (*Wald's Identities*). Consider a general random walk $\{S_n^x := x + Z_1 + \cdots + Z_n\}_{n=1}^{\infty}$ as in Corollary 11.5. Assume $\mathbb{E}Z_1 = \mu$ is finite. Then $\{S_n^x - n\mu\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale, where $\mathcal{F}_n = \sigma\{Z_1, \ldots, Z_n\}$ $(n \ge 1)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Let τ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time such that $\mathbb{E}\tau < \infty$. Then $\mathbb{E}[|S_n^x - n\mu - (S_{n-1}^x - (n-1)\mu| \mid \mathcal{F}_{n-1}] = \mathbb{E}[|Z_n - \mu| \mid \mathcal{F}_{n-1}] = \mathbb{E}|Z_n - \mu| = c$, say, is finite. Hence, the hypothesis of Corollary 11.3 is satisfied: $\mathbb{E}(c_1 + c_2 + \cdots + c_\tau)\mathbf{1}_{[\tau \ge 1]} \le \mathbb{E}c\tau = c\mathbb{E}\tau < \infty$, with $c_j \equiv c$, $\tau_1 \equiv 0$, $\tau_2 = \tau$. This yields
Wald's First Identity: If $\mathbb{E}\tau < \infty$ and Z_1 has a finite mean, then

$$\begin{aligned} \mathbb{E}(S_\tau^x - \tau\mu) = \mathbb{E}(S_0^x) = x, \quad \text{i.e.,} \\ \mathbb{E}S_\tau^x = x + \mu\mathbb{E}\tau. \end{aligned} \tag{11.14}$$

Next write $S_0' = 0$, $S_n' = S_n^x - x - n\mu = Z_1 - \mu + \cdots + Z_n - \mu$ $(n \ge 1)$. Assume $\sigma^2 := \operatorname{Var} Z_1 < \infty$. Hence $\{X_n := (S_n')^2 - n\sigma^2 : n \ge 0\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-martingale. *If* τ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time such that (i)$'$ $\mathbb{E}\tau < \infty$, (ii)$'$ $\mathbb{E}(S_\tau')^2 < \infty$, and (iii)$'$ $\mathbb{E}(S_m')^2\mathbf{1}_{[\tau > m]} \to 0$ as $m \to \infty$, then, by Theorem 11.1, one obtains $\mathbb{E}X_\tau = \mathbb{E}X_0 = 0$. Equivalently,
Wald's Second Identity: If conditions (i)$'$–(iii)$'$ hold, then

$$\mathbb{E}(S_\tau')^2 = \sigma^2\mathbb{E}\tau. \tag{11.15}$$

As a special case, let $a < x < b$ and assume Z_1 is bounded a.s., $\operatorname{Var} Z_1 > 0$. Define τ to be the first escape time of $\{S_n^x : n \ge 0\}$ from the interval (a, b). By Corollary 11.5, $\mathbb{E}\tau < \infty$ and (i)$'$–(iii)$'$ above are easily verified. Hence (11.14) and (11.15) both hold.

Example 2 (*Symmetric Simple Random Walk: Boundary Distribution and Expected Hitting Time*). Let

$$\tau = \min\{n : S_n = -a \text{ or } b\}, \tag{11.16}$$

where a and b are positive integers, and then using Stein's lemma, one has $P(\tau < \infty) = 1$. Note that $|S_\tau| \le \max\{a, b\}$, so that $\mathbb{E}|S_\tau| \le \max\{a, b\}$. Also, on the set $[\tau > m]$, one has $-a < S_m < b$, and therefore

$$\begin{aligned} |\mathbb{E}(S_m\mathbf{1}_{[\tau > m]})| &\le \max\{a, b\}\mathbb{E}\mathbf{1}_{[\tau > m]} \\ &= \max\{a, b\}P(\tau > m) \to 0 \qquad \text{as } m \to \infty. \end{aligned} \tag{11.17}$$

Thus the optional stopping Theorem 11.1 applies with $\tau_1 \equiv 0 \le \tau = \tau_2$, and therefore

$$
\begin{aligned}
0 = \mathbb{E}S_\tau &= -aP(\tau_{-a} < \tau_b) + bP(\tau_{-a} > \tau_b) \\
&= -a(1 - P(\tau_{-a} > \tau_b)) + bP(\tau_{-a} > \tau_b),
\end{aligned} \tag{11.18}
$$

which may be solved for $P(\tau_{-a} > \tau_b)$ to yield

$$
P(\tau_{-a} > \tau_b) = \frac{a}{a+b}. \tag{11.19}
$$

More generally, starting the random walk at $x \in (a, b)$, one has

$$
\psi(x) \equiv P_x(\tau_a < \tau_b) = (b-x)/(b-a) \qquad a \le x \le b. \tag{11.20}
$$

Recall that this was obtained from a discrete boundary value equation as Proposition 2.3 by conditioning on S_1, and in Chapter 7, Example 1, using the strong Markov property. Note that the optional stopping of martingales also yields the expected time to reach the boundary $\{a, b\}$ from Wald's second identity (Exercise 4) as

$$
\mathbb{E}_x\tau = (x-a)(b-x) \qquad a \le x \le b. \tag{11.21}
$$

Example 3 (*Asymmetric Simple Random Walk: Boundary Distribution and Expected Hitting Time*)**.** Consider the asymmetric simple random walk starting at an integer x, i.e., $S_n = S_n^x = x + Z_1 + \cdots + Z_n$ $(n \ge 0)$ with $P(Z_n = +1) = p$ and $P(Z_n = -1) = q \equiv 1 - p$ $(0 < p < 1)$. Write $X_n = (q/p)^{S_n}$ $(n \ge 0)$, $\mathcal{F}_n = \sigma\{Z_1, \ldots, Z_n\}$ $(n \ge 1)$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$. Then $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = (q/p)^{S_n}\mathbb{E}[(q/p)^{Z_{n+1}}|\mathcal{F}_n] = (q/p)^{S_n}\mathbb{E}(q/p)^{Z_{n+1}} = (q/p)^{S_n}$, since $\mathbb{E}(q/p)^{Z_{n+1}} = p(q/p) + q(q/p)^{-1} = q + p = 1$. Hence $\{X_n\}_{n=1}^\infty$ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-martingale. Now let a and b be two integers, $a < x < b$, and $\tau = \inf\{n \ge 0 : S_n \in \{a, b\}\}$. The hypothesis of Theorem 11.1 holds, so that $\mathbb{E}X_\tau = \mathbb{E}X_0 = (q/p)^x$. That is, writing $\psi(x)$ for the probability $\{S_n\}_{n=0}^\infty$ reaches a before b starting from x,

$$
(q/p)^x = (q/p)^a\psi(x) + (q/p)^b(1 - \psi(x)).
$$

This yields

$$
P(\{S_n^x\} \text{ hits } a \text{ before } b) = \psi(x) = \frac{(q/p)^x - (q/p)^b}{(q/p)^a - (q/p)^b} \qquad (a < x < b), \tag{11.22}
$$

which was also derived earlier as Proposition 2.1 (also see derivation by an application of the strong Markov property in Exercise 4 of Chapter 7). If one now uses the martingale $S_n' = S_n - n(p-q)$ $(n \ge 0)$, then (11.14) leads to

$$
x + (p-q)\mathbb{E}\tau = \mathbb{E}S_\tau' = a\psi(x) + b(1 - \psi(x)) = b - (b-a)\psi(x),
$$

so that

$$\mathbb{E}_x\tau = \frac{b-x}{p-q} - \frac{(b-a)}{p-q}\psi(x), \qquad a < x < b, \tag{11.23}$$

where $\psi(x)$ is as in (11.20).

Example 4 (*Gambler's Ruin*)**.** The calculations (11.20)–(11.23) may be viewed in the context of the *gambler's ruin* problem, in which a gambler has an initial capital of $S_0 = x$ dollars, $x > 0$. At each play the gambler may win 1 dollar with probability p or lose 1 dollar with probability $q = 1 - p$. The gambler's objective is to increase an initial asset to b dollars ($b > x$) and then quit. The gambler is ruined if all x dollars are lost, i.e., when the asset is reduced to zero before attaining the objective. In this case Z_n is the gambler's gain on the n-th play, with $P(Z_n = +1) = p$ and $P(Z_n = -1) = q$, and S_n represents the gambler's asset, or fortune, after n plays. The relations (11.13) and (11.22) give the probabilities for the gambler to be ruined before reaching the goal of increasing the initial assets to b, when $p = \frac{1}{2}$ and when $p \neq \frac{1}{2}$, respectively. The relations in (11.20) and (11.23) provide the corresponding expected durations of the game.

Example 5 (*Necessity of Technical Conditions for Stopping*)**.** To underscore the role played by the technical conditions (i) and (ii) of Theorem 11.1, consider again a simple symmetric random walk starting at x, and let $a < x < b$. Letting $b \uparrow \infty$ in the expression for $\psi(x)$ in (11.20), one obtains the result that $P(\tau_a < \infty) = 1$, where $\tau_j := \inf\{n \geq 0 : S_n = y\}$. Similarly, letting $a \downarrow -\infty$, one gets $P(\tau_b < \infty) = \lim_{a\downarrow-\infty}(1 - \psi(x)) = 1$. In other words, the simple symmetric random walk is recurrent. But it follows from (11.20) that $\mathbb{E}\tau_b = \infty$. If one formally applies Theorem 11.1 with $\tau_1 = 0$, $\tau_2 = \tau_b$, one arrives at the absurd identity $b = x$. Note that condition (i) of Theorem 11.1 holds, but condition (ii) fails to apply. On the other hand, (11.14) (or (11.4)) does apply to τ_b for the simple random walk with $\frac{1}{2} < p < 1$. In this case $\tau_b < \infty$ a.s. (by the SLLN, or by letting $a \downarrow -\infty$ in $1 - \psi(x)$ from (11.22)). One may check the hypothesis of Corollary11.3 in this case, with $c_n = \mathbb{E}|Z_1| + \mu = 1 + (p - q) = 2p$ for all n, and $\mathbb{E}\tau_b$ is shown to be finite by letting $a \downarrow -\infty$ in (11.20) (using the monotone convergence theorem),

$$\mathbb{E}\tau_b = \frac{b-x}{p-q} \qquad (x < b;\ p > \tfrac{1}{2}). \tag{11.24}$$

Of course, (11.24) is just (11.14) in this case.

The following corollaries of Theorem 11.1 are often referred to as *optional sampling theorems* as well.

Corollary 11.6 *(Optional Sampling).* If $\{X_n\}_{n=1}^\infty$ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-*martingale* or a *submartingale*, and τ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-stopping time, then $\{W_n := X_{\tau\wedge n}\}$ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-martingale or a submartingale, accordingly.

Proof. First note that $W_n := \sum_{m=0}^{n} X_m \mathbf{1}_{[\tau=m]} + X_n \mathbf{1}_{[\tau>n]}$ is $\mathcal{F}_n$-measurable. If $\{X_n\}_{n=1}^{\infty}$ is a submartingale, apply Theorem 11.1 with $\tau_1 = \tau \wedge n$ and $\tau_2 = \tau \wedge (n+1)$ to get

$$\mathbb{E}\left(W_{n+1}|\mathcal{F}_n\right) = \mathbb{E}\left[\mathbb{E}(W_{n+1}|\mathcal{F}_{\tau\wedge n})|\mathcal{F}_n\right] \geq \mathbb{E}(W_n|\mathcal{F}_n) = W_n, \tag{11.25}$$

with equality in the case $\{X_n\}_{n=1}^{\infty}$ is a martingale. Note that for the second equality, we have used the fact $\mathcal{F}_{\tau\wedge n} \subset \mathcal{F}_n$ (see (11.3)). ■

***Corollary 11.7** (Optional Sampling).* Let $\{\tau_n\}_{n=1}^{\infty}$ be an increasing sequence of $\mathcal{F}_n$-stopping times and $\{X_n\}_{n=1}^{\infty}$ a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-(sub)martingale. If, for each n, $\tau_n < \infty$ a.s., $\mathbb{E}|X_{\tau_n}| < \infty$, and $\mathbb{E}|X_m \mathbf{1}_{[\tau_n > m]}| \to 0$ as $m \to \infty$, then $\{W_n := X_{\tau_n}\}$ is a $\{\mathcal{F}_{\tau_n}\}$-(sub)martingale.

Proof. Take $\tau_1 = \tau_n$ and $\tau_2 = \tau_{n+1}$ in Theorem 11.1. ■

We close this chapter with an application from actuarial mathematics.

Example (*Risk, Ruin, and Reinsurance*)**.** Insurance risk is a historically important area for the development of probabilistic theory and methods. The compound Poisson process provides a basic model in insurance where premiums are collected at a constant rate c and in the course of time i.i.d. (strictly positive) claim amounts $X_1, X_2, \ldots$ occur independently of the homogeneous Poisson process $N = \{N(t) : t \geq 0\}$ with intensity parameter $\lambda > 0$ that counts the arrivals of claims. In particular, the inter-arrival times of claims are *exponentially distributed* with mean $\lambda^{-1} > 0$ in this model. For a company with an initial risk reserve amount $U(0) = u$, their capital at time t is then given by

$$U(t) = u + G(t) := u + ct - \sum_{j=0}^{N(t)} X_j, \quad t \geq 0. \tag{11.26}$$

This model is referred to as the *Cramér–Lundberg model* in insurance.

The *Sparre–Andersen* model is a more general model in which one assumes successive times $T_1, T_2, \ldots$ of claim arrival times such that the inter-arrival times $A_j = T_j - T_{j-1}$, $j \geq 1$, are i.i.d., *not* necessarily exponentially distributed. In addition, one assumes $\mathbb{E}A_1 = \lambda^{-1} < \infty$ ($T_0 = 0$).

Remark 11.1. Applications to hydrology involve *dual models* for reservoir storage by a change of signs for the claims and premium rate, in which case c is the withdrawal rate and $X_1, X_2, \ldots$ are random inputs from rainfall/runoff events, assumed to occur according to i.i.d. inter-arrival times, independently of the input amounts. Similarly for charities, the $X_1, X_2, \ldots$ are donations, and c is a spending rate.

The *probability of ruin* for the general Sparre–Andersen model is defined by

$$\psi(u) = P(\sum_{j=1}^{n} X_j > u + c\sum_{j=1}^{n} A_j \text{ for some n}) = P(\sum_{j=1}^{n} Z_j > u \text{ for some n}), \tag{11.27}$$

where

$$Z_j = X_j - cA_j, \quad j = 1, 2, \dots. \tag{11.28}$$

The common distribution of the i.i.d. $Z_1, Z_2, \dots$ is assumed to satisfy the *Net Profit Condition* (NPC):

$$(NPC) \qquad \mathbb{E}Z_1 \equiv \mathbb{E}X_1 - cA_1 < 0. \tag{11.29}$$

Note that if $\mathbb{E}Z_1$ is finite and $\mathbb{E}Z_1 \geq 0$, then by the strong law of large numbers $\psi(u) = 1$ for all u. To avoid the trivial case $\psi(u) = 0$ for all $u > 0$, one also assumes

$$P(Z_1 > 0) > 0. \tag{11.30}$$

The ruin probability provides a risk measure amenable to analysis for the Sparre–Andersen model via martingale optional stopping theory in the case of an infinite time horizon $T = \infty$, assuming the so-called *light-tailed* claim size distribution, that is, assuming

$$\mathbb{E}e^{qX_1} < \infty \quad \text{for some } q > 0. \tag{11.31}$$

Under this light-tailed assumption, there exists $0 < h \leq \infty$ such that

$$0 < m(q) \equiv \mathbb{E}e^{qZ_i} < \infty \quad 0 \leq q < h, \quad \lim_{q \downarrow h} m(q) = \infty. \tag{11.32}$$

Proposition 11.8 **(*Classical Lundberg Inequality for the Sparre–Andersen Model*).** Under the above net profit condition (11.29), non-degeneracy (11.30), and light-tailed claim assumption (11.31) for the Sparre–Andersen model, there is a unique positive solution to $m(q) = 1$, denoted $q = R > 0$. Moreover,

$$\psi(u) \leq \exp\{-Ru\}, \qquad \text{for all } u > 0. \tag{11.33}$$

Proof. Note that $m(0) = 1$, $m'(0) = \mathbb{E}Z_1 < 0$ (or, $m'(0^+) < 0$ if $m(q) = \infty$ for all $q < 0$), $m''(q) = \mathbb{E}Z_1^2 \exp\{qZ_1\} > 0$ for all $q > 0$, and $m(q) \to \infty$ as $q \uparrow h$. Hence $m(q)$ decreases from $m(0) = 1$ to a minimum at $\tilde{q} \in (0, \infty)$, before increasing strictly to infinity as $q \uparrow h$. It follows that there exists a unique $q = R > 0$ such that $m(R) = 1$. To prove (11.33), let $\tau = \inf\{n \geq 1 : S_n > u\}$, where $S_n = Z_1 + \cdots + Z_n$ $(n = 1, 2, \dots)$, $S_0 = 0$. Then τ is a stopping time with respect to the filtration $\mathcal{F}_n = \sigma\{Z_k : k = 1, \dots, n\}$ for $n \geq 1$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$, and

$$\psi(u) = P(\tau < \infty). \tag{11.34}$$

Next write $W_n = u - S_n (n \geq 1)$, $W_0 = u$. Then $U_n = \{\exp(-RW_n) : n \geq 0\}$ is a $\mathcal{F}_n$-martingale. Since $m(R) = 1$, one has

$$\begin{aligned}\mathbb{E}(U_{n+1}|\mathcal{F}_n) &= \mathbb{E}(U_n \exp\{RZ_{n+1}\}|\mathcal{F}_n) \\ &= U_n\mathbb{E}(\exp\{RZ_{n+1}\}) = U_n m(R) = U_n.\end{aligned} \tag{11.35}$$

By the optional sampling theorem, one then has

$$\begin{aligned}\exp\{-Ru\} &= \mathbb{E}U_0 = \mathbb{E}U_{\tau\wedge n} \\ &\geq \mathbb{E}(U_{\tau\wedge n}\mathbf{1}_{[\tau\leq n]}) \\ &= \mathbb{E}(U_\tau\mathbf{1}_{[\tau\leq n]}) \text{ for all } n.\end{aligned} \tag{11.36}$$

Noting that $U_\tau > 1$ on the event $[\tau < \infty]$, it follows from (11.36) that $\exp\{-Ru\} \geq \mathbb{E}\mathbf{1}_{[\tau\leq n]} = P(\tau \leq n)$ for all n. Letting $n \uparrow \infty$, one arrives at (11.33). ∎

Remark 11.2. It will be shown in Chapter 26 that the exponential rate provided by (11.33) cannot in general be improved upon. The true asymptotic rate will be shown to be given by $\psi(u) \sim d\exp\{-Ru\}$ as $u \to \infty$, for some constant $d \leq 1$. Here $\sim$ means that the ratio of the two sides converges to one as $u \to \infty$.

The parameter R is referred to as an *adjustment coefficient*. The larger R the smaller the bound on the risk of ruin for a given initial reserve u. Its use in mitigating risk can be given through considerations of *reinsurance* as follows. The purchase of reinsurance will reduce the company's profit, and however it can increase the company's security as measured by the risk of ruin. Suppose for simplicity that in the absence of reinsurance, a company has a *risk reserve process* given by the classic Cramér–Lundberg model above:

$$U(t) = u + G(t) = u + ct - \sum_{k=0}^{N_t} X_k, \quad t \geq 0.$$

A *reinsurance policy* is defined by a function ρ, which pays $\rho(x)$ on a claim amount x subject to the *feasibility condition* $0 \leq \rho(x) \leq x$, and $\rho(x) < x$ if $x > 0$. For example, the familiar *excess of loss policy with deductible b* is given by $\rho_b(x) = (x-b)^+ = \max(x-b, 0)$, $x \geq 0$. On the other hand, the *proportionate loss policy at rate* $\gamma \in (0, 1]$ is $\rho^{(\gamma)}(x) = \gamma x$, $x \geq 0$. The *reinsurance relative security loading factor* $\xi = \xi(\rho, c_\rho)$ is defined by

$$1 + \xi(\rho, c_\rho) = \frac{c_\rho}{\mathbb{E}\rho(X)},$$

where X is distributed as a claim size random variable, and $c_\rho > 0$ denotes the reinsurance premium rate. The *adjusted risk reserve* under a reinsurance policy ρ is given by

$$\tilde{U}(t) = u + \tilde{G}(t), \quad t \geq 0,$$

where

$$\tilde{G}(t) = (c - c_\rho)t - \sum_{k=0}^{N_t}(X_k - \rho(X_k)), \quad t \geq 0. \tag{11.37}$$

The following result provides a sense in which excess of loss policies are optimal from the point of view of risk reduction among all feasible reinsurance policies for the same premiums and relative security loads.

Theorem 11.9 *(Optimality of Excess of Loss Reinsurance).* Assume the Cramér–Lundberg model (11.26) for insurance capital. Let R_b denote the adjustment coefficient on the risk reserve under the excess of loss policy ρ_b with deductible b. Then for any other reinsurance policy ρ with risk reserve adjustment coefficient $\tilde{R}$ for equivalent premiums $c_\rho = c_{\rho_b}$, and the same relative security loads $\xi(\rho, c_\rho) = \xi(\rho_b, c_{\rho_b})$, one has $\tilde{R} \leq R_b$.

Proof. The risk reserve process for any policy ρ subject to the conditions of the theorem has adjustment coefficient given by the positive solution $q = \tilde{R}$ to

$$\lambda + (c - c_\rho)q = \lambda \mathbb{E} e^{q(X - \rho(X))}. \tag{11.38}$$

To see this note that under the risk reserve process, the effective premium paid by the insurance company is $c - c_\rho$, and the effective size is $x - \rho(x)$. Hence the function $m(q)$ in Proposition 11.8 becomes $\mathbb{E}\exp\{q(X - \rho(X))\} - (c - c_\rho)A\} = \frac{\lambda}{\lambda + (c - c_\rho)q}\mathbb{E}e^{-q(x - \rho(x))}$. Setting this equal to one yields (11.38).

Since the affine linear map $q \to \lambda + (c - c_\rho)q, q \geq 0$, on the left side of this equation is invariant under ρ under the conditions of the theorem, by monotonicity of the functions $q \to \lambda\mathbb{E}e^{q(X-\rho(X))}, q \geq 0$, it is sufficient to show $\mathbb{E}e^{q(X-\rho(X))} \geq \mathbb{E}e^{q(X-\rho_b(X))}$ for $q > 0$. By convexity of the exponential function, one has the line of support inequality $e^{rq} \geq e^{rq'} + re^{rq'}(q - q')$ for $r \geq 0$ and non-negative q', q. Thus using feasibility, one has $r := x - \rho(x) \geq 0$, and taking $q' := q\frac{x - \rho_b(x)}{x - \rho(x)}$ for fixed $x \geq 0$, one has for $q \geq 0$,

$$\exp\{q(x - \rho(x))\} \geq \exp\{q(x - \rho_b(x))\} + q\exp\{q(x - \rho_b(x))\}(\rho_b(x) - \rho(x)), \tag{11.39}$$

if $x - \rho(x) > 0$. If $x - \rho(x) = 0$, then (11.39) is trivial. Observe that by definition of the excess of loss policy ρ_b and non-negativity of ρ, one has

$$x - \rho_b(x) \leq b, \text{ and } x - \rho_b(x) = b \text{ if } \rho_b(x) > \rho(x).$$

Thus

$$\exp\{q(x-\rho(x))\} \geq \exp\{q(x-\rho_b(x))\} + qe^{qb}(\rho_b(x)-\rho(x)), \quad x \geq 0.$$

Apply this sample-pointwise to the random variable (function) X, taking expectations and noting that $\mathbb{E}\rho(X) = \mathbb{E}\rho_b(X)$ by the relative security loading and premium constraints. This completes the proof that $\mathbb{E}e^{q(X-\rho(X))} \geq \mathbb{E}e^{q(X-\rho_b(X))}$ for $q \geq 0$, as desired. ■

Remark 11.3. An alternative approach to mitigating[1] risk of ruin involves investment of the surplus, i.e., premiums collected minus claims, into a portfolio of a non-risky bond and a risky investment in stocks as described in Chapter 23. Also, other efforts in actuarial/financial mathematics to standardize the concept of "Value at Risk" among international institutions worth mention involve definitions based on quantiles in the profit/loss distribution (see Exercise 9). If one considers $U(t), 0 \leq t \leq T$ over a time horizon of length T (possibly infinite), then the risk reserve $u = u(r)$ at a given level of insolvency risk, say r, may be viewed in terms of a quantile of the distribution of

$$V_T = \min_{0\leq t\leq T}(c\sum_{j=1}^{N(t)} A_j - \sum_{j=0}^{N(t)} X_j)$$

via the condition that

$$P(V_T < -u) = r.$$

More specifically, the $(1-r)$-th quantile of the distribution of V_T, namely $-u = -u(r)$, defined by $P(V_T \geq -u) = 1-r$, is a form of the Value at Risk over the horizon T, often denoted $VaR_r(T)$.

Exercises

1. Let $\{X_n : n \in \mathbb{Z}_+\}$ be a sequence of measurable functions on $(\Omega, \mathcal{F})$ into $(S, \mathcal{S})$.

   (a) Let $\mathcal{G} = \sigma\{X_0, X_1, X_2, \dots\}$. If $\mathcal{G}_n = \sigma\{X_0, X_1, \cdots, X_n\}$ $(n \geq 0)$, and τ is a $\{\mathcal{G}_n\}_{n=0}^{\infty}$-stopping time, show that $\mathcal{G}_\tau \equiv \{A \in \mathcal{G} : A \cap [\tau = m] \in \mathcal{G}_m, m = 0, 1, \dots\} = \sigma\{X_{\tau\wedge n} : n \in \mathbb{Z}_+\}$.

[1] For example, see Albrecher et al. (2012) and references therein.

(b) Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a filtration such that $\{X_n\}_{n=0}^{\infty}$ is adapted to it, and τ is a $\{\mathcal{F}_n\}_{n=0}^{\infty}$-stopping time, then show that $\mathcal{F}_\tau \supset \sigma\{X_{\tau\wedge n} : n \in \mathbb{Z}_+\}$. Give an example to show that the containment "$\supset$" may be *proper*, i.e., strict.

2. (a) If $\tau_1 \leq \tau_2$ are two $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times, and $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-adapted sequence of functions with values in a measurable space $(S, \mathcal{S})$, show that $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.
   (b) Show that, with the above notation, $X_\tau \mathbf{1}_{[\tau<\infty]}$ is $\mathcal{F}_\tau$-measurable for every $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time τ.
3. Let τ_1 and τ_2 be $\{\mathcal{F}_n\}$-stopping times. Show that (i) $\tau_1 \wedge \tau_2$ and $\tau_1 \vee \tau_2$ are $\{\mathcal{F}_n\}$-stopping times. (ii) Show $\mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2} \subset \mathcal{F}_{\tau_1\wedge\tau_2}$. [*Hint*: If $G \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$, then $(G \cap [\tau_1 \leq n]) \cup (G \cap [\tau_2 \leq n]) = G \cap [\tau_1 \wedge \tau_2 \leq n] \in \mathcal{F}_n$.] (iii) Use (ii) to show that $G \cap [\tau_i \leq m] \in \mathcal{F}_{\tau_i} (i = 1, 2)$.
4. Give the details for the calculation of the expected time to reach the boundary (11.21).
5. (*Martingale Strategy; Double or Nothing*) Suppose a gambler bets 1 dollar in the first play and quits if a win (one dollar). If a loss then the bet is doubled to 2 dollars, with the same strategy of quit on a win and double bet to 4 dollars on a loss, etc. Thus the gambler quits after the first win, and bets 2^{n-1} dollars on the n-th game if the first $n - 1$ plays are losses ($n = 1, 2, \ldots$). Assume that the outcomes of the plays are independent events, and the probability of winning (for the gambler) is p in each play.

   (a) Show that with probability one, the gambler comes away with an overall gain of 1 dollar, unless $p = 0$.
   (b) Compute the expected duration of the game.
   (c) Let $\{Z_n : n \geq 1\}$ be an independent sequence with $P(Z_n = 2^{n-1}) = p$ and $P(Z_n = -2^{n-1}) = q \equiv 1 - p$ $(0 < p < 1)$. Consider the (inhomogeneous) random walk $S_n = Z_1+Z_2+\cdots+Z_n$ $(n \geq 1)$, $\mathcal{F}_n = \sigma\{Z_1, \ldots, Z_n\}$ $(n \geq 1)$, and let $\tau = \inf\{n \geq 1 : Z_n > 0\}$. Consider the martingale $X_n = S_n - \mathbb{E}S_n$ $(n \geq 1)$ and τ as defined here. Calculate the distribution of τ.
   (d) In (c), calculate $\mathbb{E}|X_m|\mathbf{1}_{[\tau>m]}$ and show that it goes to zero as $m \to \infty$ if and only if $p > \frac{1}{2}$.

6. (*Optimal Strategy for Gambler*) In Exercise 5 the gambler must have unlimited assets to cover losses as they may accrue. A more realistic assumption is that the gambler has an initial capital of x dollars ($x > 0$), and at no stage of the game she can bet more than her current asset or capital. Suppose the gambler's objective is to increase her initial capital to b $(> x)$ dollars. Suppose $0 < p \leq \frac{1}{2}$. A *feasible strategy* for the gambler is the specification of the amount Y_n, $1 \leq Y_n \leq b - S_{n-1}$ on the event $[S_{n-1} > 0]$, as her bet (using past information if necessary) for the n-th play. Then her gain Z_n, say, on the n-th play is $\pm Y_n$, so that $\mathbb{E}(Z_n||\mathcal{F}_{n-1}) = (p - q)Y_n \leq 0$ (for the standard filtration with $\mathcal{F}_0 = \{\Omega, \emptyset\}$).

(a) Show that regardless of the (feasible) strategy $S_n := x + Z_1 + \cdots + Z_n$ $(n \geq 0)$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-supermartingale, and $\tau := \inf\{n \geq 0 : S_n \in \{0, b\}\}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time.

(b) Show that if

$$r(x) \geq pr(x+y) + qr(x-y), \forall x \geq 1,$$

where $r(x) = P_x(\tau_b < \tau_0)$, then $r(S_n) = P_{S_n}(\tau_b < \tau_0), n = 0, 1, 2, \ldots,$ is a supermartingale

(c) In a *bold strategy*,[2] she does not bet any more than she needs to reach b dollars, and she quits the first time she accumulates b dollars or she loses all her capital. In the case $p = \frac{1}{2}$, prove that the probability that the gambler reaches her objective is *no more than* x/b.

(d) Assume $p = 1/2$. Show that an optimal strategy (for maximizing this probability) is the *timid strategy* of betting 1 dollar in each play until the stopping time τ is reached. [*Hint*: Note that under the timid strategy, the capital evolves according to the simple symmetric random walk on $[0, b]$ with absorbing barriers at 0, b, for which $P(\tau_b^{(1)} < \tau_0^{(1)}) = x/b$ satisfies the supermartingale inequality above (as an equality). Modify an arbitrary feasible strategy by adapting the timid strategy from the n-th step forward, $n = 1, 2, \ldots$. Use the supermartingale property for the timid strategy to show $\mathbb{E}_x P_{S_n^{(2)}}(\tau_b^{(n)} < \tau_0^{(n)}) \leq P_x(\tau_b^{(1)} < \tau_0^{(1)})$. Let $n \to \infty$.]

7. Let $\{X_n : n \geq 1\}$ be a *non-negative* $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale and τ an a.s. finite $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time. Assume $\mathbb{E}X_m \mathbf{1}_{[\tau > m]} \to 0$ as $m \to \infty$. Give a direct proof (without using Theorem 11.1) that $\mathbb{E}X_\tau \geq \mathbb{E}X_1$, with equality for the martingale case. [*Hint*: $\mathbb{E}X_\tau = \lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{E}[X_n(\mathbf{1}_{[\tau > n-1]} - \mathbf{1}_{[\tau > n]})] \geq \lim_{m\to\infty}[\mathbb{E}X_1 - \mathbb{E}X_m \mathbf{1}_{[\tau > m]}] = \mathbb{E}X_1$, with "=" in place of "≥" for martingales.]
8. Show that Theorem 11.1 holds for supermartingales with the inequalities in (11.4) reversed.
9. Consider the Sparre–Andersen model under the net profit condition and light-tailed claim assumptions. Give an upper bound on the value at risk $VaR_r(\infty)$ over an infinite time horizon $T = \infty$ in terms of the risk level r and the adjustment coefficient R.

[2] For a detailed analysis of the bold strategy, see the interesting article Billingsley (1983).

Chapter 12
The Upcrossings Inequality and (Sub)Martingale Convergence

Doob's intriguing and delicate upcrossing inequality is derived in this chapter. One of its consequences is the (sub) martingale convergence theorem, which in turn leads to a proof of the strong law of large numbers and a derivation of DeFinetti's representation of exchangeable (symmetrically dependent) sequences of random variables. Other applications include regularity of sample paths of continuous parameter stochastic processes to be derived in Chapter 13.

Recall that submartingales $\{Y_n\}_{n=1}^{\infty}$ have monotonically non-decreasing expected values. Thus, trivially, if the numerical sequence $\{\mathbb{E}|Y_n|\}_{n=1}^{\infty}$ is bounded, then $\lim_{n\to\infty} \mathbb{E}Y_n$ will exist. The remarkable fact to be proven in this chapter is that submartingales having bounded absolute means (or slightly less), and such "bounded" martingales and non-negative supermartingales, will a.s. converge to a finite limit!

Consider a sequence $\{Y_n : n \geq 1\}$ of real-valued random variables such that Y_n is $\mathcal{F}_n$-measurable for a given filtration $\{\mathcal{F}_n\}_{n=1}^{\infty}$. Let $a < b$ be an arbitrary pair of real numbers. An *upcrossing* of the interval (a, b) by $\{Y_n\}_{n=1}^{\infty}$ is a passage to a value $\geq b$ from an earlier value $\leq a$, while a *downcrossing* of (a, b) is a passage to a value $\leq a$ from an earlier value $\geq b$. It is convenient to look at the corresponding process $\{X_n := (Y_n - a)^+\}_{n=1}^{\infty}$, where $(Y_n - a)^+ = \max\{(Y_n - a), 0\}$. The upcrossings of (a, b) by $\{Y_n\}_{n=1}^{\infty}$ are the upcrossings of $(0, b - a)$ by $\{X_n\}_{n=1}^{\infty}$. The successive upcrossing times η_{2k} $(k = 0, 1, \dots)$ of $\{X_n\}_{n=1}^{\infty}$ are defined by

$$\eta_0 := 1,$$

$$\eta_1 := \inf\{n \geq 1 : X_n = 0\},$$

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_12

$$\eta_2 := \inf\{n > \eta_1 : X_n \geq b - a\},$$
$$\eta_{2k+1} := \inf\{n > \eta_{2k} : X_n = 0\},$$
$$\eta_{2k+2} := \inf\{n > \eta_{2k+1} : X_n \geq b - a\},\ k \geq 0. \tag{12.1}$$

Then each η_k is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping time. Fix a positive integer N and define

$$\tau_k := \eta_k \wedge N \equiv \min\{\eta_k, N\}, \quad (k = 0, 1, \ldots). \tag{12.2}$$

Then each τ_k is also a stopping time. Also, $\tau_k \equiv N$ for $k \geq N$, so that $X_{\tau_k} = X_N$ for $k \geq N$.

Let $U_N \equiv U_N(a, b)$ denote the number of upcrossings of (a, b) by $\{Y_n\}_{n=1}^{\infty}$ by time N, i.e.,

$$U_N(a, b) := \sup\{k \geq 0 : \eta_{2k} \leq N\}, \tag{12.3}$$

with the convention that the supremum over an empty set of 0. Thus U_N is also the number of upcrossings of $(0, b - a)$ by $\{X_n\}_{n=1}^{\infty}$ in time N.

Since $X_{\tau_k} = X_N$ for $k \geq N$, one may write

$$X_N - X_1 = \sum_{k=1}^{[N/2]+1} \left(X_{\tau_{2k-1}} - X_{\tau_{2k-2}}\right) + \sum_{k=1}^{[N/2]+1} \left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right). \tag{12.4}$$

To relate (12.4) to the number U_N, let ν denote the largest k such that $\eta_k \leq N$, i.e., ν is the last time $\leq N$ for an upcrossing or a downcrossing. Notice that $U_N = [\nu/2]$. If ν is *even*, then $X_{\tau_{2k}} - X_{\tau_{2k-1}} \geq b - a$ if $2k \leq \nu$, and is $X_N - X_N = 0$ if $2k > \nu$. Now suppose ν is *odd*. Then $X_{\tau_{2k}} - X_{\tau_{2k-1}} \geq b - a$ if $2k - 1 < \nu$, and is 0 if $2k - 1 > \nu$, and is $X_{\tau_{2k}} - 0 \geq 0$ if $2k - 1 = \nu$. Hence in every case

$$\sum_{k=1}^{[N/2]+1} \left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right) \geq [\nu/2](b - a) = U_N(b - a). \tag{12.5}$$

As a consequence,

$$X_N - X_1 \geq \sum_{k=1}^{[N/2]} \left(X_{\tau_{2k-1}} - X_{\tau_{2k-2}}\right) + (b - a)U_N. \tag{12.6}$$

Observe that (12.6) is true for an arbitrary sequence of random variables (or, real numbers) $\{Y_n\}_{n=1}^{\infty}$. Assume now that $\{Y_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale. Then $\{X_n\}_{n=1}^{\infty}$ is a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale by Proposition 10.1(b). By optional sampling (Corollary 11.6), $\{X_{\tau_k} : k \geq 1\}$ is a submartingale. Hence $\mathbb{E}X_{\tau_k}$ is non-decreasing in k, so that

$$\mathbb{E}\left(\sum_{k=1}^{[N/2]}\left(X_{\tau_{2k-1}} - X_{\tau_{2k-2}}\right)\right) \geq 0. \tag{12.7}$$

Applying this to (12.6) one obtains the following truly remarkable bound on the mean number of upcrossings.

Theorem 12.1 *(Upcrossing Inequality).* Let $\{Y_n\}_{n=1}^{\infty}$ be a $\{\mathcal{F}_n\}_{n=1}^{\infty}$-submartingale. For each pair $a < b$ the expected number of upcrossings of (a, b) by $Y_1, \ldots, Y_N$ satisfies the inequality

$$\mathbb{E}U_N(a,b) \leq \frac{\mathbb{E}(Y_N - a)^+ - \mathbb{E}(Y_1 - a)^+}{b-a} \leq \frac{\mathbb{E}Y_N^+ + |a|}{b-a} \leq \frac{\mathbb{E}|Y_N| + |a|}{b-a}. \tag{12.8}$$

As an important consequence of this result we get

Theorem 12.2 *(Submartingale Convergence Theorem).* Let $\{Y_n\}_{n=1}^{\infty}$ be a submartingale such that $\mathbb{E}(Y_n^+)$ is a bounded sequence. Then $\{Y_n\}_{n=1}^{\infty}$ converges a.s. to a limit Y_∞. If $K := \sup_n \mathbb{E}|Y_n| < \infty$, then Y_∞ is a.s. finite and $\mathbb{E}|Y_\infty| \leq K$.

Proof. Let $U(a, b)$ denote the total number of upcrossings of (a, b) by $\{Y_n\}_{n=1}^{\infty}$. Then $0 \leq U_N(a, b) \uparrow U(a, b)$ as $N \uparrow \infty$. Therefore, by the monotone convergence theorem

$$\mathbb{E}U(a,b) = \lim_{N\uparrow\infty} \mathbb{E}U_N(a,b) \leq \sup_N \frac{\mathbb{E}Y_N^+ + |a|}{b-a} < \infty. \tag{12.9}$$

In particular $U(a, b) < \infty$ almost surely, so that

$$P\left(\liminf Y_n < a < b < \limsup Y_n\right) = 0. \tag{12.10}$$

Since this holds for every pair $a, b = a + \frac{1}{m}$ with rational number a, and a positive integer m, and the set of all such pairs is countable, one must have $\liminf Y_n = \limsup Y_n$ almost surely. Let Y_∞ denote the a.s. limit. By Fatou's Lemma, $\mathbb{E}|Y_\infty| \leq \lim \mathbb{E}|Y_n|$. ■

Theorem 12.2 and Doob's L^p-maximal inequalities imply the following powerful result.

Theorem 12.3 *(L^p-Convergence of Submartingales).* Let $\{X_n : n \geq 1\}$ be a $\{\mathcal{F}_n\}$-submartingale. (a) If $\sup_n \mathbb{E}|X_n| \log^+ |X_n| < \infty$, then X_n converges a.s. and in L^1 to a random variable X. (b) If $\sup_n \mathbb{E}|X_n|^p < \infty$ for some $p > 1$, then X_n converges a.s. and in L^p to a random variable X.

Proof. In both (a) and (b), almost sure convergence follows from Theorem 12.2. Doob's maximal inequalities (Theorem 10.8) imply that in (a) $\{X_n\}$ is uniformly integrable and in (b) $|X_n|^p$ is uniformly integrable (Exercise 8). ■

An immediate consequence of Theorem 12.2 is

Corollary 12.4. (a) A non-negative martingale $\{Y_n\}_{n=1}^{\infty}$ converges almost surely to a finite limit Y_∞. Also, $\mathbb{E}Y_\infty \le \mathbb{E}Y_1$. (b) A non-negative supermartingale $\{Y_n\}_{n=1}^{\infty}$ converges almost surely to a finite limit Y_∞.

Proof. (a) For a non-negative martingale $\{Y_n\}_{n=1}^{\infty}$, $\sup_n \mathbb{E}|Y_n| = \sup_n \mathbb{E}Y_n = \mathbb{E}Y_1$. So the assertion follows from Theorem 12.2. (b) For a non-negative supermartingale $\{Y_n\}_{n=1}^{\infty}$, $\{-Y_n\}_{n=1}^{\infty}$ is a submartingale bounded above by zero, i.e., $(-Y_n)^+ = 0$ and therefore, $\sup_n \mathbb{E}(-Y_n)^+ = 0$. Hence, again by Theorem 12.2, $\lim_n Y_n = -\lim_n(-Y_n) \ge 0$ exists and is almost surely finite. ■

It follows from the Corollary that the martingales $\{Y_n := \prod_{j=1}^{n} X_j\}$ converge almost surely to an integrable random variable Y_∞, if $\{X_n\}_{n=1}^{\infty}$ is an independent non-negative sequence with $\mathbb{E}X_n = 1$ for all n. In the case $\{X_n\}_{n=1}^{\infty}$ is *i.i.d.* and $P(X_1 = 1) < 1$, it is an interesting fact that the limit of $\{Y_n\}_{n=1}^{\infty}$ is 0 a.s., as shown by the following corollary.

Corollary 12.5. Let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of non-negative random variables with $\mathbb{E}X_1 = 1$. Then $\{Y_n := \prod_{j=1}^{n} X_j\}$ converges almost surely to 0, provided $P(X_1 = 1) < 1$.

Proof. First assume $P(X_1 = 0) > 0$. Then $P(X_n = 0 \text{ for some } n) = 1 - P(X_n > 0 \text{ for all } n) = 1$, since $P(X_j > 0 \text{ for } 1 \le j \le n) = (P(X_1 > 0))^n$. But if $X_m = 0$, then $Y_n = 0$ for all $n \ge m$. Therefore, $P(Y_n = 0 \text{ for all sufficiently large } n) = 1$.

Assume now $P(X_1 > 0) = 1$. Consider the *i.i.d.* sequence $\{\log X_n\}_{n=1}^{\infty}$. Since $x \to \log x$ is concave one has, by Jensen's inequality, $\mathbb{E}\log X_1 \le \log \mathbb{E}X_1 = 0$. Since $P(X_1 = 1) < 1$, $\log X_1$ is not degenerate (i.e., not almost surely a constant). Hence the Jensen inequality is *strict*. Therefore, $\mathbb{E}\log X_n < 0$. By the strong law of large numbers,

$$\frac{1}{n}\log Y_n = \frac{1}{n}\sum_{j=1}^{n}\log X_j \xrightarrow{\text{a.s.}} \mathbb{E}\log X_1 < 0. \tag{12.11}$$

Therefore, $\log Y_n \to -\infty$, a.s., and $Y_n \to 0$ a.s. ■

Another application of Corollary 12.4 to a multiplicative process is the following.

Example 1 ***(Bienaymé–Galton–Watson Branching Process).*** Consider a simple branching process in which the total number X_n of individuals in the n-th generation evolves as follows. Each individual is replaced by a random number of offspring in the next generation. The sequence of random numbers $L_i^{(n)}, i, n \ge 1$, of offspring produced in this manner throughout the generations are i.i.d. with common probability mass function (*pmf*) f. Assume $0 < \mu := \sum_{k=0}^{\infty} kf(k) < \infty$. Let $\mathcal{F}_n := \sigma\{X_1, \dots, X_n\}$. Then $\{X_n\}_{n=1}^{\infty}$ is a Markov chain and

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} \mid \sigma(X_n)) = \mu X_n. \tag{12.12}$$

Define

$$Y_n := \mu^{-n} X_n. \tag{12.13}$$

Then

$$\mathbb{E}\,(Y_{n+1} \mid \mathcal{F}_n) = \mu^{-(n+1)}\mu X_n = \mu^{-n} X_n = Y_n. \tag{12.14}$$

Hence $\{Y_n\}_{n=1}^{\infty}$ is a non-negative martingale, which converges a.s. to a finite integrable limit Y_∞, $\mathbb{E}Y_\infty \leq \mathbb{E}Y_1 = 1$. One may then express the size X_n of the n-th generation as

$$X_n = \mu^n Y_n = \mu^n (Y_\infty + o(1)), \tag{12.15}$$

where $o(1) \to 0$ a.s. as $n \to \infty$. Suppose $\mu < 1$. Then (12.15) implies that $X_n \to 0$ a.s. (and exponentially fast). Since X_n is integer valued, this means that, with probability one, X_n is 0 for all sufficiently large n, so that $Y_\infty = 0$ a.s. That is, in case $\mu < 1$, *extinction* is certain (and the population size declines exponentially fast). As already shown, extinction is certain in the critical case $\mu = 1$. As a consequence of this and (12.15), one may deduce that $Y_\infty = 0$ (a.s.) for the case $\mu \leq 1$. To complete this picture consider the following natural property associated with the evolution of Bienaymé–Galton–Watson branching processes.

Definition 12.1. A property of Bienaymé–Galton–Watson trees is said to be inherited if every finite tree has the property, and all subtrees rooted at descendants of the root of a tree having this property also have this property.

Importantly, extinction is an inherited property.

Theorem 12.6 ***(Zero-One Law for Inherited Properties).*** The event that a Bienaymé–Galton–Watson process with non-degenerate offspring distribution has a particular inherited property has conditional probability zero or one given nonextinction.

Proof. Let G denote the event that a Bienaymé–Galton–Watson tree with k offspring of a single progenitor root possesses a given inherited property, and let $G_1, \ldots, G_k$ be the events that the respective descendent subtrees of the root have this property. Then, for arbitrary k,

$$P(G|L_1^{(1)} = k) \leq P(G_1 \cap G_2 \cap \cdots \cap G_k|L_1^{(1)} = k) = P(G)^k.$$

In particular, denoting the probability generating function (*pgf*) of the offspring by φ, one has

$$P(G) \leq \mathbb{E}(P(G))^{L_1^{(1)}} = \varphi(P(G)).$$

Let ρ denote the extinction probability. Then, as has been previously shown, ρ is a fixed point of φ. Since every finite tree has the inherited property, i.e., under inheritance, extinction implies G, one has $\varphi(\rho) = \rho \leq P(G)$. Recall that φ is strictly convex, non-decreasing, and has at most two fixed points ρ and 1 in [0, 1]. Moreover, $\varphi(x) \geq x, 0 \leq x \leq \rho$ and $\varphi(x) \leq x, \rho \leq x \leq 1$, with $\varphi(1) = 1$. Thus, $P(G) \in \{\rho, 1\}$, so that if $P(G) = \rho < 1$, then $P(G|\text{nonextinction}) = \frac{P(G)-\rho}{1-\rho} = 0$. On the other hand, if $P(G) = 1$, then $P(G|\text{nonextinction}) = 1 - P(G^c|\text{nonextinction}) = 1 - 0 = 1$. ■

Since $[Y_\infty = 0]$ is an inherited property one immediately arrives at the following.

Corollary 12.7. For any finite, positive mean offspring μ, one has conditionally on nonextinction $Y_\infty = 0$ or $Y_\infty > 0$ almost surely, i.e., $P(Y_\infty = 0)$ is either zero or ρ.

Radon–Nikodym derivatives provide an essential "change of measure" tool of both probability theory and analysis. The next application of martingale convergence provides a natural conceptual interpretation (see Exercise 7).

Corollary 12.8 (Lebesgue Decomposition). Let m, q be finite measures on a measurable space $(S, \mathcal{F})$ and assume that m is non-trivial (i.e., not identically zero) and q is normalized to a probability. Suppose that $\mathcal{F}_n, n \geq 1$ is a filtration such that $m << q$ on each $\mathcal{F}_n$ with Radon–Nikodym derivative $Q_n = dm/dq$ on $\mathcal{F}_n$. Define

$$Q_\infty(x) = \limsup_{n\to\infty} Q_n(x), \quad x \in S.$$

Then

$$m(A) = \int_A Q_\infty(x)q(dx) + m(A \cap [Q_\infty = \infty]), \quad A \in \mathcal{F}.$$

In particular

a. $m << q \iff Q_\infty < \infty \; m - \text{a.e.} \iff \mathbb{E}_q Q_\infty = m(S)$
b. $m \perp q \iff Q_\infty = \infty \; m - \text{a.e.} \iff \mathbb{E}_q Q_\infty = 0.$

Proof. First observe that the sequence $\{Q_n : n \geq 1\}$ defined on the probability space $(S, \mathcal{F}, q)$ is a non-negative martingale with respect to the filtration $\mathcal{F}_n, n \geq 1$ since for any bounded $\mathcal{F}_n$-measurable, and hence $\mathcal{F}_{n+1}$-measurable, function G on S one has

$$\int_S G\mathbb{E}(Q_{n+1}|\mathcal{F}_n)dq = E_q(GQ_{n+1}) = \int_S GQ_{n+1}dq = \int_S GQ_n dq = \int_S Gdm.$$

That is, $\mathbb{E}(Q_{n+1}|\mathcal{F}_n)$ is a version of dm/dq on $\mathcal{F}_n$ and hence agrees with Q_n. Thus it follows from the martingale convergence theorem that $Q_\infty = \lim_n Q_n < \infty$ a.s. with respect to q. Assume without loss of generality that m is a probability; else replace m by its normalization to a probability in what follows. A standard trick to find measures λ_n to dominate (in the sense of absolute continuity) both $m_n := m|_{\mathcal{F}_n}$

and $q_n := q|_{\mathcal{F}_n}$ is to define

$$\lambda_n = \frac{m_n + q_n}{2} \quad \text{and} \quad \lambda = \frac{m+q}{2}.$$

Then $m_n << \lambda_n$, and $q_n << \lambda_n$ on $\mathcal{F}_n$ and λ_n coincides with the restriction of λ to $\mathcal{F}_n$. Now

$$M_n = \frac{dm_n}{d\lambda_n} \quad \text{and} \quad R_n = \frac{dq_n}{d\lambda_n}$$

define non-negative bounded martingales $\{M_n : n \geq 1\}$ and $\{R_n : n \geq 1\}$ on the probability space $(S, \mathcal{F}, \lambda)$, which may be checked along the same lines as above (Exercise 6). Thus the following limits exist a.s. with respect to λ:

$$M_\infty := \lim_{n\to\infty} M_n \quad R_\infty := \lim_n R_n.$$

Moreover one has (Exercise 6)

$$M_\infty = \frac{dm}{d\lambda} \quad R_\infty = \frac{dq}{d\lambda}.$$

Thus $Q_n = M_n/R_n$ and therefore $Q_\infty = M_\infty/R_\infty$ a.s. with respect to λ. Note that $\lambda([M_\infty = R_\infty = 0]) = 0$ since $M_n + R_n = 2\lambda$ for all n. Thus we have arrived at

$$\begin{aligned} m(A) &= \int_A \frac{dm}{d\lambda} d\lambda = \int_A M_\infty d\lambda \\ &= \int_A \frac{M_\infty}{R_\infty} \mathbf{1}_{[R_\infty > 0]} R_\infty d\lambda + \int_A M_\infty \mathbf{1}_{[R_\infty = 0]} d\lambda \\ &= \int_A Q_\infty dq + \int_A \mathbf{1}_{[R_\infty = 0]} dm. \end{aligned}$$

Now notice that since $Q_n R_n = M_n$ for each n, $\mathbf{1}_{[R_\infty=0]} = \mathbf{1}_{[Q_\infty=\infty]}$ a.s. with respect to m, completing the proof of the decomposition. The various equivalent conditions for absolute continuity and/or mutual singularity can be directly read off of this representation (Exercise 6). ■

Corollary 12.9. Let Y be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_n\}_{n=1}^\infty$ an increasing sequence of sub-σ-fields of $\mathcal{F}$. Then $\{Y_n := \mathbb{E}(Y \mid \mathcal{F}_n)\}_{n=1}^\infty$ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-martingale and $\{Y_n\}_{n=1}^\infty$ converges almost surely, and in L^1, to $Y_\infty := \mathbb{E}(Y \mid \mathcal{F}_\infty)$, where $\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n \equiv \sigma\{\cup_{n=1}^\infty \mathcal{F}_n\}$.

Proof. The martingale property is clear. Also, by Jensen's inequality applied to the function $x \to |x|$ one obtains $|Y_n| \equiv |\mathbb{E}(Y \mid \mathcal{F}_n)| \le \mathbb{E}(|Y| \mid \mathcal{F}_n)$, so that

$$\mathbb{E}|Y_n| \le \mathbb{E}|Y| \quad \text{for all } n. \tag{12.16}$$

Therefore, by Theorem 12.2, $\{Y_n\}_{n=1}^\infty$ converges almost surely to an integrable random variable Y_∞, say. In order to show that $\{Y_n\}_{n=1}^\infty$ converges to Y_∞ in L^1, we need to prove that $\{Y_n\}_{n=1}^\infty$ is *uniformly integrable*. This follows from

$$\begin{aligned}\int_{[|Y_n|>\lambda]} |Y_n| dP &= \int_{[|Y_n|>\lambda]} |\mathbb{E}(Y \mid \mathcal{F}_n)| dP \le \int_{[|Y_n|>\lambda]} \mathbb{E}(|Y| \mid \mathcal{F}_n) dP \\ &= \int_{[|Y_n|>\lambda]} |Y| dP.\end{aligned}$$

Now

$$\sup_n P(|Y_n| > \lambda) \le \sup_n \mathbb{E}|Y_n|/\lambda = \mathbb{E}[|\mathbb{E}(Y \mid \mathcal{F}_n)|] \le \mathbb{E}[\mathbb{E}(|Y| \mid \mathcal{F}_n)]/\lambda = \mathbb{E}|Y|/\lambda. \tag{12.17}$$

Therefore, $\int_{[|Y_n|>\lambda]} |Y_n| dP \to 0$ (Exercise). Now, for any given m, $\mathbb{E}Y\mathbf{1}_A = \mathbb{E}Y_n\mathbf{1}_A$ for all $A \in \mathcal{F}_m$ provided $n \ge m$. Letting $n \to \infty$, one gets $\mathbb{E}Y\mathbf{1}_A = \mathbb{E}Y_\infty\mathbf{1}_A$ for all $A \in \mathcal{F}_m$. Since m is arbitrary, the last equality holds for all $A \in \cup_{m=1}^\infty \mathcal{F}_m$. ■

Another useful convergence result related to martingales is

Proposition 12.10 *(Convergence of Reverse Martingales).* Let $\{\mathcal{F}_n : n = 0, 1, 2, \ldots\}$ be a decreasing sequence of σ-fields $\subset \mathcal{F}$, and let $\{Y_n : n = 0, 1, \ldots\}$ be a $\{\mathcal{F}_n\}_{n=1}^\infty$-reverse martingale, i.e., $\mathbb{E}(Y_n \mid \mathcal{F}_{n+1}) = Y_{n+1}$ $(n \ge 0)$.

a. Then $\{Y_n\}_{n=1}^\infty$ converges in L^1 to $\mathbb{E}(Y_0 \mid \mathcal{F}_\infty)$ where $F_\infty = \cap_{n=0}^\infty \mathcal{F}_n$.
b. In particular, given Y_0 integrable, $\mathbb{E}(Y_0 \mid \mathcal{F}_n)$ converges in L^1 to $\mathbb{E}(Y_0 \mid \mathcal{F}_\infty)$.
c. The convergence in (a), (b) is a.s.

Proof.

a. Note that $Y_n = \mathbb{E}(Y_0 \mid \mathcal{F}_n)$ $(n \ge 1)$. This is easy to check directly, or use the fact that $\{Y_n, Y_{n-1}, \ldots, Y_0\}$ is a martingale with respect to the increasing family of σ-fields $\{\mathcal{F}_n, \mathcal{F}_{n-1}, \ldots, \mathcal{F}_0\}$. It follows that $\{Y_n : n \in \mathbb{Z}_+\}$ is uniformly integrable, using (12.17) (with Y_0 in place of Y) and Chebyshev's inequality exactly as in the proof of Corollary 12.9. Hence there exists a subsequence $Y_{n_1}, Y_{n_2}, \cdots$ $(n_1 < n_2 < \cdots)$ which converges in L^1 to some random variable Y_∞, say. Clearly, Y_∞ is $\mathcal{F}_\infty$-measurable. Also, by the definition of conditional expectations,

$$\int_A Y_0 dP = \int_A Y_n dP \qquad \forall\, n \ge 1, \ \forall\, A \in \mathcal{F}_\infty, \tag{12.18}$$

since $A \in \mathcal{F}_n$ for all n. Letting $n \to \infty$ through the subsequence $\{n_j\}$ one gets $\int_A Y_0 dP = \int_A Y_\infty dP$ for every $A \in \mathcal{F}_\infty$, proving that $\mathbb{E}(Y_0 \mid \mathcal{F}_\infty) = Y_\infty$. Since the limit Y_∞ is independent of the subsequence $n_1 < n_2 < \cdots$, the proof is complete.

b. Given Y_0 integrable, $\{\mathbb{E}(Y_0 \mid \mathcal{F}_n) : n \geq 0\}$ (with $\mathcal{F}_0 = \mathcal{F}$, say) is a $\{\mathcal{F}_n\}_{n=1}^\infty$-reverse martingale, so that (a) applies.

c. The proof of a.s. convergence is almost the same as in the case of a submartingale (Theorem 12.2). If $U_N = U_N(a, b)$ is the number of upcrossings of (a, b) by the martingale $\{Y_{-N}, Y_{-N+1}, \dots, Y_1\}$, then $\mathbb{E}U_N(a, b) \leq (\mathbb{E}|Y_1| + |a|)/(b - a)$, by (12.8) (relabeling the random variables Y_{-N} as $Y_1, \dots, Y_1$ as Y_N). The rest of the proof is the same as that of Theorem 12.2.

■

Corollary 12.11 *(Strong Law of Large Numbers—SLLN).* Let $\{Z_n : n \geq 1\}$ be an i.i.d. sequence of integrable random variables. Then $S_n/n \to \mathbb{E}Z_1$ a.s. and in L^1 as $n \to \infty$, where $S_n = Z_1 + \cdots + Z_n$ $(n \geq 1)$.

Proof. Let $\mathcal{F}_n := \sigma\{S_m : m \geq n\}$ $(n = 1, 2, \cdots)$. Observe that the distribution of the sequence $(Z_1, Z_2, \dots, Z_n, S_n, S_{n+1}, \dots)$ is the same as that of the sequence $(Z_{\pi(1)}, Z_{\pi(2)}, \dots, Z_{\pi(n)}, S_n, S_{n+1}, \dots)$ for every permutation $\pi(1), \dots, \pi(n)$ of $1, 2, \dots, n$. In particular,

$$\mathbb{E}(Z_j \mid \mathcal{F}_n) = \mathbb{E}(Z_1 \mid \mathcal{F}_n) \quad (1 \leq j \leq n). \tag{12.19}$$

Averaging (12.19) over $j = 1, \dots, n$, one has

$$\mathbb{E}\left(\frac{S_n}{n} \,\Big|\, \mathcal{F}_n\right) = \mathbb{E}(Z_1 \mid \mathcal{F}_n) \quad (n \geq 1). \tag{12.20}$$

But $\mathbb{E}(S_n/n \mid \mathcal{F}_n) = S_n/n$. Now apply Proposition 12.10 to the sequence $Y_n := \mathbb{E}(Z_1 \mid \mathcal{F}_n)$ $(n \geq 1)$ to get S_n/n converges a.s. and in L^1 to $\mathbb{E}(Z_1 \mid \mathcal{F}_\infty)$ where $\mathcal{F}_\infty = \cap_{n \geq 1}\mathcal{F}_n$. However, since $\liminf S_n/n$ is measurable with respect to the *tail σ-field* of $\{Z_n : n \geq 1\}$, it follows from Kolmogorov's zero-one law that the limit is a constant a.s., namely, $\mathbb{E}Z_1$. ■

To conclude this chapter we consider an application of martingales to arbitrary "symmetrically dependent," or *exchangeable* sequences.

Example (*Exchangeable Sequences of Random Variables and DeFinetti's Theorem*). Recall that a S-valued sequence of random variables $\{X_n : n \geq 1\}$ is *exchangeable* if its distribution is *permutation–invariant*, i.e., for any given $n \in \mathbb{N}$ the distribution of $(X_1, X_2, \dots, X_n)$ is the same as that of $(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)})$ for all permutations $(\pi(1), \dots, \pi(n))$ of $(1, 2, \dots, n)$. One way to construct such sequences is to first pick at random a probability measure Q_θ on $(S, \mathcal{S})$ from a family $\{Q_\theta : \theta \in \Theta\}$, with the random parameter having a distribution $\mu(d\theta)$ on a measurable space $(\Theta, \mathcal{A})$. Given θ, let $X_1, X_2, \dots,$ be an i.i.d. sequence with common distribution Q_θ. The (unconditional) distribution of $\{X_n : n \geq 1\}$ is then permutation invariant. Remarkably, the following theorem asserts that essentially all exchangeable sequences are of this form.

Theorem 12.12 *(DeFinetti's Theorem).* Let $\{X_n : n \geq 1\}$ be exchangeable with values in $(S, \mathcal{S})$. Define $\mathcal{F}_n$ to be the σ-field of events invariant under the permutation of $X_1, \ldots, X_n$, $\mathcal{F}_\infty = \cap_{n=1}^\infty \mathcal{F}_n$.

a. *(SLLN)* if g is a measurable real-valued function on S such that $\mathbb{E}|g(X_1)| < \infty$, then $(1/n)\sum_{j=1}^n g(X_j)$ converges a.s. and in L^1 to $\mathbb{E}(g(X_1) \mid \mathcal{F}_\infty)$.

b. $\{X_n : n \geq 1\}$ is i.i.d., conditionally given $\mathcal{F}_\infty$, i.e., for arbitrary k and bounded measure $g_1, \ldots, g_m$ on S one has, a.s.,

$$\mathbb{E}\left[\prod_{i=1}^m g_i(X_j) \mid \mathcal{F}_\infty\right] = \prod_{i=1}^m \mathbb{E}[g_i(X_j) \mid \mathcal{F}_\infty] = \prod_{i=1}^m \mathbb{E}[g_i(X_1) \mid \mathcal{F}_\infty]. \tag{12.21}$$

c. If S is Polish, $\mathcal{S} = \mathcal{B}(S)$, then the distribution γ of $(X_1, X_2, \ldots, \ldots)$ is a mixture of product measures Q^∞ on $(S^\infty, \mathcal{S}^{\otimes\infty})$, i.e.,

$$\gamma(A) = \int_{\mathcal{P}(S)} Q^\infty(A)\nu(dQ) \quad (A \in \mathcal{S}^{\otimes\infty}), \tag{12.22}$$

where $Q \to Q^\infty(A)$ is, for each $A \in \mathcal{S}^{\otimes\infty}$, a measurable map on the space $\mathcal{P}(S)$ of all probability measures (on S) endowed with the Borel σ-field under the weak topology.

Proof.

a. The $\{\mathcal{F}_n\}_{n=1}^\infty$-reverse martingale $E[g(X_1) \mid \mathcal{F}_n]$ converges a.s. and in L^1 to $E[g(X_1) \mid \mathcal{F}_\infty]$, by Proposition 12.10. By symmetry, $E[g(X_1) \mid \mathcal{F}_n] = (1/n)\sum_{j=1}^n g(X_j)$, as in the first part of the proof of Corollary 12.11.

b. Fix $n \geq m$. For every permutation $(\pi(1), \ldots, \pi(m))$ of $(1, 2, \ldots, n)$ taken m at a time, one has

$$\mathbb{E}\left[\prod_{i=1}^m g_i(X_i) \mid \mathcal{F}_n\right] = \mathbb{E}\left[\prod_{i=1}^m g_i(X_{\pi(i)}) \mid \mathcal{F}_n\right] \quad \text{a.s.} \tag{12.23}$$

The average of the right side over the $(n)_m := n(n-1)\cdots(n-m+1)$ permutations π, as $n \to \infty$, is (Exercise 4)

$$\frac{1}{(n)_m}\prod_{i=1}^m \left\{\sum_{j=1}^n g_i(X_j)\right\} + \frac{O(n^{m-1})}{(n)_m} \quad \text{a.s.,} \tag{12.24}$$

which converges a.s. to $\prod_{i=1}^m \mathbb{E}[g(X_1) \mid \mathcal{F}_\infty]$ as $n \to \infty$, by (a). Since the left side of (12.23) converges a.s. to $\mathbb{E}[\prod_{i=1}^m g_i(X_i) \mid \mathcal{F}_\infty]$, by Proposition 12.10, the first and third terms in (12.21) equal. Now use $\mathbb{E}[g_i(X_1) \mid \mathcal{F}_\infty] = \mathbb{E}[g_i(X_j) \mid \mathcal{F}_\infty]$ a.s., due to exchangeability (e.g., it is true with $\mathcal{F}_\infty$ replaced by $\mathcal{F}_n$ where $n \geq j$; now let $n \to \infty$).

c. Since S is Polish and $\mathcal{S}$ its Borel σ-field, given $\mathcal{F}_\infty$ regular conditional distributions $Q_1(\omega)$, say, of X_1 and $Q_{1,2,\ldots,m}(\omega)$ of $(X_1, \ldots, X_m)$ exist ($m \geq 1$), and $Q_{1,2,\ldots,m}(\omega)$ is the m-fold product probability $Q_1(\omega) \times Q_1(\omega) \times \cdots \times Q_1(\omega)$ ($m \geq 1$), outside a P-null set. Hence a regular conditional distribution $Q_\infty(\omega)$, say, exists for $(X_1, X_2, \ldots)$ (given $\mathcal{F}_\infty$) and it is given by the infinite product probability $Q_1^\infty(\omega)$, on $(S^\infty, \mathcal{S}^{\otimes\infty})$. Since $\gamma(A) = \mathbb{E}(Q_1^\infty(\omega)(A))$, (12.23) follows with $\omega \to Q_1^\infty(\omega)$ a map on $(\Omega, \mathcal{F}, P)$ into $\mathcal{P}(S^\infty)$ having distribution ν, say.

■

Exercises

1. Prove that a non-negative $\{\mathcal{F}_n\}_{n=1}^\infty$-supermartingale $\{X_n : n \in \mathbb{N}\}$ converges a.s. to a random variable X_∞ satisfying $\mathbb{E}X_\infty \leq \mathbb{E}X_1$. [*Hint*: Apply Theorem 12.2, noting that $\mathbb{E}X_n \leq \mathbb{E}X_1\ \forall n$.]
2. Let $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}[0, 1)$, and $P =$ Lebesgue measure on [0, 1]. Suppose f is integrable on $(\Omega, \mathcal{F}, P)$.

   (a) Suppose $0 = x_0 < x_1 < \cdots < x_k = 1$ is a partition of [0, 1), and $\mathcal{F}$ is the σ-field on [0, 1] generated by $\{[x_i, x_{i+1}) : i = 0, 1, \ldots, k-1\}$. Prove that $\mathbb{E}(f \mid \mathcal{F})(x) = f_i$ for $x \in [x_i, x_{i+1})$, where $f_i = \int_{[x_i, x_{i+1})}, f(x)dx$ ($i = 0, 1, \cdots, k-1$).

   (b) Let $0 = x_0^{(n)} < x_1^{(n)} < \cdots < x_{k_n}^{(n)} = 1$ denote a sequence of partitions $\pi_n = \{x_0^{(n)}, \ldots, x_{k_n}^{(n)}\}$ such that $\pi_n \subset \pi_{n+1}\ \forall\ n$ and $\delta_n := \max\{x_{i+1}^{(n)} - x_i^{(n)} : 0 \leq i \leq k_n - 1\} \to 0$ as $n \to \infty$. Prove that $\mathbb{E}(f \mid \mathcal{F}_n) = \sum_{i=0}^{k_n-1} f_i^{(n)} \mathbf{1}_{[x_i, x_{i+1}^{(n)})}(x) \to f(x)$ a.s. (λ) and in L^1 as $n \to \infty$, and $\sum_{i=1}^{k_{n+1}} f_i^{(n)}(x_{i+1}^{(n)} - x_i^{(n)}) \to \int_{[0,1]} f(x)dx$. Here $f_i^{(n)} := \int_{[x_i^{(n)}, x_{i+1}^{(n)})} f(x)dx$.

   (c) State and prove proper extensions of (a), (b) when λ is replaced by an arbitrary probability measure on [0, 1).

3. Let $\{X_n : n \in \mathbb{N}\}$ be a uniformly integrable $\{\mathcal{F}_n\}_{n=1}^\infty$-submartingale.

   (a) Show that X_n converges a.s. and in L^1 to a random variable X_∞, say, and $\mathbb{E}(X_\infty \mid \mathcal{F}_n) \geq X_n$ for all n, a.s.

   (b) If, in addition, $\{X_n\}_{n=1}^\infty$ is a $\{\mathcal{F}_n\}_{n=1}^\infty$-martingale, show that $\mathbb{E}(X_\infty \mid \mathcal{F}_n) = X_n$ a.s. for all n.

4. Show that
   (a) the average of the right side of (12.23) over the $(n)_m$ permutations π (of $(1, 2, \ldots, n)$ taken m at a time) is of the form (12.24). [*Hint*: The number of terms in the expansion of the product $\prod_{i=1}^{m}\{\sum_{j=1}^{n} g_i(X_j)\}$ which have repeated indices j is of the order $O(n^{m-1})$.]
   (b) Write out this average over $(n)_m$ permutations, with the leading term given in (12.24) followed by terms with repeated indices.
5. Consider an exchangeable sequence of Bernoulli $0-1$ valued random variables $X_1, X_2, \ldots$.
   (a) Use deFinetti's theorem to show that there is probability measure ν on $[0, 1]$ such that for any $\epsilon_j \in \{0, 1\}, 1 \leq j \leq n, \sum_{j=1}^{n} \epsilon_j = k$, $P(X_1 = \epsilon_1, \ldots, X_n = \epsilon_n) = \int_{[0,1]} x^k (1-x)^{n-k} \nu(dx)$.
   (b) Use the Riesz representation theorem to provide an alternative proof to (i). [*Hint*: Starting with $\ell(x^0) = 1, \ell(x^n) = P(X_1 = 1, \ldots, X_n = 1), n \geq 1$, construct a well-defined bounded linear functional on $C[0, 1]$ using Weierstrass approximation and the inclusion-exclusion formula.]
6. In reference to the Lebesgue decomposition Corollary 12.8
   (i) Show that $\{M_n\}_{n=1}^{\infty}$ and $\{R_n\}_{n=1}^{\infty}$ define non-negative bounded martingales whose respective limits M_∞ and R_∞ exist a.s. with respect to λ.
   (ii) Show that $M_\infty = dm/d\lambda$ and $R_\infty = dq/d\lambda$. [*Hint*: Use dominated convergence to argue that $m(A) = \int_A M_\infty d\lambda$ on the π-system $\cup_n \mathcal{F}_n$ which generates $\mathcal{F}$ and apply the $\pi - \lambda$-theorem. The argument is the same for R_∞.]
   (ii) Prove the asserted equivalent conditions for absolute continuity and mutual singularity in Corollary 12.8.
7. (*Radon–Nikodym Derivatives*) This exercise provides a conceptual representation of Radon–Nikodym derivatives based on the Corollary 12.8. For a given integer $b \geq 2$ consider successive partitions of the unit interval $S = [0, 1)$ into subintervals $\Delta_{k,n} := [kb^{-n}, (k+1)b^{-n}), k = 0, 1, \ldots b^n - 1)$. Let $\mathcal{F}_n = \sigma(\Delta_{k,n} : 0 \leq k \leq b^n - 1$. Suppose that μ, ν are two positive finite measures on the Borel σ-field of S with $\mu << \nu$. Define a sequence of piecewise constant functions on S by the ratios $\rho_n(x) := \mu(\Delta_{k,n})/\nu(\Delta_{k,n})$ for $x \in \Delta_{k,n}$. Show that $\rho_\infty = \lim_{n\to\infty} \rho_n$ exists ν-a.e. and $d\mu/d\nu = \rho_\infty$.
8. Under the hypothesis of Theorem 12.3(a), show that $\{X_n\}$ is uniformly integrable. [*Hint*: $|X_n| \leq M_n \uparrow M$ almost surely, and by the monotone convergence theorem, $\mathbb{E}M \leq \sup_n \mathbb{E}|X_n| \log^+ |X_n| \infty$.] Similarly, prove that, under the hypothesis of Theorem 12.3(b), $\{|X_n|^p\}$ is uniformly integrable. [*Hint*: Note $|X_n|^p \leq M^p$ for all n, and $\mathbb{E}M^p < \infty$, by monotone convergence.]

Chapter 13
Continuous Parameter Martingales

In this chapter some of the main theorems for discrete parameter martingales obtained in previous chapters are extended to continuous parameter martingales. A central point is the use of martingale theory for the regularization of sample paths of stochastic processes.

The index set T here is an interval contained in $\{0, \infty)$; most often, $T = [0, \infty)$. $\overline{T}$ denotes the closure of T in $[0, \infty]$, which may be thought of as a one-point compactification of $[0, \infty)$.

Theorem 13.1. Let $\{X_t : t \in T\}$ be a right-continuous non-negative $\{\mathcal{F}_t : t \in T\}$–submartingale, with $\mathbb{E}|X_t|^p < \infty$ for some $p \geq 1$. Then $M_t := \sup\{X_s : s \leq t\}$ is $\mathcal{F}_t$–measurable ($t \in T$), and
(a) for all $\lambda > 0$,

$$P(M_t > \lambda) \leq \frac{1}{\lambda^p} \int_{[M_t > \lambda]} X_t^p \leq \frac{1}{\lambda^p} \mathbb{E} X_t^p; \tag{13.1}$$

(b) if $\mathbb{E}|X_t|^p < \infty$ for some $p > 1$, then

$$\|M_t\|_p := (\mathbb{E} M_t^p)^{1/p} \leq q \|X_t\|_p, \tag{13.2}$$

where q is the conjugate of p, i.e., $(1/p) + (1/q) = 1$.

Proof. (a) Let $t_{1,n} < t_{2,n} < \cdots < t_{k_n,n} = t$ be in T, $\{t_{j,n} : 1 \leq j \leq k_n\} \uparrow$ with respect to inclusion as $n \uparrow$, and $\cup_{n=1}^{\infty}\{t_{j,n} : 1 \leq j \leq k_n\}$ dense in $\{s \in T : s \leq t\}$. Write $M_{t,n} := \max\{X_{t_{j,n}} : 1 \leq j \leq k_n\}$. For $\lambda > 0$, $[M_{t,n} > \lambda] \uparrow [M_t > \lambda]$ as $n \uparrow \infty$, by right-continuity of $\{X_s : s \in T\}$. In particular, M_t is $\mathcal{F}_t$–measurable.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_13

Also, by Theorem 10.6,

$$P(M_{t,n} > \lambda) \leq \frac{1}{\lambda^p} \int_{[M_{t,n} > \lambda]} X_t^p \, dP \leq \frac{1}{\lambda^p} \mathbb{E} X_t^p. \tag{13.3}$$

Letting $n \uparrow \infty$, one obtains (13.1).

(b) is obtained similarly from Theorem 10.6(b). ∎

To state the optional stopping and optional sampling theorems we need to define $\{\mathcal{F}_t : t \in T\}$–stopping times. The intuitive idea of τ as a stopping time strategy is that to "stop by time t, or not," according to τ, is determined by the knowledge of the past up to time t, and does not require "a peek into the future."

Definition 13.1. Let $\{\mathcal{F}_t : t \in T\}$ be a filtration on a probability space $(\Omega, \mathcal{F}, P)$, with T a linearly ordered index set to which one may adjoin, if necessary, a point "∞" as the largest point of $T \cup \{\infty\}$. A random variable $\tau : \Omega \to T \cup \{\infty\}$ is a $\{\mathcal{F}_t : t \in T\}$-*stopping time* if $[\tau \leq t] \in \mathcal{F}_t \; \forall \, t \in T$. If $[\tau < t] \in \mathcal{F}_t$ for all $t \in T$, then τ is called an *optional time*.

Most commonly, T in this definition is $\mathbf{N}$ or $\mathbf{Z}^+$, or $[0, \infty)$, $T = [0, C]$ for $0 < C < \infty$, and τ is related to a$\{\mathcal{F}_t : t \in T\}$-adapted process $\{X_t : t \in T\}$.

Example 1. Let $\{X_t : t \in T\}$ be a $\{\mathcal{F}_t : t \in T\}$-adapted process with values in a measurable space $(S, \mathcal{S})$, with a linearly ordered index set. (a) If $T = \mathbf{N}$ or $\mathbf{Z}^+$, then for every $B \in \mathcal{S}$,

$$\tau_B := \inf\{t \geq 0 : X_t \in B\} \tag{13.4}$$

is a $\{\mathcal{F}_t : t \in T\}$-stopping time. (b) If $T = \mathbb{R}_+ \equiv [0, \infty)$, S is a metric space $\mathcal{S} = \mathcal{B}(S)$, B is *closed,* and $t \to X_t$ is continuous, then τ_B is a $\{\mathcal{F}_t\}_{t \in T}$-stopping time (Exercise 2(i)) (c) If $T = \mathbb{R}_+$, S is a topological space, $t \to X_t$ is right continuous, and B is *open* then $[\tau_B < t] \in \mathcal{F}_t$ for all $t \geq 0$, i.e., τ_B is an optional time (Exercise 2(ii)).

Definition 13.2. Let $\{\mathcal{F}_t : t \in T\}$ be a filtration on $(\Omega, \mathcal{F})$. Suppose that τ is a $\{\mathcal{F}_t\}$–stopping time. The *pre-τ sigmafield* $\mathcal{F}_\tau$ comprises all $A \in \mathcal{F}$ such that $A \cap [\tau \leq t] \in \mathcal{F}_t$ for all $t \in T$.

Heuristically, $\mathcal{F}_\tau$ comprises events determined by information available only up to time τ. For example, if T is discrete with elements $t_1 < t_2 < \cdots$, and $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t) \subset \mathcal{F}, \forall t$, where $\{X_t : t \in T\}$ is a process with values in some measurable space $(S, \mathcal{S})$, then $\mathcal{F}_\tau = \sigma(X_{\tau \wedge t} : t \in T)$; (Exercise 8). The stochastic process $\{X_{\tau \wedge t} : t \in T\}$ is referred to as the *stopped process*.

The $\mathcal{F}_\tau$–measurability of τ is easy to verify. Also if τ_1, τ_2 are two $\{\mathcal{F}_t : t \in T\}$–stopping times and $\tau_1 \leq \tau_2$, then it is simple to check that (Exercise 1)

$$\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}. \tag{13.5}$$

Suppose $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$–adapted process with values in a measurable space $(S, \mathcal{S})$, and τ is a $\{\mathcal{F}_t : t \in T\}$–stopping time. Unlike the discrete parameter case, the joint measurability $(t, \omega) \to X_t(\omega)$ is needed for many purposes.

Suppose $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$-adapted process with values in a measurable space $(S, \mathcal{S})$, and τ is a $\{\mathcal{F}_t : t \in T\}$-stopping time. For many purposes the following notion of adapted joint measurability of $(t, \omega) \to X_t(\omega)$ is important.

Definition 13.3. A stochastic process $\{X_t : t \in T\}$ with values in a measurable space $(S, \mathcal{S})$ is *progressively measurable* with respect to $\{\mathcal{F}_t : t \in T\}$ if, for each $t \in T$, the map $(s, \omega) \to X_s(\omega)$ is measurable with respect to the $\sigma-$fields $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ (on $[0, t] \times \Omega$) and $\mathcal{S}$ (on S). Here $\mathcal{B}[0, t]$ is the Borel $\sigma-$field on $[0, t]$, $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ is the usual product $\sigma-$field.

Proposition 13.2. (a) Suppose $\{X_t : t \in T\}$ is progressively measurable, and τ is a stopping time. Then X_τ is $\mathcal{F}_\tau$–measurable, i.e., $[X_\tau \in B] \cap [\tau \le t] \in \mathcal{F}_t$ for each $B \in \mathcal{S}$ and each $t \in T$. (b) Suppose S is a metric space and $\mathcal{S}$ its Borel $\sigma-$field. If $\{X_t : t \in T\}$ is right-continuous, then it is progressively measurable.

Proof. (a) Fix $t \in T$. On the set $\Omega_t := [\tau \le t]$, X_τ is the composition g of the maps (i) $f(\omega) = (\tau(\omega), \omega)$, on Ω_t into $[0, t] \times \Omega_t$, and (ii) $g(s, \omega) = X_s(\omega)$ on $[0, t] \times \Omega_t$ into S. Now f is $\widetilde{\mathcal{F}}_t$–measurable on Ω_t, where $\widetilde{\mathcal{F}}_t$ is the *trace $\sigma-$field* $\{A \cap \Omega_t : A \in \mathcal{F}_t\}$ on Ω_t, and $B[0, t] \otimes \widetilde{\mathcal{F}}_t$ is the $\sigma-$field on $[0, t] \times \Omega_t$. Next, the map $g(s, \omega) = X_s(\omega)$ on $[0, t] \times \Omega$ into S is $\mathcal{B}[0, t] \otimes \mathcal{F}_t$–measurable. Therefore, the restriction of this map to the measurable subset $[0, t] \times \Omega_t$ is measurable on the *trace $\sigma-$field* $\{A \cap ([0, t] \times \Omega_t) : A \in \mathcal{B}[0, t] \otimes \mathcal{F}_t\}$. Therefore, the composition X_τ is $\widetilde{\mathcal{F}}_t$–measurable on Ω_t, i.e., $[X_\tau \in B] \cap [\tau \le t] \in \widetilde{\mathcal{F}}_t \subset \mathcal{F}_t$. (b) Fix $t \in T$. Define, for each positive integer n, the stochastic process $\{X_s^{(n)} : 0 \le s \le t\}$ by

$$X_s^{(n)} := X_{j2^{-n}t} \quad \text{for} \quad (j-1)2^{-n}t \le s < j2^{-n}t \quad (1 \le j \le 2^n), \;\; X_t^{(n)} = X_t. \tag{13.6}$$

Since $\{(s, \omega) : X_s^{(n)}(\omega) \in B\} = \cup_{j=1}^{2^n}([j-1)2^{-n}t, j2^{-n}t) \times \{\omega : X_{j2^{-n}t}(\omega) \in B\}) \cup (\{t\} \times \{\omega : X_t(\omega) \in B\}) \in \mathcal{B}[0, t] \otimes \mathcal{F}_t$, $\{X_t^{(n)} : t \ge 0\}$ is progressively measurable. Now $X_t^{(n)}(\omega) \to X_t(\omega)$ for all (t, ω) as $n \to \infty$, in view of the right-continuity of $t \to X_t(\omega)$. Hence $\{X_t : t \in T\}$ is progressively measurable. ∎

Remark 13.1. It is often important to relax the assumption of 'right-continuity' of $\{X_t : t \in T\}$ to "a.s. right-continuity". To ensure *progressive measurability* in this case, it is convenient to take $\mathcal{F}$, $\mathcal{F}_t$ to be *P–complete,* i.e., if $P(A) = 0$ and $B \subset A$ then $B \in \mathcal{F}$ and $B \in \mathcal{F}_t \, \forall t$. Then modify X_t to equal $X_0 \; \forall t$ on the P-null set $N = \{\omega : t \to X_t(\omega)$ is not right-continuous$\}$. This modified $\{X_t : t \in T\}$, $\{\mathcal{F}_t : t \in T\}$ satisfy the hypothesis of part (b) of Proposition 13.2.

Theorem 13.3 *(Optional Stopping).* Let $\{X_t : 0 \le t < \infty\}$ be a right-continuous $\{\mathcal{F}_t : t \in T\}$–submartingale and $\tau_1 \le \tau_2$ two a.s. finite $\{\mathcal{F}_t : t \in T\}$–stopping times such that $X_{\tau_i \wedge m} \to X_{\tau_i}$ in L^1 $(i = 1, 2)$, as $m \to \infty$ (or, equivalently, $\mathbb{E}|X_{\tau_i}| < \infty$ $(i = 1, 2)$ and $\mathbb{E}(|X_m|\mathbf{1}_{[\tau_2 > m]}) \to 0$ as $m \to \infty$). Then

$$\mathbb{E}\left(X_{\tau_2} \mid \mathcal{F}_{\tau_1}\right) \geq X_{\tau_1} \text{ a.s.,} \tag{13.7}$$

with equality if $\{X_t : t \in T\}$ is a $\{\mathcal{F}_t : t \in T\}$–martingale.

Proof. Define, for each positive integer n, and each $i = 0, 1$, the stopping time

$$\tau_i^{(n)} := \frac{k}{2^n} \quad \text{on} \quad \left[\frac{k-1}{2^n} < \tau_i \leq \frac{k}{2^n}\right] \quad (k = 0, 1, \ldots). \tag{13.8}$$

Then $\tau_i^{(n)}$ is a $\{\mathcal{F}_t : t \in T\}$–stopping time, $\tau_i \leq \tau_i^{(n)} \leq \tau_i + 2^{-n}$ for all n $(i = 1, 2)$. Since $\tau_i^{(n)}$ is a stopping time with respect to the discrete parameter family $\{\mathcal{F}_{k2^{-n}} : k = 0, 1, 2, \ldots\}$, for each positive integer n, so is $\tau_i^{(n)} \wedge m$ for each positive integer m, Theorem 11.1 may be applied (or, use Corollary 11.2) to obtain, in the *martingale case,*

$$\mathbb{E}\left(X_{\tau_2^{(n)} \wedge m} \cdot \mathbf{1}_A\right) = \mathbb{E}\left(X_{\tau_1^{(n)} \wedge m} \cdot \mathbf{1}_A\right) \quad (A \in \mathcal{F}_{\tau_1 \wedge m}). \tag{13.9}$$

Here we have used the fact $\mathcal{F}_{\tau_1 \wedge m} \subset \mathcal{F}_{\tau_1^{(n)} \wedge m}$. For each m, we will now show that $\{X_{\tau_1^{(n)} \wedge m} : n \geq 1, i = 1, 2\}$ is a uniformly integrable family. Consider for this the two–element martingale $\{X_{\tau_i^{(n)} \wedge m}, X_m\}$ with σ–fields $\mathcal{F}_{\tau_i^{(n)} \wedge m}, \mathcal{F}_m$ (Theorem 11.1). It follows from the submartingale property of $\{|X_{\tau_i^{(n)} \wedge m}|, |X_m|\}$ that

$$\begin{aligned} \mathbb{E}\left|X_{\tau_i^{(n)} \wedge m}\right| \mathbf{1}_{[|X_{\tau_i^{(n)} \wedge m}| > \lambda]} &\leq \mathbb{E}|X_m| \mathbf{1}_{[|X_{\tau_i^{(n)} \wedge m}| > \lambda]}, \\ P\left(|X_{\tau_i^{(n)} \wedge m}| > \lambda\right) &\leq \frac{1}{\lambda}\mathbb{E}\left|X_{\tau_i^{(n)} \wedge m}\right| \leq \frac{1}{\lambda}\mathbb{E}|X_m|, \end{aligned} \tag{13.10}$$

proving the desired uniform integrability. Now, since $X_{\tau_i^{(n)} \wedge m} \to X_{\tau_i \wedge m}$ a.s. as $n \to \infty$, by right-continuity of $\{X_t : t \in T\}$, this convergence is also in L^1. Therefore (13.9) yields

$$\mathbb{E}\left(X_{\tau_2 \wedge m} \cdot \mathbf{1}_A\right) = \mathbb{E}\left(X_{\tau_1 \wedge m} \cdot \mathbf{1}_A\right) \qquad A \in \mathcal{F}_{\tau_1 \wedge m}. \tag{13.11}$$

By hypothesis, $X_{\tau_i \wedge m} \to X_{\tau_i}$ in L^1 as $m \to \infty$. Hence taking $A = B \cap [\tau_1 \leq m]$ with $B \in \mathcal{F}_{\tau_1}$ (so that $A \in \mathcal{F}_{\tau_1 \wedge m}$), one obtains from (13.11), in the limit as $m \to \infty$,

$$\mathbb{E}X_{\tau_2}\mathbf{1}_B = \mathbb{E}X_{\tau_1}\mathbf{1}_B \qquad \forall\ B \in \mathcal{F}_{\tau_1}, \tag{13.12}$$

completing the proof for the martingale case.

For the case of a submartingale $\{X_t : t \in T\}$, consider for each $c \in \mathbb{R}$, and given n and m, the two-component submartingale $\{X_{1,n} := \max(X_{\tau_1^{(n)} \wedge m}, c),\ X_{2,n} := \max(X_{\tau_2^{(n)} \wedge m}, c)\}$ (Theorem 11.1 and Corollary10.2). Then the equality in (13.9) becomes an inequality "$\geq$" with $X_{i,n}$ in place of $X_{\tau_i^{(n)} \wedge m}$ $(i = 1, 2)$. Since $\{X_{i,n} : n \geq 1, i = 1, 2\}$ is uniformly integrable (Look at the two–element martingale $\{X_{i,n}, X_m := \max(X_{m,C})\}$ for each $i = 1, 2$.), one obtains in place of (13.11) the inequality $\mathbb{E}X_2\mathbf{1}_A \geq \mathbb{E}X_1\mathbf{1}_A$ with $X_i := \max(X_{\tau_i \wedge m}, c)$ $(i = 1, 2)$, $A = B \cap [\tau_1 \leq m]$, $B \in \mathcal{F}_{\tau_1}$. Letting $m \to \infty$, and by hypothesis, one obtains $\mathbb{E}Z_2\mathbf{1}_B \geq \mathbb{E}Z_1\mathbf{1}_B$ where $Z_i := \max(X_{\tau_i}, c)$, $\forall\ B \in \mathcal{F}_{\tau_1}$. Now let $c \downarrow -\infty$ and use the monotone convergence theorem to get the desired relation $\mathbb{E}X_{\tau_2}\mathbf{1}_B \geq \mathbb{E}X_{\tau_1}\mathbf{1}_B$ $\forall\ B \in \mathcal{F}_{\tau_1}$. ■

The equivalence of the conditions (a) $X_{\tau_i \wedge m} \to X_{\tau_i}$ in L^1 as $m \to \infty$ $(i = 1, 2)$, and (b) $\mathbb{E}|X_{\tau_i}| < \infty$ $(i = 1, 2)$ and $\mathbb{E}|X_m|\mathbf{1}_{[\tau_2 > m]} \to 0$ as $m \to \infty$, is proved in exactly the same manner as the lemma following the statement of Theorem 11.1.

Definition 13.4. A stochastic process $\{Y_t : t \in T\}$ is said to be a *modification* or *version* of a stochastic process $\{X_t : t \in T\}$ if $P(Y_t = X_t) = 1\ \forall t$, and $\{Y_t : t \in T\}$ is an *indistinguishable version* of $\{X_t : t \in T\}$ if $P(Y_t = X_t\ \forall t) = 1$.

Note that if $\{Y_t : t \in T\}$ is a version of $\{X_t : t \in T\}$, then the finite-dimensional distributions of $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ are the same.

Remark 13.2. It follows from Remark 13.1 that if $\mathcal{F}$, $\mathcal{F}_t$ are P–complete and $\{X_t : t \in T\}$ is an *a.s.* right-continuous $\{\mathcal{F}_t : t \in T\}$–(sub)martingale, then it has an indistinguishable right-continuous version which is a $\{\mathcal{F}_t : t \in T\}$–(sub)martingale, and Theorem 13.3 applies to this version. But, in fact, the proof of the theorem remains valid, if one replaces "right-continuous" by "a.s. right-continuous" in the statement, provided $X_{\tau_i}\mathbf{1}_{[\tau_2 < \infty]}$ is *assumed* $\mathcal{F}_{\tau_i}$–measurable $(i = 1, 2)$. Indeed, (13.12) (with inequality for the submartingale case) holds even without this assumption of $\mathcal{F}_{\tau_i}$–measurability of X_{τ_i}.

Corollary 13.4. If $\{X_t : 0 \leq t < \infty\}$ is a right-continuous $\{\mathcal{F}_t : t \in T\}$–(sub)martingale, and τ is a $\{\mathcal{F}_t : t \in T\}$–stopping time, then $\{X_{\tau \wedge t} : t \geq 0\}$ is a $\{\mathcal{F}_t : t \in T\}$–(sub)martingale.

Again, as indicated in Remark 13.2 above, one may assume a.s. right-continuity in the corollary provided $X_{\tau \wedge t}$ is $\mathcal{F}_t$–measurable $\forall t$.

Example 2 *(Hitting by Brownian Motion of a Two-Point Boundary)*. Let $\{B_t^x : t \geq 0\}$ be a one dimension standard Brownian motion starting at x, and let $c < x < d$. Let τ denote the stopping time, $\tau = \inf\{t \geq 0 : B_t^x = c \text{ or } d\}$. Then $\mathbb{E}B_\tau^x = \mathbb{E}B_0^x = x$, or writing $\psi(x) := P(\{B_t^x\}_{t \geq 0}$ reaches d before $c)$, $d\psi(x) + c(1 - \psi(x)) = x$, or,

$$\psi(x) = \frac{x - c}{d - c} \qquad c < x < d. \tag{13.13}$$

Applying Theorem 13.3 to the martingale $X_t := (B_t^x - x)^2 - t$, one gets $\mathbb{E}X_\tau = 0$, or $(d-x)^2\psi(x) + (x-c)^2(1-\psi(x)) = \mathbb{E}\tau$, so that $\mathbb{E}\tau = [(d-x)^2 - (x-c)^2]\psi(x) + (x-c)^2$, or

$$\mathbb{E}\tau = (d-x)(x-c). \tag{13.14}$$

Consider now a Brownian motion $\{Y_t^x : t \geq 0\}$ with nonzero drift μ and diffusion coefficient $\sigma^2 > 0$, starting at x. Then $\{Y_t^x - t\mu : t \geq 0\}$ is a martingale, so that $\mathbb{E}(Y_\tau^x - \mu\tau) = x$, i.e., $d\psi_1(x) + c(1-\psi_1(x)) - \mu\mathbb{E}\tau = x$, or,

$$(d-c)\psi_1(x) - \mu\mathbb{E}\tau = x - c, \tag{13.15}$$

where $\psi_1(x) = P(Y_\tau^x = d)$, i.e., $\{Y_t^x : t \geq 0\}$ reaches d before c. There are two unknowns, ψ_1 and $\mathbb{E}\tau$ in (13.15), so we need one more relation to solve for them. Consider the exponential martingale $Z_t := \exp\{\xi(Y_t^x - t\mu) - \frac{\xi^2\sigma^2}{2}t\}$ $(t \geq 1)$ (See Exercise 11(i)). $Z_0 = e^{\xi x}$, so that $e^{\xi x} = \mathbb{E}Z_\tau = E\exp\{\xi(d-\tau\mu) - \xi^2\sigma^2\tau/2\}\mathbf{1}_{[Y_\tau^x = d]} + \mathbb{E}[\exp\{\xi(c-\tau\mu) - \xi^2\sigma^2\tau/2\}\mathbf{1}_{[Y_\tau^x = d]}]$. Take $\xi \neq 0$ such that the coefficient of τ in the exponent is zero, i.e., $\xi\mu + \xi^2\sigma^2/2 = 0$, or $\xi = -2\mu/\sigma^2$. Then Theorem 13.3 yields

$$\begin{aligned} e^{-2\mu x/\sigma^2} &= \exp\{\xi d\}\psi_1(x) + \exp\{\xi c\}(1-\psi_1(x)) \\ &= \psi_1(x)\left[\exp\left\{-\frac{2\mu d}{\sigma^2}\right\} - \exp\left\{-\frac{2\mu c}{\sigma^2}\right\}\right] = \exp\left\{+\frac{2\mu c}{\sigma^2}\right\}, \end{aligned}$$

or,

$$\psi_1(x) = \frac{\exp\{-2\mu x/\sigma^2\} - \exp\{-2\mu c/\sigma^2\}}{\exp\{-\frac{2\mu d}{\sigma^2}\} - \exp\{-\frac{2\mu c}{\sigma^2}\}}, \tag{13.16}$$

one may use this to compute $\mathbb{E}\tau$,

$$\mathbb{E}\tau = \frac{(d-c)\psi_1(x) - (x-c)}{\mu}. \tag{13.17}$$

Checking the hypotheses of Theorem 13.3 for the validity of the relations (13.13)–(13.17) are left to Exercise 11(ii).

The following property of Brownian motion is basic to the theory of stochastic differential equations (*SDE*). It is included here as an application of the use of the maximal inequality in the analysis of Brownian motion sample path structure (also see Proposition 6.6).

Theorem 13.5. (a) Outside a P–null set, $\{B_t : t \geq 0\}$ is of unbounded variation on every non-degenerate interval, and (b) for every $t > 0$,

$$\max_{1\le N\le 2^n}\left|\sum_{i=1}^{N}\left(B_{i2^{-n}t}-B_{(i-1)2^{-n}t}\right)^2-N2^{-n}t\right|\longrightarrow 0,\quad \text{a.s. as } n\to\infty. \tag{13.18}$$

Proof. (a) Let $V_n := \sum_{i=1}^{2^n}|B_{i2^{-n}}-B_{(i-1)2^{-n}}|$, $(n=1,2,\cdots)$. Then $V_{n+1}\ge V_n$ for all n. Since $B_{i2^{-n}}-B_{(i-1)2^{-n}}$ $(i=1,2,\cdots,2^n)$ are i.i.d. $N(0,2^{-n})$, one has $\mathbb{E}V_n = 2^n\mathbb{E}|B_{2^{-n}}-B_0| = 2^{n/2}\mathbb{E}|B_1|$ as $\mathbb{E}|B_{2^{-n}}| = \mathbb{E}|B_1/2^{n/2}|$. Therefore by Chebyshev's inequality, for all $M>0$, $2^{n/2}>M$,

$$\begin{aligned}P(V_n\le M)&\equiv P(V_n-\mathbb{E}V_n\le M-2^{n/2}\mathbb{E}|B_1|)\\&\le P(|V_n-\mathbb{E}V_n|\ge 2^{n/2}\mathbb{E}|B_1|-M)\le 1/(2^{n/2}\mathbb{E}|B_1|-M)^2,\end{aligned}$$

which goes to zero as $n\to\infty$. Therefore, $P(V_n>M)\to 1$ as $n\to\infty$, i.e., $V_n\uparrow\infty$ a.s. Applying the same argument to dyadic intervals $[a,b]$ (instead of $[0,1]$), one gets the desired result. (b) Since, for each n, the finite sequence $Y_N := \sum_{i=1}^{N}(B_{i2^{-n}t}-B_{(i-1)2^{-n}t})^2-N2^{-n}t$ $(N=1,2,\ldots,2^n)$ is a martingale, Doob's maximal inequality with $p=2$ yields (see Theorem 10.6, or Kolmogorov's maximal inequality Chapter 10, Exercise 7).

$$P\left(\max_{1\le N\le 2^n}|Y_N|>n2^{-n/2}\right)\le\frac{ct^2}{n^2}\qquad (c:=\mathbb{E}(B_1^2-1)^2=2). \tag{13.19}$$

Now apply the Borel–Cantelli lemma to get (13.18). ■

Remark 13.3. If one *defines* the *quadratic variation* $v([s,t],f)$ of a function f on $[s,t]$, $s<t$, to be the limit, if it exists, of

$$v_n([s,t],f):=\sum_{i=1}^{2^n}\left[f(s+i2^{-n}(t-s))-f(s+(i-1)2^{-n}(t-s))\right]^2$$

as $n\to\infty$, then Theorem 13.5(b) implies, in particular, that for every $s<t$ the standard Brownian motion has a finite quadratic variation $t-s$ on the interval $[s,t]$, outside a P–null set. This fact leads to the following distinctive symbolism often employed in the calculus of stochastic differential equations:

$$(dB_t)^2=dt\quad\text{a.s.} \tag{13.20}$$

It follows from Theorem 13.5(b) that outside a P-null set, the *p-th variation* of Brownian motion on any interval $[s,t]$, $s<t$, defined by the limit (if it exists)

$$v_{n,p}\left([s,t],\{B_t:t\ge 0\}\right):=\sum_{i=1}^{2^n}\left|B_{s+i2^{-n}(t-s)}-B_{s+(i-1)2^{-n}(t-s)}\right|^p \tag{13.21}$$

is, outside a P–null set, infinite if $p < 2$ and zero if $p > 2$ (Exercise 9).

We next show that the crucial assumption of right-continuity of a submartingale required in this chapter is satisfied by an appropriate modification of the submartingale.

Definition 13.5. Let $\{\mathcal{F}_t : t \in T\}$ be a filtration on an interval T. Define $\mathcal{F}_{t+} := \cap_{s>t}\mathcal{F}_s$ (with $\mathcal{F}_{t+} := \mathcal{F}_t$ if t is a right end point), $\mathcal{F}_{t-} = \sigma(\cap_{s<t}\mathcal{F}_s)$ (with $\mathcal{F}_{t-} := \mathcal{F}_t$ if t is a left end point). A filtration $\{\mathcal{F}_t : t \in T\}$ is said to be *right*– (or, *left–) continuous* if $\mathcal{F}_{t+} = \mathcal{F}_t \; \forall t$ (resp., $\mathcal{F}_{t-} = \mathcal{F}_t \; \forall t$).

Remark 13.4. In the case of filtrations generated by a stochastic process, i.e., $\mathcal{F}_t = \sigma(X_s : s \le t), t \ge 0$, say, the notions of left and right-continuous filtrations correspond to the same sample path behaviors.

Note that given any filtration $\{\mathcal{F}_t : t \in S\}$ on an interval S, the filtration $\{\mathcal{F}_{t+}\}$ (or, $\{\mathcal{F}_{t-}\}$) is right– (resp., left–) continuous (Exercise 5).

Theorem 13.6 *(Regularization of Submartingales).* For $0 < s \le \infty$, let $S = [0, s]$, and let $\{X_t : t \in S\}$ be a submartingale or supermartingale with respect to an increasing family of sigmafields $\{\mathcal{F}_t\}$. Suppose that $\{X_t : t \in S\}$ is continuous in probability at each $t \ge 0$. Then there is a stochastic process $\{\tilde{X}_t : t \in S\}$ such that:

a. (*Stochastic Equivalence*) $\{\tilde{X}_t\}$ is a version of $\{X_t\}$ in the sense that $P(X_t = \tilde{X}_t) = 1$ for each $t \ge 0$.
b. (*Sample Path Regularity*) With probability 1 the sample paths of $\{\tilde{X}_t\}$ are bounded on compact intervals $a \le t \le b$, $(a, b \ge 0)$, and are right-continuous and have left-hand limits at each $t > 0$.
c. $\{\tilde{X}_t : t \in S\}$ is a $\{\mathcal{F}_{t+}\}$-submartingale.

Proof. Fix $0 < T < s$ and let Q_T denote the set of rational numbers in $[0, T]$. Write $Q_T = \bigcup_{n=1}^\infty R_n$, where each R_n is a *finite* subset of $[0, T]$ and $s \in R_1 \subset R_2 \subset \cdots$. By Doob's maximal inequality we have

$$P\left(\max_{t\in R_n} |X_t| > \lambda\right) \le \frac{\mathbb{E}|X_T|}{\lambda}, \qquad n = 1, 2, \ldots.$$

Therefore,

$$P\left(\sup_{t\in Q_T} |X_t| > \lambda\right) \le \lim_{n\to\infty} P\left(\max_{t\in R_n} |X_t| > \lambda\right) \le \frac{\mathbb{E}|X_T|}{\lambda}.$$

In particular, the paths of $\{X_t : t \in Q_T\}$ are bounded with probability 1. Let (c, d) be any interval in $\mathbb{R}$ and let $U^{(T)}(c, d)$ denote the number of upcrossings of (c, d) by the process $\{X_t : t \in Q_T\}$. Then $U^{(T)}(c, d)$ is the limit of the number $U^{(n)}(c, d)$ of upcrossings of (c, d) by $\{X_t : t \in R_n\}$ as $n \to \infty$. By Doob's upcrossing inequality one has

$$\mathbb{E}U^{(n)}(c,d) \le \frac{\mathbb{E}|X_T| + |c|}{d-c}.$$

Since, the $U^{(n)}(c,d)$ are non-decreasing with n, it follows that $U^{(T)}(c,d)$ is a.s. finite. Taking unions over all (c,d), with c,d rational, it follows that with probability one $\{X_t : t \in Q_T\}$ has only finitely many upcrossings of any interval. In particular, therefore, left- and right-hand limits must exist at each $t < T$ a.s. To construct a right-continuous version of $\{X_t : t \le T\}$, define $\tilde{X}_t = \lim_{r\to t^+, r\in Q_T} X_r$ for $t < T$. That $\{\tilde{X}_t : t \le T\}$ is in fact stochastically equivalent to $\{X_t : t \le T\}$ now follows from continuity in probability; i.e., $\tilde{X}_t = \lim_{r\to t^+} X_r = X_t$ since a.s. limits and limits in probability must a.s. coincide. Since $\tilde{X}_t = X_t$ a.s., to prove (c) it is sufficient to note that $\tilde{X}_t$ is $\mathcal{F}_{t+}$ measurable since X_{t+} is measurable with respect to $\mathcal{F}_{t+\frac{1}{n}}$ for all $n = 1, 2, \ldots$. Since T is arbitrary, the theorem follows. ■

Exercises

1. Let $\tau_1 \le \tau_2$ be two $\{\mathcal{F}_t : t \in T\}$–stopping times with values in $[0,\infty]$ interval. (a) Prove that $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$. (b) Prove that a $\{\mathcal{F}_t\}$-stopping time is an $\{\mathcal{F}_t\}$-optional time.
2. (i) Prove that τ_B defined in (13.4) is a $\{\mathcal{F}_t\}$-stopping time for a continuous parameter stochastic process $t \to X_t (t \in [0,\infty))$ if X_t takes values in a metric space (S,ρ) and $t \to X_t$ is continuous.
   (ii) Prove that if $t \to X_t$ is right-continuous with values in (S,ρ) and B is open then τ_B is an optional time. [*Hint*: (i) $[\tau_B \le t] = \cap_{n\in\mathbb{N}} \cup_{r\in Q\cap[0,t]} [\rho(X_r, B) \le \frac{1}{n}]$, where Q is the set all rational numbers. (ii) $[\tau_B < t] = \cup_{r\in Q\cap(0,t)} [X_r \in B]$.]
3. Check that the hypothesis of Theorem 13.3 holds in the derivations of (13.15)–(13.17).
4. (i) Prove that if τ is an optional time with respect to a filtration $\{\mathcal{F}_t : 0 \le t < \infty\}$ then it is a stopping time with respect to the filtration $\{\mathcal{G}_t : 0 \le t < \infty\}$, where $\mathcal{G}_t = \mathcal{F}_{t^+}$.
   (ii) Prove that τ_B is a $\{\mathcal{F}_{t+}\}$-stopping time if B is an open or closed, provided that $t \to X_t$ is continuous with values in a metric space.
5. Given a filtration $\{\mathcal{F}_t : t \in S\}$, S an interval, prove that $\{\mathcal{F}_{t+}\}$ is right–continuous.
6. Let f be a real–valued function on $Q \cap [0,\infty)$ which has no discontinuities of the second kind, and which is bounded on compacts. Define $g(t) := \liminf f(s)$ as $s \downarrow t$. Show that
   (a) g is right–continuous on $[0,\infty]$,
   (b) g has left–hand limits, and

(c) g is bounded on compacts.

7. Let $\{X_t : t \in T\}$ be a continuous $\{\mathcal{F}_t : t \in T\}$–martingale on $T = [0, \infty)$, $X_0 = x$ a.s. Suppose $c < x < d$ and $\tau := \inf\{t \geq 0 : X_t \in \{c, d\}\} < \infty$ a.s. Prove $P(\{X_t : t \in T\}$ reaches d before $c) = \frac{x-c}{d-c}$.
8. Let $\{X_t : t \in T\}$ be a stochastic process on $(\Omega, \mathcal{F})$ with values in some measurable space $(S, \mathcal{S})$, T a discrete set with elements $t_1 < t_2 < \cdots$. Define $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t) \subset \mathcal{F}, t \in T$. Assume that τ is an $\{\mathcal{F}_t : t \in T\}-$stopping time and show that $\mathcal{F}_\tau = \sigma(X_{\tau\wedge t} : t \in T)$; i.e., $\mathcal{F}_\tau$ is the $\sigma-$field generated by the *stopped process* $\{X_{\tau\wedge t} : t \in T\}$.
9. Show that the p-th variation of Brownian motion paths is a.s. infinite for $p < 2$, and zero for $p > 2$.
10. Suppose $\{G(t) : t \geq 0\}$ is a process with stationary independent increments having right-continuous sample paths with $G(0) = 0$. Assume that $\mathbb{E}G(t) = ct, t \geq 0$, with $c > 0$ and $\mathbb{E}e^{-qG(t)} < \infty$ for some $q > 0$. Show

    (i) $\tilde{m}(q) := \frac{1}{t}\log\mathbb{E}e^{-qG(t)}, \quad t \geq 0$, does not depend on t. [*Hint*: Exploit stationarity and independence of increments since for fixed $q > 0$ $f(t + s) := \mathbb{E}e^{-qG(t+s)} = \mathbb{E}e^{-q(G(t+s)-G(s))}e^{-q(G(s)-G(0))} = f(t)f(s), s, t \geq 0$. Exploit log-linearity of the positive right-continuous function f on $[0, \infty)$.]
    (ii) $P(u + G(t) < 0$ for some $t > 0) \leq e^{-Ru}$ where $R = \sup\{q : m(q) \geq 0\}$.[*Hint*: Use the optional stopping theorem.]
    (iii) If $\{G(t) : t \geq 0\}$ has continuous sample paths then one gets the equality $P(u + G(t) < 0$ for some $t > 0) = e^{-Ru}$. [*Hint*: The super-martingale is actually a martingale when $q = R$.]

11. Consider the Brownian motion $\{Y_t^x : t \geq 0\}$ having nonzero drift μ and diffusion coefficient $\sigma^2 > 0$, starting at x.

    (i) Show that $Z_t = \exp\{\xi(Y_t - t\mu) - \frac{\xi^2\sigma^2}{2}t\}, t \geq 0$, is a martingale.
    (ii) Verify the hypothesis of Theorem 13.3 for the validity of (13.13)–(13.17).

Chapter 14
Growth of Supercritical Bienaymé–Galton–Watson Simple Branching Processes

In this chapter the goal is to provide a precise calculation of the rate of growth of a branching process under a mild moment condition on the offspring distribution. The method illustrates the use of the size-bias change of measure technique for an alternative model of branching processes that permits more detailed description of the genealogy. For a given single progenitor the tree graph structure then makes it possible to uniquely trace any given progeny to its root. A natural distance between trees is also introduced that makes the collection of all such family trees a complete and compact metric space.

Let us recall some general facts about the Bienaymé–Galton–Watson simple branching process. Starting from a single progenitor $X_0 = 1$, let $X_1, \dots$ be the successive generation sizes having offspring distribution $f(k), k = 0, 1, \dots,$ with positive finite mean

$$0 < \mu := \sum_{k=0}^{\infty} k f(k) < \infty. \tag{14.1}$$

Assume throughout, but without further mention, that $0 < f(0) + f(1) < 1$ so to avoid considerations of special degeneracies. Letting $\{L_j^{(n)} : j, n \geq 1\}$ be a collection of i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, P)$ distributed according to f, one may define $\{X_n : n \geq 0\}$ according to the stochastic recursion (beginning with $X_0 = 1$),

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_14

$$X_{n+1} = \begin{cases} \sum_{j=1}^{X_n} L_j^{(n+1)}, & \text{if } X_n \geq 1, \\ 0 & \text{if } X_n = 0. \end{cases}$$

Let us also recall that the event $A := [\lim_{n\to\infty} X_n = 0]$ of *extinction* has a probability $\rho = P(A)$ which may be computed as the smallest fixed point of the probability generating function $\hat{f}(s) = \sum_{k=0}^{\infty} s^k f(k)$ of the offspring distribution. In particular $\rho = 1$ in the *subcritical* ($\mu < 1$) and in the *critical* ($\mu = 1$) cases. However, in the *supercritical* case ($\mu > 1$) one has $0 \leq \rho < 1$. Also notice that since the generation sizes X_n are non-negative integer valued, the *extinction time* $T = \inf\{n : X_n = 0\}$ must be finite on the event A as well. Recall the following simple observation pertaining to the mean rate.

Proposition 14.1. $\lim_n \frac{X_n}{\mu^n} = X_\infty$ exists a.s. where $P(0 \leq X_\infty < \infty) = 1$, *and* $\mathbb{E}X_\infty \leq 1$ if $X_0 = 1$.

Proof. Simply recall that $\{\frac{X_n}{\mu^n} : n \geq 0\}$ is a non-negative martingale and apply the martingale convergence theorem, together with Corollary 12.4. ■

Notice that this Proposition 14.1 also provides an alternative proof that $\rho = 1$ in the subcritical case $\mu < 1$ since $\mu^{-n} \to \infty$ in this case and the indicated limit must exist a.s. In any case, as already established in Theorem 9.1 for $\mu \leq 1$, one has

$$P(X_\infty = 0) = 1 = \rho \quad \text{for} \quad \mu \leq 1. \tag{14.2}$$

The main goal of this chapter is to determine the extent to which the mean number of offspring determines the rate of population growth $X_0, X_1, \ldots$ on the event A^c of *non-extinction* of a supercritical process. Our goal is to prove the following cornerstone result of the theory. We let L denote a generic random variable having the given offspring distribution.

***Theorem 14.2** (Kesten–Stigum).* Assume $\mu > 1$, and define X_∞ according to Proposition 14.1. The following are equivalent: (a) $P(X_\infty = 0) = \rho$; (b) $\mathbb{E}X_\infty = 1$; (c) $EL\log^+ L < \infty$.

Accordingly, the $L\log L$ condition (c) on the offspring distribution is, by (a), the condition under which the almost sure growth rate is precisely given by μ, and by (b) the condition for uniform integrability of the sequence $\{X_n/\mu^n : n \geq 1\}$. The proof will rest on martingale theory and a size-biased change of measure.[1]

In order to facilitate this approach we require a more detailed representation of the branching process as a probability distribution on the space Ω of family

[1]The use of this approach was noticed by Lyons et al. (1995). As a proof technique it has also come to be known as the *distinguished path technique*, and has also had applications in a number of other contexts, such as multiplicative cascades, tree polymers, coalescence, branching random walk, and branching Brownian motion, that will be treated as special topics in Chapter 21 of this text. Also see Arratia et al. (2019).

trees furnished by the *genealogy (family trees)* of the initial individuals in a branching process. Thus we will first develop another model (or probability space) for branching processes in which the possible outcomes are coded by the possible family trees. To describe the family trees corresponding to the branching process $\{X_n : n \geq 0\}$ defined in the previous chapter by (9.1), suppose that initially there are $X_0 = j$ individuals, encoded from left to right as $\langle 1\rangle, \langle 2\rangle, \dots, \langle j\rangle$, and referred to as *roots* or *progenitors*. Let $L_{\langle i\rangle}$ denote the number of offspring of individual $\langle i\rangle$, $i = 1, \dots, j$, and encode the $L_{\langle i\rangle}$ offspring comprising the first generation of $\langle i\rangle$ as $\langle i1\rangle, \langle i2\rangle, \dots, \langle iL_{\langle i\rangle}\rangle$ if $L_{\langle i\rangle} \geq 1$. If $L_{\langle i\rangle} = 0$, then the family tree of $\langle i\rangle$ stops there. One now proceeds recursively. Let $\{L_{\langle v\rangle} : v \in \cup_{n=1}^{\infty}\mathbb{N}^n\}$ be a collection of i.i.d. random variables distributed as $P(L_{\langle v\rangle} = k) = f(k)$, $k = 0, 1, 2, \dots$. If $\langle iv\rangle \in \mathbb{N}^{n+1}$ is an n-th generation member of the family tree of $\langle i\rangle$ and $L_{\langle iv\rangle} \geq 1$, then the $(n+1)$st generation offspring of $\langle iv\rangle$ are encoded[2] as $\langle iv1\rangle, \langle iv2\rangle, \dots, \langle ivL_{\langle iv\rangle}\rangle$. Thus the *family tree* $\tau(\langle i\rangle)$ of the initial $(n = 0)$ generation progenitor $\langle i\rangle$ is a set of finite sequences $v \in \cup_{n\geq 1}\mathbb{N}^n$, to be referred to as *vertices*, such that:

a. $\langle j\rangle \in \tau(\langle i\rangle)$ iff $j = i$, for $i, j \in \mathbb{N}$.
b. If $\langle vk\rangle \in \tau(\langle i\rangle)$, then $\langle vj\rangle \in \tau(\langle i\rangle)$ for $j \leq k$, $j, k \in \mathbb{N}$.
c. If $\langle vk\rangle \in \tau(\langle i\rangle)$, then $\langle v\rangle \in \tau(\langle i\rangle)$, $i, k \in \mathbb{N}$.

Condition (a) identifies $\langle i\rangle$ as the progenitor (root) of the family tree $\tau(\langle i\rangle)$. Condition (b) defines the left to right orientation of the (lexicographic) code, and condition (c) defines the orientation with respect to successive generations. The *generation height* of a root vertex is $|\langle i\rangle| = 0$, and the height of a subsequent offspring $v = \langle ii_1 \dots i_n\rangle$ is denoted $|v| \equiv |\langle ii_1 \dots i_n\rangle| = n$.

By assigning an *edge* between any pair of vertices $u, v \in \tau(\langle i\rangle)$ such that either $u = \langle vj\rangle$ or $v = \langle uj\rangle$ for some j, the set $\tau(\langle i\rangle)$ is a connected graph with no cycles (i.e., a *tree graph*). Apart from the progenitors, the vertices represent individual offspring as depicted in Figure 14.1. The *forest* $\tau(\langle i\rangle)$, $i = 1, 2, \dots, X_0$, depicts the overall histories of an initially prescribed X_0 individuals. Note that each offspring vertex of the tree may be uniquely associated with its parental edge. We complete this picture by assigning a "ghost" edge to each progenitor $\langle i\rangle$.

One may now view the *Bienaymé–Galton–Watson simple branching model started from a single progenitor* as a probability distribution P on the space Ω of family trees τ rooted at $\langle 1\rangle \in \tau$ as defined above and equipped with a Borel $\sigma-$field $\mathcal{B}$ as follows: Let d denote the non-negative function on $\Omega \times \Omega$ defined (with the conventions $\inf\emptyset = \infty$, and $\frac{1}{1+\infty} = 0$) by

$$d(\tau, \eta) = \frac{1}{1 + \inf\{k : \tau|k \neq \eta|k\}}, \qquad \tau, \eta \in \Omega, \tag{14.3}$$

[2]This is often referred to as the Harris-Ulam labeling.

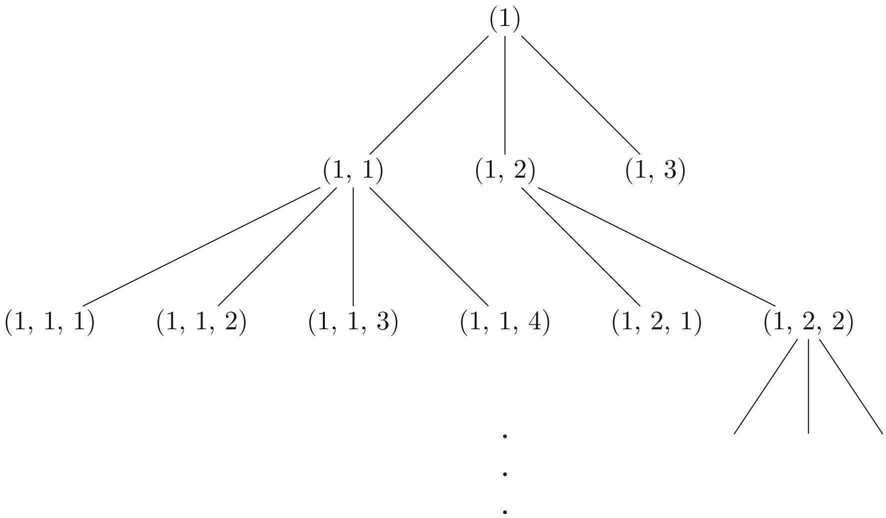


Fig. 14.1 Bienaymé–Galton–Watson Tree Graph

where $\tau|0 = \eta|0 = \langle 1\rangle$, $\tau|k$, $\eta|k$ denote the subtrees of τ, η, comprising the vertices present in their respective first k generations, i.e., $\tau|k = \{\langle 1v_1 \cdots v_j\rangle \in \tau : j \leq k\}, k \geq 1$. That is, the distance between two trees τ and η is measured in terms of shortest length of the common ancestral lines originating with the common progenitor. Then one may check that d defines a metric which makes Ω a complete and separable compact metric space (Exercise 1). The open ball centered at $\tau \in \Omega$ of radius $\frac{1}{n+1}$ is given by

$$B_{\frac{1}{n+1}}(\tau) = \{\eta \in \Omega : \eta|n = \tau|n\}. \tag{14.4}$$

For emphasis, note again that the set $B_{\frac{1}{n+1}}(\tau)$ only specifies the structure of the tree in the first n generations which, in turn, only specifies the offspring of vertices belonging to the first $n-1$ generations. In particular, $B_1(\{\langle 1\rangle\}) = \Omega$, and the open ball $B_{\frac{1}{n+1}}(\tau)$ of radius $\frac{1}{n+1}$ centered at τ defines the finite-dimensional event $[\tau]_n$ that the first n generations of the tree are given by $\tau|n$. In this framework

$$X_n(\tau) = \begin{cases} \mathrm{card}\{\langle 1\rangle\} = 1, & \text{if } n = 0, \\ \mathrm{card}\{v \in \tau : \langle 1v_1 \dots v_n\rangle \in \tau\} & \text{if } n \geq 1. \end{cases}$$

Given an offspring probability mass function $f(k) : k = 0, 1, \dots$ the probability measure P is uniquely defined on $\mathcal{B}$ (Exercise 2) by prescribing

$$P(B_1(\{\langle 1\rangle\})) = 1, \quad P(B_{\frac{1}{n+1}}(\tau)) = \prod_{v \in \tau|(n-1)} f(\#v), \quad n \geq 1, \tag{14.5}$$

where the indicated product for $n \geq 1$ is over all vertices in the first $n-1$ generations of τ and $\#v$ denotes the number of offspring of $v = \langle 1v_1 \cdots v_k \rangle$ in τ, i.e., $\#v$ is the cardinality of $\{j \in \mathbb{N} : \langle 1v_1 \cdots v_k j \rangle \in \tau\}$. Also the product is restricted to $n-1$ since the numbers of $n-$generation offspring is not specified by $B_{\frac{1}{n+1}}(\tau)$. In view of the following proposition, the successive generation sizes $X_n, n \geq 0$, of Chapter 9 are readily identified within this framework.

Proposition 14.3. For a given offspring distribution $f(k) \geq 0, \sum_{k=0}^{\infty} f(k) = 1$, let $(\Omega, \mathcal{B}, P)$ be defined by prescribing probabilities according to (14.5). The stochastic process of successive generation sizes defined by $X_0(\tau) = 1, \tau \in \Omega$, and $X_n(\tau) = \text{card}\{v : |v| = n, \langle 1v_1 \cdots v_n \rangle \in \tau\}, \tau \in \Omega, n = 1, 2 \ldots$, is a Markov chain on the state space of non-negative integers with initial distribution $P(X_0 = 1) = 1$ and one-step transition probabilities $P(X_{n+1} = j | X_n = i) = f^{*i}(j), i, j = 0, 1, 2, \ldots$.

Proof. The assertion is that

$$P(X_0 = 1, \ldots, X_n = i_n) = P(X_0 = 1, \ldots, X_{n-1} = i_{n-1}) f^{*i_{n-1}}(i_n). \tag{14.6}$$

First note that $P(X_0 = 1) = P(B_1(\{\langle 1 \rangle\})) = 1$, and

$$\begin{aligned} P(X_0 = 1, X_1 = i_1) &= P(B_{\frac{1}{2}}(\{\langle 1 \rangle, \langle 11 \rangle, \ldots, \langle 1i_1 \rangle\})) \\ &= f(\#\langle 1 \rangle) = f(i_1), i_1 = 0, 1, \ldots. \end{aligned} \tag{14.7}$$

Next, the event $[X_0 = 1, X_1 = i_1, X_2 = i_2]$ is a disjoint union of balls of radius $1/3$ defined by the integral partitions of i_2 into i_1 terms. Thus,

$$\begin{aligned} &P(X_0 = 1, X_1 = i_1, X_2 = i_2) \\ &= f(\#\langle 1 \rangle) \sum_{\#\langle 11 \rangle + \cdots + \#\langle 1i_1 \rangle = i_2} f(\#\langle 11 \rangle) f(\#\langle 12 \rangle) \cdots f(\#\langle 1i_1 \rangle) \\ &= f(i_1) f^{*i_1}(i_2). \end{aligned} \tag{14.8}$$

So the assertion holds for $n = 1, 2$. The result follows as in these cases since $[X_0 = 1, \ldots, X_n = i_n, X_{n+1} = i_{n+1}]$ is the disjoint union

$$\cup_{\tau \in T(1, i_1, \ldots, i_n)} \cup_\pi \{\eta \in \Omega : \eta|(n+1) = \tau_\pi|(n+1)\} = \cup_{\tau \in T(1, i_1, \ldots, i_n)} \cup_\pi B_{\frac{1}{n+2}}(\tau_\pi),$$

where $T(1, i_1, \ldots, i_n)$ is the (finite) collection of all n generation trees $\tau = \tau(1, i_1, \ldots, i_n)$ having the successive generation sizes $i_0 = 1, i_1, \ldots, i_n$ subject to coding rules (a)–(c), with no further offspring, and τ_π denotes a family tree extension of such a $\tau \in T(1, i_1, \ldots, i_n)$ to $n+1$ generations determined by integral partition π of i_{n+1} into i_n terms. That is, with $i_0 = 1$, and noting that $P(B_{\frac{1}{n+2}}(\tau_\pi)) = \prod_{v \in \tau_\pi | n} f(\#v)$,

$$
\begin{aligned}
&P(X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}) \\
&= \sum_{\tau \in T(1,i_1,\ldots,i_n)} \sum_{\pi} P(B_{\frac{1}{n+2}}(\tau_\pi)) \\
&= \sum_{\tau \in T(1,i_1,\ldots,i_n)} \prod_{v \in \tau|n-1} f(\#v) \sum_{\pi} \prod_{v \in \tau_\pi, |v|=n} f(\#v) \\
&= \sum_{\tau \in T(1,i_1,\ldots,i_n)} P(B_{\frac{1}{n+1}}(\tau)) f^{*i_n}(i_{n+1}) \\
&= P(X_0 = 1, X_1 = i_1, \ldots, X_n = i_n) f^{*i_n}(i_{n+1}).
\end{aligned}
$$

■

The event A^c of non-extinction corresponds to the occurrence of infinite trees τ, or equivalently, trees with at least one infinite *line of descent path* emanating from the root $\langle 1 \rangle$. Any non-degenerate asymptotic number $X_n(\tau)$ of progeny at level n will depend on the preponderance of such infinite paths. Thus we will analyze the distribution of trees from the perspective of a "randomly selected (distinguished) infinite path." This technique relies heavily on martingale convergence and, especially, the Lebesgue decomposition of measures given in Corollary 12.8, which the reader may wish to review at the outset.

For a tree graph structure, that is a connected graph without loops rooted at $\langle 1 \rangle$, any vertex $v = \langle 1v_1 \ldots v_m \rangle$ defines a unique path of vertices $\langle 1 \rangle, \langle 1v_1 \rangle, \ldots, \langle 1v_1 \ldots v_m \rangle$ between itself and the root. Similarly, an infinite line of descent path from the root $\langle 1 \rangle$ of an infinite tree may be defined by an infinite sequence $\gamma = \langle 1\gamma_1\gamma_2 \ldots \rangle$, $\gamma_j \in \mathbb{N}$, $j \geq 1$. The intuitive idea is to first enlarge the probability model $(\Omega, \mathcal{B}, P)$ to encode trees with infinite line of descent paths as pairs (τ, γ), where $\tau \in \Omega$ is an infinite tree having an infinite line of descent path γ, i.e., $\langle 1\gamma_1\gamma_2 \ldots \gamma_m \rangle \in \tau$ for all $m \geq 1$, denoted by a slight abuse of notation as $\gamma \in \tau$. The Figure 14.2 is an aid to understanding this enlargement.

Let Ω_∞ denote the subspace of Ω consisting of infinite trees. Then Ω_∞ is a closed set for the metric d given by (14.3) and therefore Borel measurable (Exercise 4). Let $\Gamma := \{1\} \times \mathbb{N}^{\mathbb{N}}$ be the collection of all infinite paths emanating from the root $\langle 1 \rangle$. Let Ω^* denote the space of tree-path pairs (τ, γ) where $\tau \in \Omega_\infty$, $\gamma \in \tau$. The $\sigma-$field $\mathcal{B}^*$ will denote the $\sigma-$field of subsets of $\mathcal{B}^*$ generated by finite-dimensional events of the form

$$[(\tau, \gamma)]_n := \{(\eta, \alpha) \in \Omega^* : \eta|n = \tau|n, \alpha|n = \gamma|n\}, \tag{14.9}$$

where $\gamma|n = \langle 1\gamma_1 \ldots \gamma_n \rangle$, $\gamma \in \Gamma$. The probabilities assigned to such an extension are intuitively a manifestation of the conditional probability formula

$$\text{“prob}(\tau, \gamma) = \text{prob}(\tau|\gamma)\text{prob}(\gamma)''. \tag{14.10}$$

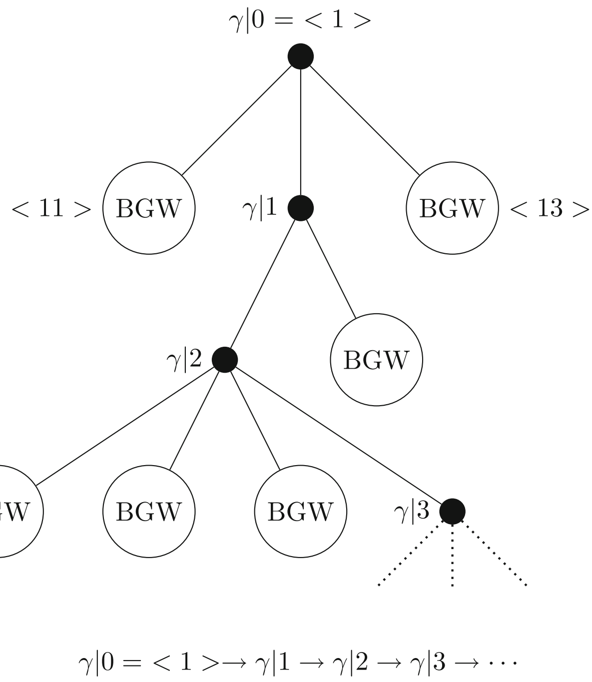


Fig. 14.2 Size-biasing along distinguished path

Considering that the infinite paths through a given tree are statistically indistinguishable under the definition (14.5), paths should be "selected uniformly in trees." However, conditioning on the existence of an infinite line of descent will re-weight the probability (14.5). Observe that for a specified tree τ having $X_1(\tau) = k$ first generation offspring, there will be k corresponding subtrees $\tau^{(j)}$ rooted at $\langle 1j \rangle$, $j = 1, \ldots, k$. A given path $\gamma \in \tau$ will have $\gamma \in \tau^{(j)}$ for one and only one j, say $j = l$. According to the model (14.5) one has for such fixed τ,

$$P([\tau]_{n+1}) = f(k) \prod_{j=1}^{k} P([\tau^{(j)}]_n) = kf(k) \frac{1}{k} P([\tau^{(l)}]_n) \prod_{j \neq l}^{k} P([\tau^{(j)}]_n). \tag{14.11}$$

Multiplying this equation by $\frac{1}{\mu^{n+1}}$ and defining

$$\mathcal{Q}([\tau, \gamma]_n) := \frac{P([\tau]_n)}{\mu^n}, \tag{14.12}$$

one has the consistent definition of a probability $\mathcal{Q}$ on Ω^* such that

$$\mathcal{Q}([\tau,\gamma]_{n+1}) = \frac{kf(k)}{\mu}\frac{1}{k}\mathcal{Q}([\tau^{(l)},\gamma]_n)\prod_{j\neq l}^{k} P([\tau^{(j)}]_n). \tag{14.13}$$

Note that $\tilde{f}(k) = kf(k)/\mu, k = 0, 1, 2, \ldots,$ is a new offspring probability distribution for production along the randomly selected path γ, while off the path offspring are generated according to f. In particular, along the path γ $\tilde{f}(0) = 0$, so that at least one offspring is sure to occur at each generation along the path. Now, since each n-th generation vertex defines a unique path to the root, summing over the paths γ such that $\gamma|n \in \tau|n$ in (14.12), yields the marginal P^* distribution is given by

$$P^*([\tau]_n) = \frac{X_n(\tau)}{\mu^n}P([\tau]_n) \equiv Q_n(\tau)P([\tau]_n), \tag{14.14}$$

where $Q_n(\tau) = \frac{X_n(\tau)}{\mu^n}, \tau \in \Omega_\infty$. In particular P^* is absolutely continuous with respect to P on each $\sigma-$field $\mathcal{F}_n$ with Radon-Nikodym derivative $Q_n(\tau)$, i.e.,

$$\frac{dP^*}{dP} = Q_n \quad \text{on } \mathcal{F}_n, n = 0, 1, 2, \ldots. \tag{14.15}$$

The probabilities $\frac{kf(k)}{\mu}, k = 1, 2, \ldots$ comprise the size-biased offspring distribution, and P^* is naturally referred to as the corresponding *size-bias change of the probability P*.

Example 1. Consider the Bernoulli offspring distribution $1 - f(0) = f(2) = p$, $p > 1/2$, with mean $\mu = 2p$. Then the size-biased offspring distribution is deterministic $\hat{f}(k) = kf(k)/\mu = \delta_{2,k}, k \geq 0$. Thus, two offspring are produced at each generation along a distinguished path, while off the path the offspring either doubles or dies with probabilities $p, q = 1-p$, respectively. In particular, since the root belongs to every path, the process begins with two offspring of the root. This may be verified from (14.14) since for $[\tau]_1 = \{\langle 1\rangle, \langle 11\rangle, \langle 12\rangle\}$ one has $P^*([\tau]_1) = \frac{X_1(\tau)}{2p}P([\tau]_1) = \frac{2}{2p}p = 1$. For $[\tau]_2 = \{\langle 1\rangle, \langle 11\rangle, \langle 12\rangle\langle 111\rangle, \langle 112\rangle\}$, $P^*([\tau]_2) = \frac{X_2(\tau)}{(2p)^2}P([\tau]_2) = \frac{2}{4p^2}p^2q = q/2$. Similarly, for $[\tau]_2 = \{\langle 1\rangle, \langle 11\rangle, \langle 12\rangle\langle 121\rangle, \langle 122\rangle\}$, while for $[\tau]_2 = \{\langle 1\rangle, \langle 11\rangle, \langle 12\rangle, \langle 111\rangle, \langle 112\rangle, \langle 121\rangle, \langle 122\rangle\}$, one has $P^*([\tau]_2) = \frac{4}{4p^2}p^3 = p$. In particular, these three events exhaust the total P^* probability, i.e., $p + q/2 + q/2 = 1$.

A major mathematical advantage of the size-bias change of probability is the induced exchange of the determination of "vanishing/non-vanishing" of limits with that of "non-finiteness/finiteness"; i.e., notice that $dP/dP^* = 1/Q_n$ on $\mathcal{F}_n$. This reasoning is made precise using the *Lebesgue decomposition* of measures given in Corollary 12.8. In particular, taking $m = P^*$ and $q = P$ there, we have $P^* << P$ on each $\mathcal{F}_n$ with Radon-Nikodym derivative $Q_n = dP^*/dP = \frac{X_n}{\mu^n}$ on $\mathcal{F}_n$. As per Corollary 12.8 define

$$Q_\infty(x) = \limsup_{n\to\infty} Q_n(x), \quad x \in S.$$

Then, in accordance with the Lebesgue decomposition into its absolutely continuous and singular components,

$$P^*(A) = \mathbb{E}_P(\mathbf{1}_A Q_\infty) + P^*(A \cap [Q_\infty = \infty]), \quad A \in \mathcal{F}.$$

In particular

a. $P^* << P \iff P^*([Q_\infty = \infty]) = 0 \iff \mathbb{E}_P Q_\infty = P^*(\Omega) \quad (= 1)$
b. $P^* \perp P \iff P^*([Q_\infty < \infty]) = 0 \iff \mathbb{E}_P Q_\infty = 0.$

Thus, as noted earlier, there is a trade between 0 and ∞ such that $Q_\infty = 0$, P-a.s. if and only if $Q_\infty = \infty$, P^*-a.s., and $Q_\infty > 0$ with positive P-probability if and only if $Q_\infty < \infty$ with positive P^*-probability. So the task for the proof of the Kesten–Stegum theorem is to determine when $Q_\infty = \lim_n \frac{X_n}{\mu^n}$ is $P^* - a.s.$ finite or not. For this we employ the following simple *first departure bounds* based on the departures along a given line of descent path into branching subtrees; also referred to as a *spine decomposition*.

Proposition 14.4. For any $(\tau, \gamma) \in \Omega^*$

$$\frac{X_{n,n-1}(\tau)}{\mu^n} \le \frac{X_n(\tau)}{\mu^n} \le \frac{1}{\mu^n} + \sum_{j=1}^{n-1} \frac{X_{n,j}(\tau)}{\mu^{n-j}} \frac{1}{\mu^j},$$

where $X_{n,j}(\tau)$ is the number of offspring descendants of $\langle 1\gamma_1 \ldots \gamma_j \rangle$ in the $n-$th generation of τ.

Proof. The lower bound is obvious from the definition of X_n since every path to the $n-$th generation must pass through the $(n-1)-$st generation. For the upper bound, any path to the $n-$th generation must either coincide with γ or have a height $1 \le j \le n-1$ of first departure. ■

We may now prove the Kesten–Stigum Theorem 14.2:

Proof. Assume mean offspring $\mu > 1$ and define X_∞ according to Proposition 14.1. Suppose first that $\mathbb{E}_{P^*} \log L \equiv \frac{1}{\mu}\mathbb{E}_P L \log^+ L = \infty$. In particular $X_{n,n-1}, n \ge 1$, are i.i.d. under P^* with infinite mean. Thus

$$\limsup_n \frac{X_{n,n-1}}{\mu^n} = \limsup_n e^{\log \frac{X_{n,n-1}}{\mu^n}} = \limsup_n e^{n(\frac{\log X_{n,n-1}}{n} - \log\mu)} = \infty \tag{14.16}$$

since, by the Borel–Cantelli lemma, $\limsup \frac{\log X_{n,n-1}}{n} = \infty$, P^*-a.s. (Exercise 5). Thus by the lower bound in the first departure decomposition, it follows that $Q_\infty = \infty$ $(P^*-$a.s.), i.e., $P^*([Q_\infty < \infty]) = 0$. In particular, therefore, by the Lebesgue decomposition it follows that $\mathbb{E}_P X_\infty \equiv \mathbb{E}_P Q_\infty = 0$. Conversely suppose that $\mathbb{E}_{P^*} \log L \equiv \frac{1}{\mu}\mathbb{E}_P L \log^+ L < \infty$. Let us check that the upper bound sum $\sum_{j=0}^{n-1} \frac{X_{n,j}}{\mu^n}$ is conditionally under Q given $\sigma(L_{\gamma|j}, j \geq 0)$ a non-negative submartingale with bounded expectation and therefore (conditionally) a.s. convergent. To see the martingale property note that for each fixed j, $(1 \leq j \leq n-1)$, under Q, $\mathbb{E}(\frac{X_{n+1,j}}{\mu^{n+1}}|\mathcal{F}_n^*, \gamma) = \frac{X_{n,j}\mu}{\mu^{n+1}} = \frac{X_{n,j}}{\mu^n}$. $\mathbb{E}(\frac{X_{j+1,j}}{\mu^{j+1}}|\mathcal{F}_j, \gamma) = \frac{L_{\gamma|j}}{\mu^j}$ as well. In fact, $\frac{X_{n,j}}{\mu^{n-j}}, n \geq j+1$, is the non-negative martingale associated with a Bienaymé–Galton–Watson process having offspring distribution f and $L_{\gamma|j}$ initial progeny under Q given $\sigma(L_{\gamma|j} : j \geq 0)$. Thus

$$\mathbb{E}_Q\left\{\sum_{j=0}^{n-1} \frac{X_{n,j}}{\mu^{n-j}}\frac{1}{\mu^j}|\sigma(L_{\gamma|k}, k \geq 0)\right\} = \sum_{j=1}^{n-1} \frac{L_{\gamma|j}}{\mu^j}.$$

Now using the condition $\mathbb{E}_{P^*} \log L < \infty$ and Borel–Cantelli lemma it follows (Exercise 5) that $\sum_{j=1}^{n-1} \frac{L_{\gamma|j}}{\mu^j}$ converges P^*-a.s. to a finite non-negative limit; i.e., the conditional expectations are (conditionally) a.s. bounded by $\sum_{j=1}^{\infty} \frac{L_{\gamma|j}}{\mu^j} < \infty$, as asserted. Now, taking expectations,

$$\mathbb{E}_{P^*}\sum_{j=0}^{n-1} \frac{X_{n,j}}{\mu^n} < \infty.$$

Thus P^*-a.s. $Q_\infty < \infty$, i.e., $P^*([Q_\infty = \infty]) = 0$. So, using the Lebesgue decomposition, we have that $\mathbb{E}_P X_\infty \equiv \mathbb{E}_P Q_\infty = 1$. ■

Corollary 14.5. Assume $\mu > 1$, $X_0 = 1$, as well as the $\mathbb{E}L\log^+ L < \infty$ condition of the Kesten-Stigum theorem. Then X_∞ solves the stochastic fixed point equation

$$X_\infty =^{\text{dist}} \frac{1}{\mu}\sum_{j=1}^{X_1} X_\infty^{(j)},$$

where $X_\infty^{(j)}$, $j = 1, 2, \ldots$, are independent of X_1 and distributed as X_∞. Let $\gamma(t) = \mathbb{E}e^{-tX_\infty}$, $t \geq 0$. Then γ is the unique solution to the functional equation

$$\gamma(\mu t) = \hat{f} \circ \gamma(t), \ |\gamma'(0^+)| = \mathbb{E}X_\infty < \infty \quad t \geq 0.$$

Proof. Both stochastic fixed point equation and the equation for the Laplace transform are obtained in the Kesten-Stigum limit from the recursion

$$\frac{X_{n+1}}{\mu^{n+1}} =^{\text{dist}} \frac{1}{\mu}\sum_{j=1}^{X_1} \frac{X_n^{(j)}}{\mu^n},$$

where $\{X_n^{(j)} : n \geq 0\}$, $j = 1, \ldots,$ are i.i.d. distributed as $\{X_n : n \geq 0\}$, $X_n^{(j)}, n \geq 0$, being the branching process originating from the j-th offspring of the root. Suppose that γ and $\tilde{\gamma}$ are solutions to the functional equation with $|\tilde{\gamma}'(0^+)| = |\gamma'(0^+)| < \infty$. Then, since $\frac{\hat{f}(t)-\hat{f}(s)}{t-s} \leq \hat{f}'(1) = \mu, 0 \leq s < t \leq 1$, by convexity, one has after iteration that

$$
\begin{aligned}
|\tilde{\gamma}(t) - \gamma(t)| &= |\hat{f} \circ \tilde{\gamma}(\frac{t}{\mu}) - \hat{f} \circ \gamma(\frac{t}{\mu})| \\
&\leq \mu|\tilde{\gamma}(\frac{t}{\mu}) - \gamma(\frac{t}{\mu})| \\
&\leq \cdots \leq \mu^n|\tilde{\gamma}(\frac{t}{\mu^n}) - \gamma(\frac{t}{\mu^n})| \\
&= \mu^n|\tilde{\gamma}(\frac{t}{\mu^n}) - \tilde{\gamma}(0) - (\gamma(\frac{t}{\mu^n}) - \gamma(0))| \\
&= t\frac{\left|\tilde{\gamma}(\frac{t}{\mu^n}) - \tilde{\gamma}(0) - (\gamma(\frac{t}{\mu^n}) - \gamma(0))\right|}{\frac{t}{\mu^n}}| \\
&\to t|\tilde{\gamma}'(0^+) - \gamma'(0^+)| = 0,
\end{aligned}
$$

in the limit as $n \to \infty$. This proves uniqueness. ■

Exercises

1. Show that the function d defined by (14.3) makes Ω a complete and separable metric space. [*Hint*: The set $D = \{\tau : ||\tau|| < \infty\}$ is countable and it is dense in Ω. If $\{\tau_k : k \geq 1\}$ is Cauchy, then for every $n = 0, 1, 2, \ldots, \tau_k|n = \tau_m|n$ for all sufficiently large m, k (depending on n).]
2. Show that (14.5) uniquely specifies the probability measure P on $\mathcal{B}$. [*Hint*: Apply the Caratheodory construction[3]]
3. (*Seneta's Theorem*) A Bienaymé–Galton–Watson branching process with immigration is defined as follows: Let $Y_1, Y_2, \ldots$ be i.i.d. non-negative integer-valued random variables. The branching process begins with the immigration of Y_1 initial progenitors and each independently of the others produces a random number of offspring according to the offspring distribution $f(k), k \geq 0$. The second generation consists of these offspring together with an additional independent number of Y_2 immigrants, and so on as defined by the stochastic recursion

[3]See BCPT pp. 225–228.

$$X_0 = Y_1, \quad X_{n+1} = \sum_{j=1}^{X_n} L_j^{(n+1)} + Y_{n+2}, n \geq 0,$$

where $\{L_j^{(n)} : j \geq 0, n \geq 1\}$ are i.i.d. distributed according to the offspring distribution f and independent of $Y_1, Y_2, \ldots$. Assume $\mu = \sum_{k=1}^{\infty} kf(k) > 1$. Show that if $E \log^+ Y_1 < \infty$, then $\lim_n \frac{X_n}{\mu^n}$ exists a.s. and is a.s. finite. Show that if $E \log^+ Y_1 = \infty$ then $\limsup_n \frac{X_n}{m^n} = \infty$ a.s. for any constant $m > 0$. [*Hint*: Use the size-bias change of measure to mimic the proof of the Kesten-Stigum theorem.]

4. Show that Ω_∞ is a closed set.
5. Show that if $L_1, L_2, \ldots$ is an i.i.d. sequence of non-negative random variables, then $\limsup_n \frac{L_n}{n}$ is a.s. either 0 or ∞ according to whether $\mathbb{E}L_1 < \infty$ or $\mathbb{E}L_1 = \infty$, respectively. [*Hint*: Use the Borel–Cantelli lemma.] Use this to show $\limsup_n \frac{X_{n,n-1}}{\mu^n} = \infty$, P^*-a.s. if $\mathbb{E}_P L \log^+ L = \infty$, and to show $\sum_{j=1}^{n-1} \frac{L_{\gamma|j}}{\mu^j}$ converges P^*-a.s. when $\mathbb{E}_P L \log^+ L < \infty$.
6. Suppose that the offspring distribution is given by the geometric distribution $f(k) = qp^k, k = 0, 1, 2 \ldots$ where $q = 1 - p, 0 < p < 1$. Show that

   (i) $\hat{f}(s) = \sum_{k=0}^{\infty} s^k f(k) = \dfrac{q}{1 - ps}$

   (ii) $\hat{f}^{o(n)}(s) = \dfrac{q\mu^n(1 - s) + ps - q}{p\mu^n(1 - s) + ps - q}$, $\mu = \frac{p}{q}$ where the composite function $\hat{f}^{o(n)}(s)$ is inductively defined $\hat{f}^{o(1)}(s) := \hat{f}(s)$, $\hat{f}^{o(n+1)}(s) := \hat{f}(\hat{f}^{o(n)}(s))$.

   (iii) $\mathbb{E}_{X_0=1}\, s^{X_n} = \hat{f}^{o(n)}(s)$.

   (iv) In the supercritical case $\mu > 1$ show that $\frac{X_n}{\mu^n}$ converges in distribution to a random variable Z_∞ exponentially distributed as $P(X_\infty > x) = (1 - \rho)e^{-(1-\rho)x}, \quad x > 0$, where $P(X_\infty = 0) = \rho = \frac{1-\sqrt{1-4pq}}{2p}$.

   (v) In the subcritical case $\mu < 1$, show that $P\,(X_n = k|X_n > 0) \longrightarrow (1 - \mu)\mu^{k-1}$.

7. Let $\psi = \hat{f}^{-1}$ on $[q, 1]$, and $\psi_n = \psi^{\circ n}$ the iterated compositions, with $\psi_0(s) = s$. Show that $M_n(s) = \psi_n(s)^{X_n}, n = 0, 1, 2, \ldots$, is a positive martingale for each $q \leq s < 1$, bounded by one.

Chapter 15
Stochastic Calculus for Point Processes and a Martingale Characterization of the Poisson Process

The main purpose of this chapter is to provide a martingale characterization of the Poisson process obtained in Watanabe (1964). This will be aided by the development of a special stochastic calculus[1] that exploits its non-decreasing, right-continuous, step-function sample path structure when viewed as a counting process; i.e., for which stochastic integrals can be defined in terms of standard Lebesgue integration theory.

We begin with two general definitions.

Definition 15.1. A non-negative integer-valued stochastic process $N = \{N(t); t \geq 0\}$, $N(0) = 0$, defined on a probability space $(\Omega, \mathcal{F}, P)$, whose sample paths are right-continuous step functions with unit jumps, is referred to as a *simple point process* or *counting process*

Remark 15.1. The terminology of "simple point process"places emphasis on the *points* of discontinuity at which the (unit) jumps occur in the counting process. Equivalently, one may view the counting process as a random measure with unit atoms at each point of discontinuity. More *general* notions of point processes permit integer jumps of magnitude greater than one. The mathematical representation in terms of random measures, even random linear functionals, finds numerous applications.[2]

[1]A more comprehensive treatment of point processes from a martingale perspective is given in Brémaud (1981).

[2]For example, see Le Cam (1960b), Sun and Stein (2015), and extensive list of the references therein.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_15

Definition 15.2. Let $\mathcal{F}_t, t \geq 0$ be a filtration on $(\Omega, \mathcal{F})$, and let λ be a non-negative measurable function on $[0, \infty)$. A *Poisson process with intensity function* λ *with respect to the filtration* $\mathcal{F}_t, t \geq 0$ is a simple point process N adapted to this filtration, such that $N(t) - N(s)$ is independent of $\mathcal{F}_s$ for each $0 \leq s < t$, and

$$P(N(t) - N(s) = n|\mathcal{F}_s) = \frac{(\int_s^t \lambda(r)dr)^n}{n!} \exp\{-\int_s^t \lambda(r)dr\}, n = 0, 1, 2, \ldots .$$

Observe that a sample path $t \to N(\omega, t), t \geq 0$, of a simple point process may be viewed as the distribution function of a measure, again denoted N, on $[0, \infty)$, with atoms at $0 < T_1 < T_2 < \cdots$. Thus, for each $\omega \in \Omega$ the integral $\int_0^t f(s)dN(s) \equiv \int_0^t f(s)N(ds)$ may be viewed sample pathwise as defined by a Lebesgue–Stieltjes integral[3] for non-negative, measurable (deterministic) integrands f, and more generally as usual for $f = f^+ - f^-$ in the case $f^\pm$ both have finite integrals. Of course this interpretation can be achieved whenever the sample paths of the integrator N are functions of bounded variation on bounded intervals. The measurability issues arise in connection with stochastic integrands f as will be defined below. However, it should be emphasized that the "stochastic calculus"developed here essentially depends only on standard Lebesgue integration concepts from real analysis. In particular, if N is a counting process with jumps at $0 < T_1 < T_2 \cdots$, and if $\varphi : [0, \infty) \to \mathbb{R}$ is an arbitrary left-continuous function, then

$$\int_0^t \varphi(s)dN(s) = \sum_{j=1}^{N(t)} \varphi(T_j), \quad t \geq 0. \tag{15.1}$$

The following definition is introduced for extensions to stochastic integrands with a goal of preserving certain martingale structure.

Definition 15.3. Given a filtration $\mathcal{F}_t, t \geq 0$, on $(\Omega, \mathcal{F})$, the σ-field of subsets of $(0, \infty) \times \Omega$ generated by sets of the form, $(s, t] \times A, 0 \leq s \leq t, A \in \mathcal{F}_t$, is referred to as the $\mathcal{F}_t$*−predictable* σ-field on $(0, \infty) \times \Omega$. A real-valued stochastic process $\varphi = \{\varphi(t) : t \geq 0\}$ such that $\varphi(0)$ is $\mathcal{F}_0$-measurable, and $(t, \omega) \to \varphi(t, \omega), t > 0, \omega \in \Omega$, is measurable with respect to the predictable filtration on $(0, \infty) \times \Omega$, is said to be an $\mathcal{F}_t$*−predictable process*.

Note that a predictable process would also be a progressively measurable process as defined by Definition 13.3. The following theorem is useful in keeping track of predictability.

Theorem 15.1. If a real-valued stochastic process φ, adapted to a filtration $\mathcal{F}_t, t \geq 0$, has left-continuous sample paths, then φ is $\mathcal{F}_t$-predictable on $(0, \infty) \times \Omega$.

[3]See BCPT p. 228, for Lebesgue–Stieltjes measure and integration.

Proof. For each $n = 1, 2, \ldots$, one has by left-continuity that $\varphi(t,\omega) = \lim_{n\to\infty} \varphi(2^{-n}[2^n t], \omega)$ for each $(t,\omega) \in (0,\infty) \times \Omega$. Moreover, $2^{-n}[2^n t] \le t$. Thus φ can be expressed as a point-wise limit of $\mathcal{F}_t$-predictable processes $\varphi_n(t,\omega) = \varphi(2^{-n}[2^n t], \omega)$. ■

Let us recall Definition 13.5 that a filtration $\mathcal{F}_t$, $t \ge 0$ is said to be *left-continuous* if

$$\mathcal{F}_t = \mathcal{F}_{t-} = \sigma(\cap_{s<t}\mathcal{F}_s,)\ t > 0, \quad \mathcal{F}_{0-} = \mathcal{F}_0.$$

In general, if N is a simple point process and φ is a $\mathcal{F}_t$-predictable process, then the stochastic differential equation

$$dX(t) = \varphi(t^-)dN(t),\ X(0) = x, \tag{15.2}$$

is defined by

$$X(t) = x + \int_0^t \varphi(s^-)dN(s) = x + \sum_{j=1}^{N(t)} \varphi(T_j^-). \tag{15.3}$$

Theorem 15.2. Suppose that $M = \{M(t) : t \ge 0\}$ is a martingale with respect to a filtration $\mathcal{F}_t, t \ge 0$, on $(\Omega, \mathcal{F}, P)$. Assume that the sample paths of M are of bounded variation on bounded intervals. If $\varphi = \{\varphi(t) : t \ge 0\}$ is a $\mathcal{F}_t$-predictable process such that $\int_0^t |\varphi(s)||M|(ds) < \infty$ for each $t \ge 0$, where $|M| = M^+ + M^-$, then $X(t) = \int_0^t \varphi(s^-)M(ds), t \ge 0$, defines a martingale, where the stochastic integrals are *sample pathwise Lebesgue–Stieltjes integrals*. In particular $\mathbb{E}\int_0^t \varphi(s^-)M(ds) = 0, t \ge 0$.

Proof. Suppose that G_s is a bounded $\mathcal{F}_s$-measurable function. Then, for $0 \le s < t$,

$$\begin{aligned}
\mathbb{E}(G_s X(t)) &= \mathbb{E}\int_0^t G_s\varphi(u^-)M(du) \\
&= \mathbb{E}\{G_s \int_0^s \varphi(u^-)M(du)\} + \mathbb{E}\{G_s\mathbb{E}(\int_s^t \varphi(u^-)M(du)|\mathcal{F}_s)\} \\
&= \mathbb{E}\{G_s \int_0^s \varphi(u^-)M(du)\} = \mathbb{E}(G_s X(s)).
\end{aligned} \tag{15.4}$$

The martingale property implies $\mathbb{E}X(t) = \mathbb{E}X(0) = 0$. ■

***Lemma 1** (A Simple Stochastic Lebesgue–Stieltjes Lemma).* Suppose that h is an arbitrary measurable function, and suppose γ is $\mathcal{F}_t$-predictable. Consider,

$$dX(t) = \gamma(t^-)dN(t),\ t > 0, X(0) = x.$$

Then

$$h(X(t)) - h(x) = \int_0^t \{h(X(s^-) + \gamma(s)) - h(X(s^-))\}dN(s). \tag{15.5}$$

Proof. Under the hypothesis one has

$$X(t) = x + \sum_{j=1}^{N(t)} \gamma(T_j^-), \quad t \geq 0. \tag{15.6}$$

Moreover,

$$N(T_j) = 1 + N(T_j^-), \quad j = 1, 2, \ldots. \tag{15.7}$$

In particular,

$$X(T_j) = x + \sum_{k=1}^{N(T_j^-)+1} \gamma(T_k^-) = X(T_j^-) + \gamma(T_j), \ j = 1, 2, \ldots. \tag{15.8}$$

Thus, noting that with $T_0 = 0$, one has

$$\begin{aligned} h(X(t)) &= h(x) + \sum_{j=1}^{N(t)} \{h(X(T_j)) - h(X(T_{j-1}))\} \\ &= h(x) + \sum_{j=1}^{N(t)} \{h(X(T_j^-) + \gamma(T_j)) - h(X(T_j^-))\} \\ &= h(x) + \int_0^t \{h(X(s^-) + \gamma(s)) - h(X(s^-))\}dN(s). \end{aligned} \tag{15.9}$$

■

Theorem 15.3 *(Watanabe's Martingale Characterization of Poisson Processes).* Suppose that $N = \{N(t) : t \geq 0\}$ is a simple point process and $t \to \lambda(t), t \geq 0$, is a non-negative measurable and locally integrable function such that $M(t) = N(t) - \int_0^t \lambda(s)ds$, $t \geq 0$, is a martingale with respect to a filtration $\mathcal{F}_t$, $t \geq 0$, on $(\Omega, \mathcal{F}, P)$. Then N is a Poisson process with respect to the filtration $\mathcal{F}_t$, $t \geq 0$, with intensity λ.

Proof. For the proof we exploit the uniqueness theorem[4] for Laplace transform by showing, for r on a half-line, and bounded $\mathcal{F}_s$-measurable G_s, for fixed $s \geq 0$,

[4]See Feller (1971), p. 430.

$$\mathbb{E}(G_s e^{r(N(t)-N(s))}) = \exp\{(e^r - 1)\int_s^t \lambda(u)du\}\mathbb{E}G_s, \quad 0 \le s < t. \tag{15.10}$$

Note that the finiteness of $\mathbb{E}N(t)$ is implicit in the martingale condition. Applying Lemma 1 to

$$Y(s,t) = G_s e^{r(N(t)-N(s))} = G_s e^{-rN(s)} e^{rN(t)}, \quad t \ge s,$$

one has, taking $h(N(t)) = e^{rN(t)}$, $\gamma \equiv 1$, that in differential form (with respect to t for fixed s),

$$\begin{aligned} d_t Y(s,t) &= G_s e^{-rN(s)}\{e^{r(N(t^-)+1)} - e^{rN(t^-)}\}dN(t) \\ &= G_s e^{-rN(s)} e^{rN(t^-)}(e^r - 1)dN(t) \\ &= Y(s,t^-)(e^r - 1)dN(t). \end{aligned} \tag{15.11}$$

It is notationally convenient to continue to express the integral equations in their differential form. In this spirit, one may write $dN(t) = \lambda(t)dt + dM(t), t \ge 0$, so that

$$d_t Y(s,t) = Y(s,t^-)(e^r - 1)\lambda(t)dt + Y(s,t^-)(e^r - 1)dM(t), t \ge s.$$

Now, the meaning is that for $0 \le s \le t$, noting $Y(s,s) = G_s$,

$$Y(s,t) = G_s + (e^r - 1)\int_s^t Y(s,u^-)\lambda(u)du + (e^r - 1)\int_s^t Y(s,u^-)dM(u).$$

Taking expectations one has, writing $g_s = \mathbb{E}G_s$, $y(s,t) = \mathbb{E}Y(s,t)$, $0 \le s \le t$,

$$y(s,t) = g_s + (e^r - 1)\int_s^t \lambda(u)y(s,u^-)du = g_s + (e^r - 1)\int_s^t \lambda(u)y(s,u)du,$$

where continuity of $t \to y(s,t)$ follows from the first integral. Differentiating with respect to t, yields the equation

$$\frac{d}{dt}y(s,t) = (e^r - 1)\lambda(t)y(s,t), 0 \le s \le t, \quad y(s,s) = g_s,$$

and the uniquely determined solution follows by a simple integration of dy/y. Namely,

$$\begin{aligned} \mathbb{E}G_s e^{r(N(t)-N(s)} &= y(s,t) \\ &= \mathbb{E}G_s \exp\{(e^r - 1)\int_s^t \lambda(u)du\}. \end{aligned} \tag{15.12}$$

■

The function $A(t) = \int_0^t \lambda(s)ds$, $t \geq 0$, for which $M(t) = N(t) - A(t)$, $t \geq 0$, is a martingale is referred to as a *compensator* for the simple point process N. More generally,

Definition 15.4. Let N be a counting process. An increasing, predictable process $A(t)$, $t \geq 0$, such that $M(t) = N(t) - A(t)$, $t \geq 0$, is a martingale, is referred to as a compensator for N.

An example of a more general compensator than in the Poisson case is given in the Exercise 2.

Exercises

1. Suppose that $Z = \{Z(t) : t \geq 0\}$ is a (cadlag) predictable process of bounded variation, and $Y = \{Y(t) : t \geq 0\}$ is predictable with $\mathbb{E} \int_0^t |Y(s)||Z|(ds) < \infty$, for each $t \geq 0$, where integration is defined sample pathwise with respect to the Lebesgue–Stieltjes measure Z. Show that $X(t) = \int_0^t Y(s)Z(ds)$, $t \geq 0$ is a predictable process (See Remark 7.4).
2. (*Cox process or Doubly Stochastic Poisson Process*) Let $\Lambda(dt)$ be a locally finite random measure on the Borel σ-field of $[0, \infty)$. The *Cox process* is the counting process N that conditionally given Λ, has a Poisson distribution with intensity measure $\Lambda(dt)$. Show that $A(t) = \Lambda[0, t]$, $t \geq 0$, defines a compensator for N.
3. Let N be a Poisson process with constant intensity parameter $\lambda > 0$.
   (i) Show directly that $Y(t) = e^{\mu N(t) - \lambda(e^{\mu}-1)t}$, $t \geq 0$, is a positive martingale.
   (ii) Compute $\lim_{t\to\infty} Y(t)$.
4. Let N be a simple point process on $[0, \infty)$ with jumps at $0 < T_1 < T_2 < \cdots$, and let μ, γ, x be arbitrary fixed real numbers. Consider $dX(t) = \mu X(t^-)dt + \gamma dN(t)$, $t > 0$, $X(0) = x$. Show that $X(t) = xe^{\mu t} + \gamma \int_0^t e^{\mu(t-s)} dN(s)$, $t \geq 0$. [*Hint*: Multiply the equation by $e^{-\mu t}$ and consider $Y(t) = e^{-\mu t}X(t)$, $t \geq 0$.]
5. Let N be a simple point process on $[0, \infty)$ with jumps at $0 < T_1 < T_2 < \cdots$, and let μ, γ, x be arbitrary fixed real numbers. Consider $dX(t) = \mu X(t^-)dt + \gamma X(t^-)dN(t)$, $X(0) = x$, $t \geq 0$.
   (i) Show that $X(t) = xe^{\mu t} \equiv xe^{\mu t^-}$, $0 \leq t < T_1$, and $X(T_1) = (1 + \gamma)X(T_1^-) = x(1+\gamma)e^{\mu T_1}$.
   (ii) Show for $T_1 \leq t < T_2$, $X(t) = xe^{\mu(t-T_1)}(1+\gamma)e^{\mu T_1} = xe^{\mu t}(1+\gamma)$, $T_1 \leq t < T_2$, and $X(T_2) = (1+\gamma)X(T_2^-) = x(1+\gamma)^2 e^{\mu T_2}$.
   (iii) Show $X(t) = xe^{\mu t}(1+\gamma)^{N(t)}$, $t \geq 0$.
   (iv) Assume that N is a Poisson process with intensity $\lambda > 0$. Calculate $\mathbb{E}X(t)$.
   (v) Assume that μ, γ are predictable and show $X(t) = xe^{\int_0^t \mu(s)ds} \prod_{n:T_n \leq t}(1 + \gamma(T_n)) = xe^{\int_0^t \mu(s)ds + \int_0^t \ln(1+\gamma(s))dN(s)}$, $t \geq 0$.

Chapter 16
First Passage Time Distributions for Brownian Motion with Drift and a Local Limit Theorem

A local limit theorem for convergence of probability density functions is provided as a tool for the computation of hitting time distributions for Brownian motion, with or without drift, as a limit of hitting times for random walk, and other asymptotic limit theorems of this nature.

The "first passage time" refers to the time of first arrival to a point in space by the stochastic process, in this case Brownian motion. The purpose of this chapter is two-fold. First we will show that the pdf of the first passage time of Brownian motion without drift, computed (identified) in Corollary 7.14 by the reflection principle, is also the limit of the first passage time densities for associated re-scaled random walks. Second, we will apply the local limit theorem to compute, and hence identify, the first passage time density for Brownian motion with drift.

Remark 16.1. In physical sciences the pdf, when it exists, of the first passage time to a point y is sometimes referred to as "breakthrough curve" at y. The first passage time to $a > 0$ distribution for Brownian motion with drift $\mu > 0$ is also referred to as the *inverse-Gaussian distribution* with parameters a, μ. It models the concentration of particles which arrive at y for the first time at time t as a function of t. By removal of particles upon their arrival at y, one may empirically estimate the first passage time density of a sufficiently dilute initial injection. Apart from such modeling and prediction considerations, the first passage times play a basic role in various aspects of the mathematical analysis of Brownian motion and related stochastic processes.

The two-fold goals of this chapter rely on a "histogram approximation" of distributions of a sequence of discrete random variables X_n ($n \geq 1$) to derive its convergence in distribution. For later purposes we permit random variables which

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_16

are possibly defective, i.e., allow $P(X_n \in \mathbb{R}) \leq 1$. Let the set of values of X_n be contained in a discrete set $L_n = \{x_i^{(n)} : i \in \mathcal{I}_n\}$, where $\mathcal{I}_n$ is a countable index set. Write $p_i^{(n)} = P(X_n = x_i^{(n)})$.

Assume that there exist non-overlapping intervals $A_i^{(n)}$ of lengths $|A_i^{(n)}| > 0$, $i \in \mathcal{I}_n$, which partition an interval $J \subset \mathbb{R}$ such that (i) $x_i^{(n)} \in A_i^{(n)}$, (ii) $\delta_n := \sup\{|A_i^{(n)}| : i \in \mathcal{I}_n\} \to 0$ as $n \to \infty$, (iii) for every $t \in J$ outside a set of Lebesgue measure zero, and with the index $i = i(n,t)$ such that $t \in A_i^{(n)}$, one has

$$p_{i(n,t)}^{(n)} \big/ |A_{i(n,t)}^{(n)}| \longrightarrow f(t) \qquad \text{as } n \to \infty, \tag{16.1}$$

and (iv) $1 \geq \alpha_n := \sum_{i \in \mathcal{I}_n} p_i^{(n)} \to \alpha := \int_J f(t)dt > 0$ as $n \to \infty$.

Proposition 16.1 *(Local Limit Theorem).* Under the assumptions (i)–(iv) above, $\sum_{\{i \in \mathcal{I}_n : x_i^{(n)} \leq t\}} p_i^{(n)} \to \int_{J \cap (-\infty, t]} f(y)dy$ for every $t \in J$. In particular if X_n are proper random variables, i.e., $\sum_{i \in \mathcal{I}_n} p_i^{(n)} = 1$, then X_n converges in distribution to the random variable with density f.

Proof. On J define the density function $f_n(t) = p_i^{(n)}/|A_i^{(n)}|$ if $t \in A_i^{(n)}$ $(t \in J)$. By assumption (iv) and Scheffé's theorem,[1] $\int_J |f_n(y) - f(y)|dy \to 0$. On the other hand, since $p_i^{(n)} = \int_{A_i^{(n)}} f_n(y)dy$ for all $i \in \mathcal{I}_n$,

$$\left| \sum_{\{i \in \mathcal{I}_n : x_i^{(n)} \leq t\}} p_i^{(n)} - \int_{J \cap (-\infty, t]} f_n(y)dy \right| \leq \left| A_{i(n,t)}^{(n)} \right| |f_n(t)| \leq \delta_n |f_n(t)| \longrightarrow 0.$$

■

Remark 16.2. In the commonly stated version of Scheffé's theorem, one would require $\alpha_n = \alpha$ for all n, instead of condition (iv). But dividing $p_i^{(n)}$ by α_n and $f(t)$ by α, this requirement is easily met. Thus the proof of Proposition 16.1 goes through with this minor modification.

Remark 16.3. Note also that this extends with virtually the same proof to higher dimensions $\mathbb{R}^k$ where $A_i^{(n)}$ is a rectangle of positive k-dimensional volume $|A_i^{(n)}|$.

We have seen in Chapter 3, relation (3.7), that for a simple symmetric random walk starting at zero, the first passage time T_y to the state $y \neq 0$ has the distribution

$$P(T_y = N) = \frac{|y|}{N} \binom{N}{\frac{N+y}{2}} \frac{1}{2^N}, \quad N = |y|, |y|+2, |y|+4, \ldots. \tag{16.2}$$

[1] See BCPT pp. 14–15.

To apply the local limit theory take $X_n = \frac{1}{n} \cdot T_{[\sqrt{n}z]}$, $z \neq 0$ fixed, $\mathcal{I}_n := \{0, 1, 2, \dots\}$, $x_i^{(n)} = \frac{|[\sqrt{n}z]|+2i}{n}$, $A_i^{(n)} = (\frac{1}{n}(|[\sqrt{n}z]| + 2i - 2), \frac{1}{n}(|[\sqrt{n}z]| + 2i)]$ for $i \geq 1$, $A_0^{(n)} = (0, \frac{1}{n}|[\sqrt{n}z]|]$. Then $J = (0, \infty)$. Note that if $t > 0$ and $n \geq 1$, then $t \in A_i^{(n)}$ means $\frac{1}{2}(nt - |[\sqrt{n}z]|) \leq i < \frac{1}{2}(nt - |[\sqrt{n}z]|) + 1$, so that $i(n, t)$ differs from $\frac{1}{2}(nt - |[\sqrt{n}z])$ by at most 1. Also, $p_i^{(n)} \equiv P(X_n = x_i^{(n)}) = P(T_{[\sqrt{n}z]} = |[\sqrt{n}z]| + 2i)$ and $f_n(t) = p_i^{(n)}/|A_i^{(n)}| = \frac{n}{2} p_i^{(n)}$ with $i = i(n, t)$. To check (16.1) let $y = [\sqrt{n}z]$, $N = |[\sqrt{n}z]| + 2i(n, t)$ in (16.2). Observe that N differs from nt by at most 2, so that $y^2/N \to z^2/t$, and both N and $N \pm y$ tend to infinity as $n \to \infty$. Thus by Stirling's formula,

$$
\begin{aligned}
f_n(t) &= \frac{n}{2} \cdot \frac{|y|}{N} \binom{N}{\frac{N+y}{2}} 2^{-N} \\
&= \frac{n \cdot |y|}{2(2\pi)^{1/2} N} \frac{e^{-N} N^{N+\frac{1}{2}} 2^{-N}}{e^{-(N+y)/2} \left(\frac{N+y}{2}\right)^{(N+y)/2+\frac{1}{2}} e^{-(N-y)/2} \left(\frac{N-y}{2}\right)^{(N-y)/2+\frac{1}{2}}} \\
&\quad \times (1 + \mathrm{o}(1)) \\
&= \frac{n \cdot 2|y|}{(2\pi)^{1/2} N^{3/2}} \left(1 + \frac{y}{N}\right)^{-(N+y)/2-\frac{1}{2}} \left(1 - \frac{y}{N}\right)^{-(N-y)/2-\frac{1}{2}} (1 + \mathrm{o}(1)) \\
&= \frac{|z|}{\sqrt{2\pi} t^{3/2}} \left(1 + \frac{y}{N}\right)^{-(N+y)/2} \left(1 - \frac{y}{N}\right)^{-(N-y)/2} (1 + \mathrm{o}(1)), \qquad (16.3)
\end{aligned}
$$

where $\mathrm{o}(1)$ denotes a quantity whose magnitude is bounded above by a quantity $\epsilon_n(t, z)$ that depends only on n, t, z and which goes to zero as $n \to \infty$. Also,

$$
\begin{aligned}
\log\left[\left(1 + \frac{y}{N}\right)^{-\frac{N+y}{2}} \left(1 - \frac{y}{N}\right)^{-\frac{N-y}{2}}\right] &= -\frac{N+y}{2}\left[\frac{y}{N} - \frac{y^2}{2N^2} + \mathrm{O}\left(\frac{|y|^3}{N^3}\right)\right] \\
&\quad + \frac{N-y}{2}\left[\frac{y}{N} + \frac{y^2}{2N^2} + \mathrm{O}\left(\frac{|y|^3}{N^3}\right)\right] \\
&= -\frac{y^2}{2N} + \theta(N, y) \to -z^2/2t. \qquad (16.4)
\end{aligned}
$$

Here $|\theta(N, y)| \leq n^{-1/2} c(t, z)$ and $c(t, z)$ is a constant depending only on t and z. Combining (16.3) and (16.4) and Corollary 7.14, we arrive at the following.

Proposition 16.2. The histogram approximation (16.3) converges to the density of the first passage time τ_z to z of standard Brownian motion starting at zero.

$$
\lim_{n\to\infty} f_n(t) = \frac{|z|}{\sqrt{2\pi} t^{3/2}} \exp\left\{-\frac{z^2}{2t}\right\}. \qquad (16.5)
$$

In particular

$$P(\tau_z^{(n)} \le t) \to P(\tau_z \le t).$$

Remark 16.4. The first passage time distribution for the slightly more general case of Brownian motion $\{X_t : t \ge 0\}$ with zero drift and diffusion coefficient $\sigma^2 > 0$, starting at the origin, may be obtained by applying the formula for the standard Brownian motion $\{(1/\sigma)X_t : t \ge 0\}$. In particular, the first passage time to z for $\{X_t : t \ge 0\}$ is the first passage time to z/σ for the standard Brownian motion. So the probability density function $f_{\sigma^2}(t)$ of τ_z is therefore given as follows.

Proposition 16.3 ***(First Passage Time to z for σB).*** The first passage time to z for Brownian motion with zero drift and diffusion coefficient $\sigma > 0$ has density

$$f_{\sigma^2}(t) = \frac{|z|}{\sqrt{2\pi\sigma^2}t^{3/2}} e^{-\frac{z^2}{2\sigma^2 t}}, \quad t > 0. \tag{16.6}$$

Note that for large t the *tail* of the pdf $f_{\sigma^2}(t)$ is of the order of $t^{-3/2}$. Therefore, although $\{X_t : t \ge 0\}$ will reach z in a finite time with probability 1, the *expected time* is infinite (Exercise 1).

Consider now the first passage time distribution for a Brownian motion $\{X_t : t \ge 0\}$ with a nonzero drift μ and diffusion coefficient σ^2 that starts at the origin. One may note that if, for example, $\mu > 0$, then by the transience property obtained in the previous chapter (or simply by the SLLN), there is a positive probability that the process may not ever reach a given $z < 0$, i.e., one may have a defective first passage time distribution in the sense that $P(\tau_z < \infty) = \int_0^\infty f_{\sigma^2,\mu}(t)dt < 1$.

Although the random walk asymptotics for the proof will follow from the local limit Proposition 16.1 as above, at this stage one is unable to compare the limit to a previous computation for Brownian motion having nonzero drift. To make this connection we will anticipate the FCLT from the next Chapter 17 which shows how functionals (such as first passage times) of the random walk converge to those of the Brownian motion (Also see Remark 1.4 and Remark 16.5.)

Proposition 16.4 ***(First Passage Time Distribution Under Drift).*** Let $\sigma > 0$, $z \in \mathbb{R}$. Also let $\mu \in \mathbb{R}$ and let $\{B_t : t \ge 0\}$ The first passage time τ_z of the diffusion $\{X_t := \mu t + \sigma B_t : t \ge 0\}$ starting at zero with drift μ and diffusion coefficient σ^2 at z has a possibly defective pdf given by

$$f_{\sigma^2,\mu}(t) = \frac{|z|}{(2\pi\sigma^2)^{1/2}t^{3/2}} \exp\left\{-\frac{1}{2\sigma^2 t}(z-\mu t)^2)\right\} \quad (t > 0). \tag{16.7}$$

Proof. Using Corollary 17.6 from Chapter 7, the polygonal process $\{\tilde{X}_t^{(n)} : t \ge 0\}$ corresponding to the simple random walk $S_{m,n} = Z_{1,n} + \cdots + Z_{m,n}$, $S_{0,n} = 0$, with $P(Z_{m,n} = 1) = p_n = \frac{1}{2} + \mu/(2\sigma\sqrt{n})$, converges in distribution to $\{W_t = X_t/\sigma : t \ge 0\}$, which is a Brownian motion with drift μ/σ and diffusion coefficient 1. On the other hand, writing $T_{y,n}$ for the first passage time of $\{S_{m,n} : m = 0, 1, \ldots\}$ to y, one has, by relation (3.4) of Chapter 5,

$$
\begin{aligned}
P(T_{y,n} = N) &= \frac{|y|}{N}\binom{N}{\frac{N+y}{2}} p_n^{(N+y)/2} q_n^{(N-y)/2} \\
&= \frac{|y|}{N}\binom{N}{\frac{N+y}{2}} 2^{-N}\left(1+\frac{\mu}{\sigma\sqrt{n}}\right)^{(N+y)/2}\left(1-\frac{\mu}{\sigma\sqrt{n}}\right)^{(N-y)/2} \\
&= \frac{|y|}{N}\binom{N}{\frac{N+y}{2}} 2^{-N}\left(1-\frac{\mu^2}{\sigma^2 n}\right)^{N/2} \\
&\quad \times \left(1+\frac{\mu}{\sigma\sqrt{n}}\right)^{y/2}\left(1-\frac{\mu}{\sigma\sqrt{n}}\right)^{-y/2}. \qquad (16.8)
\end{aligned}
$$

For $N \neq 0$, the first passage time to w for $\{\tilde{X}_t^{(n)} : t \geq 0\}$ is (asymptotically) the same as $\frac{1}{n}T_{[w\sqrt{n}],n}$. Therefore, one may seek to apply Proposition 16.1 with $X_n = \frac{1}{n}T_{[w\sqrt{n}],n}$, and $x_i^{(n)}$, $A_i^{(n)}$, $i(n,t)$ as above, but with z replaced by w. Also, $p_i^{(n)}$ here is given by (16.8). For $y = [w\sqrt{n}]$ for some given nonzero w, and $N = [nt]$ for some given $t > 0$, one has

$$
\begin{aligned}
&\left(1-\frac{\mu^2}{\sigma^2 n}\right)^{N/2}\left(1+\frac{\mu}{\sigma\sqrt{n}}\right)^{y/2}\left(1-\frac{\mu}{\sigma\sqrt{n}}\right)^{-y/2} \\
&= \left(1-\frac{\mu^2}{\sigma^2 n}\right)^{nt}\left(1+\frac{\mu}{\sigma\sqrt{n}}\right)^{w\sqrt{n}/2}\left(1-\frac{\mu}{\sigma\sqrt{n}}\right)^{-w\sqrt{n}/2}(1+\mathrm{o}(1)) \\
&= \exp\left\{-\frac{t\mu^2}{\sigma^2}\right\}\exp\left\{\frac{\mu w}{2\sigma}\right\}\exp\left\{\frac{\mu w}{2\sigma}\right\}(1+\mathrm{o}(1)) \\
&= \exp\left\{-\frac{t\mu^2}{\sigma^2}+\frac{\mu w}{\sigma}\right\}(1+\mathrm{o}(1)), \qquad (16.9)
\end{aligned}
$$

where $\mathrm{o}(1)$ represents a term that goes to zero as $n \to \infty$. The first passage time τ_z to z for $\{X_t : t \geq 0\}$ is the same as the first passage time to $w = z/\sigma$ for the process $\{W_t : t \geq 0\} = \{X_t/\sigma : t \geq 0\}$. It follows from the hitting time formulae in Proposition 7.21 of Chapter 7, also see Chapter 7, Exercise 15, that if

i. $\mu < 0$ *and* $z > 0$ or if **ii.** $\mu > 0$ *and* $z < 0$,

then there is a positive probability that the process $\{W_t : t \geq 0\}$ will never reach $w = z/\sigma$ (i.e., $\tau_z = \infty$), but nonetheless one has $\lim_{n\to\infty} P(X_n < \infty) \equiv \lim_{n\to\infty} P(T_{y,n} < \infty) = P(\tau_z < \infty)$, with $y = [w\sqrt{n}]$ (Exercise 3). By

Proposition 16.1 this is adequate for the local limit theory. Moreover, in view of (16.3), (16.4), (16.8) and (16.9) we have for $x_i^{(n)} = \frac{[\sqrt{n}w]+2i}{n}$,

$$\lim_{n\to\infty} f_n(t) = \frac{|w|}{\sqrt{2\pi}t^{3/2}} e^{-\frac{w^2}{2t}} \cdot e^{-\frac{t\mu^2}{2\sigma^2}+\frac{\mu w}{\sigma}}. \tag{16.10}$$

Thus the probability density function of τ_z is given by

$$f_{\sigma^2,\mu}(t) = \frac{|z|}{(2\pi\sigma^2)^{1/2}t^{3/2}} \exp\left\{\frac{\mu z}{\sigma^2} - \frac{z^2}{2\sigma^2 t} - \frac{\mu^2}{2\sigma^2}t\right\} \tag{16.11}$$

as asserted. ■

Observe that letting $p(t; 0, y)$ denote the pdf of the normal distribution $\Phi_{\sigma^2,\mu}$ of the position X_t at time t, (16.10) can be expressed as

$$f_{\sigma^2,\mu}(t) = \frac{|z|}{t} p(t; 0, z). \tag{16.12}$$

Also as mentioned before, the integral of $f_{\sigma^2,\mu}(t)$ is less than 1 if either

$$i.\ \mu > 0, z < 0 \quad \text{or} \quad ii.\ \mu < 0, z > 0.$$

In all other cases, Proposition 16.4 provides a proper probability density function.

It is also noteworthy that from the first passage time distribution to a point for one-dimensional Brownian motion, one may readily obtain the hitting time of a line by a two-dimensional Brownian motion and, more generally, the hitting time of a $(k-1)-$dimensional hyperplane by a k-dimensional Brownian motion; Exercise 2.

Note that we have also obtained the distribution of $\max_{0\le t\le T} X_t$ as follows.

Corollary 16.5. Under the conditions of Proposition 16.4 one has the following immediate formula.

$$P(\max_{0\le t\le T} X_t > z) = \int_0^T \frac{|z|}{(2\pi\sigma^2)^{1/2}t^{3/2}} \exp\left\{-\frac{1}{2\sigma^2 t}(z-\mu t)^2)\right\} dt, \quad z \ge x.$$

Recalling the transience of Brownian motion with drift obtained in Chapter 7 let us also note the following.

Proposition 16.6 ***(Time of Last Visit to Zero for Brownian Motion with Drift).*** Fix $\mu \in \mathbb{R}$, $\mu \ne 0$, and $\sigma^2 > 0$ and define

$$\Gamma_0 := \sup\{t : \sigma B_t + \mu t = 0\}.$$

Then, Γ_0 has the pdf $\frac{\mu}{\sigma\sqrt{2\pi}t^{\frac{3}{2}}} e^{-\frac{\mu^2}{2\sigma^2 t}}$, $t \ge 0$.

Proof. Without essential loss of generality, let $\mu > 0$, First recall that the process defined by

$$Z_t = tB_{\frac{1}{t}}, t > 0, \quad Z_0 = 0,$$

is distributed as standard Brownian motion (using the symmetry of Brownian motion). Now observe that for $\sigma^2 > 0$,

$$\begin{aligned}\Gamma_0 &:= \sup\{t : \sigma B_t + \mu t = 0\} = \sup\{t : \frac{1}{t}B_t + \frac{\mu}{\sigma} = 0\} \\ &= \sup\{t : Z_{\frac{1}{t}} = -\frac{\mu}{\sigma}\} = \frac{1}{\inf\{t : Z_t = -\frac{\mu}{\sigma}\}}.\end{aligned} \tag{16.13}$$

An explicit formula for the distribution of τ_y^x was obtained in Corollary 7.14 from which the distribution of Γ_0 follows. (Exercise 4). ■

For the record, the local (central) limit theorem for simple symmetric random walk may be expressed as the following useful Gaussian approximation (Exercise 8).

Proposition 16.7. Let $S_n, n = 0, 1, 2, \ldots$ denote the simple symmetric random walk. Then, for n, j both even or both odd, $\frac{\sqrt{n}}{2}P(S_n = j) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{j}{\sqrt{n}})^2}(1 + o(1))$.

The following general remark is applicable to this chapter and elsewhere.

Remark 16.5. Suppose $X^{(n)}$, $n \geq 1$, is a sequence of processes converging in distribution to a stochastic process X (on $C[0, 1]$ or $C[0, \infty)$), and g is a functional which is continuous a.s. with respect to the distribution of X. Then $g(X^{(n)})$ converges in distribution to $g(X)$, by the Mann-Wald theorem (Proposition 17.4). If by some means (e.g., a local limit theorem) one is able to compute the limit of the distribution function of $g(X^{(n)})$, then, in view of weak convergence, it is the distribution function of $g(X)$ at all points of continuity of the latter. Since this set of continuity points is dense in $\mathbb{R}$ (or $\mathbb{R}^k$ if g is $\mathbb{R}^k$−valued), and since the distribution function of $g(X)$ is right-continuous, it must equal the right-continuous limit everywhere.

Exercises

1. Show that $\mathbb{E}\tau_z = \infty$, where τ_z is the first passage time from Brownian motion with zero drift and diffusion coefficient $\sigma^2 > 0$.
2. (i) Let $L(a, b)$ denote the line $\{(x, y) \in \mathbb{R}^2 : y = ax + b\}$ and let $\mathbf{B}$ denote a 2-dimensional Brownian motion starting at the origin. Compute the distribution of $\tau := \inf\{t : \mathbf{B}_t \in L(a, b)\}$.

(ii) Generalize (i) to the hitting time of the hyperplane $H(a_1, \ldots, a_k, b) = \{(x_1, \ldots, x_k) : \sum_{j=1}^{k} a_j x_j = b\}$ by a k-dimensional Brownian motion.

3. To complete the proof of Proposition 16.4, show that $\lim_{n\to\infty} P(X_n < \infty) \equiv \lim_{n\to\infty} P(T_{[w\sqrt{n},n} < \infty) = P(\tau_z < \infty)$, with $w = z/\sigma$.[*Hint*: The limit can be computed, without presuming the FCLT. The limit, with $x = 0$, $d = z$, $\mu < 0$, is the same as $\int_0^\infty f(t)dt$, as required by the hypothesis of the local limit theorem (Proposition 16.4), may be proved by a direct integration of (16.7).]

4. (i) Complete the proof of Proposition 16.6 for the distribution of the time Γ_0 of the last visit to 0 by a Brownian motion with drift $\mu \neq 0$ and diffusion coefficient σ^2 defined in (16.13).
   (ii) Show that $\mathbb{E}\Gamma_0 = \infty$.

5. Let $\{B_t : t \geq 0\}$ be standard Brownian motion starting at 0 and let $a, b > 0$.

   (i) Calculate the probability that $-at < B_t < bt$ for all sufficiently large t.
   (ii) Calculate the probability that $\{B_t : t \geq 0\}$ last touches the line $y = -at$ instead of $y = bt$. [*Hint*: Consider the process $\{Z_t : t \geq 0\}$ defined by $Z_0 = 0$, $Z_t = tB_{1/t}$ for $t > 0$.]

6. Let $\{(B_t^{(1)}, B_t^{(2)}) : t \geq 0)\}$ be a two-dimensional standard Brownian motion starting at (0, 0). Let $\tau_y = \inf\{t \geq 0 : B_t^{(2)} = y\}$, $y > 0$. Show that $B_{\tau_y}^{(1)}$ has a symmetric Cauchy distribution. [*Hint*: Condition on τ_y and evaluate the integral by substituting $u = (x^2 + y^2)/t$.]

7. Let τ_a be the first passage time to a for a standard Brownian motion starting at 0 with zero drift.

   (i) Verify that $\mathbb{E}\tau_a$ is *not* finite.
   (ii) Show that $(1/n)\tau_{[a\sqrt{n}]} \Rightarrow \tau_a$ in distribution as $n \to \infty$. [*Hint*: Use the continuity theorem[2] for Laplace transforms and Proposition 7.15. The determination of the constant $k > 0$ is not required here.]

8. Provide a proof for Proposition 16.7. [*Hint*: Use Stirling formula approximations to get in the even parity case that $\frac{\sqrt{2n}}{2}\binom{2n}{n+j}2^{-2n} = \frac{1}{\sqrt{2\pi}}(\frac{n^2}{(n+j)(n-j)})^{\frac{1}{2}} \frac{1}{(1+\frac{j}{n})^{n+j}(1-\frac{j}{n})^{n-j}}(1 + o(1))$ followed by grouping terms in the Taylor expansion of the logarithm to get $(n + j)\ln(1 + \frac{j}{n}) + (n - j)\ln(1 - \frac{j}{n}) = \frac{j^2}{n}O(1)$.]

[2] See Feller (1971), p. 431.

Chapter 17
The Functional Central Limit Theorem (FCLT)

The functional central limit theorem, or invariance principle, refers to convergence in distribution of centered and rescaled random walks having finite second moments to Brownian motion. This provides a tool for computing asymptotic limits of functionals of rescaled random walks by analyzing the corresponding functional of Brownian motion. The term "invariance principle"refers to the invariance of the distribution of the limit, namely Brownian motion, regardless of the specific random walk increments, with a finite second moment. The proof given here is by a beautiful technique of Skorokhod in which the random walk paths are embedded within the Brownian motion.

Consider a sequence of i.i.d. random variables $\{Z_m\}_{m=1}^{\infty}$ and assume for the present that $\mathbb{E}Z_m = 0$ and $\mathrm{Var}\, Z_m = \sigma^2 > 0$. Define the random walk

$$S_0 = 0, \quad S_m = Z_1 + \cdots + Z_m \quad (m = 1, 2, \ldots). \tag{17.1}$$

Define, for each value of the *scale parameter* $n \geq 1$, the stochastic process

$$X_t^{(n)} = \frac{S_{[nt]}}{\sqrt{n}} \qquad (t \geq 0), \tag{17.2}$$

where $[nt]$ is the *integer part* of nt. The process $\{S_{[nt]} : t \geq 0\}$ records the discrete time random walk $\{S_m : m = 0, 1, 2, \ldots\}$ on a continuous *time scale* such that in one unit ($t = 1$) of continuous time there will be contributions from n discrete time units. The process $X^{(n)} := \{X_t^{(n)} : t \geq 0\} = \{(1/\sqrt{n})S_{[nt]} : t \geq 0\}$ further *scales*

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_17

distance in such a way that one unit of distance in the new scale equals $\sqrt{n}$ spatial units used for the random walk. This is a convenient normalization since (for large n)

$$\mathbb{E}X_t^{(n)} = 0, \quad \operatorname{Var} X_t^{(n)} = \frac{[nt]\sigma^2}{n} \simeq \sigma^2 t. \tag{17.3}$$

Since the sample paths of $X^{(n)} = \{X_t^{(n)} : t \geq 0\}$ have jumps (though small for large n) and are, therefore, discontinuous, it is technically more convenient to *linearly interpolate* the random walk between one jump point and the next, using the same space-time scales as used for $\{X_t^{(n)} : t \geq 0\}$. The resulting *polygonal process* $\tilde{X}^{(n)} := \{\tilde{X}_t^{(n)} : t \geq 0\}$ is formally defined by (see Figure 17.1)

$$\tilde{X}_t^{(n)} = \frac{S_{[nt]}}{\sqrt{n}} + (nt - [nt])\frac{Z_{[nt]+1}}{\sqrt{n}} \quad t \geq 0. \tag{17.4}$$

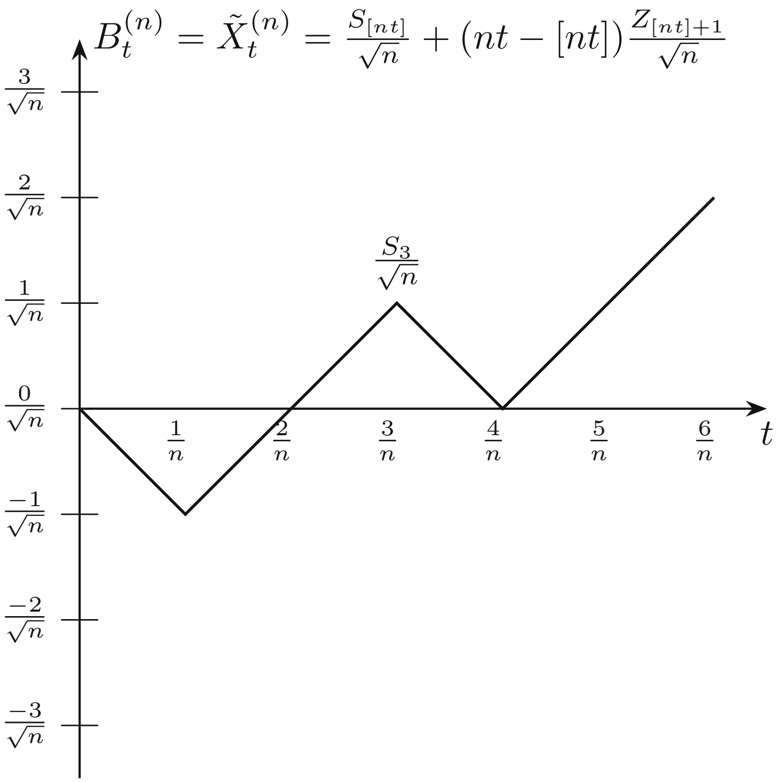


Fig. 17.1 Rescaled Simple Random Walk

In this way, just as for the limiting Brownian motion process, the paths of $\{\tilde{X}_t^{(n)} : t \geq 0\}$ are continuous, i.e., $\tilde{X}^{(n)}$ takes its values in the same space $C[0, \infty)$ as the Brownian motion process.

In any given interval $[0, T]$ the maximum difference between values of the two processes $\{X_t^{(n)} : t \geq 0\}$ and $\{\tilde{X}_t^{(n)} : t \geq 0\}$ cannot exceed

$$\epsilon_n(T) = \max\left(\frac{|Z_1|}{\sqrt{n}}, \frac{|Z_2|}{\sqrt{n}}, \ldots, \frac{|Z_{[nT]+1}|}{\sqrt{n}}\right).$$

Now, for each $\delta > 0$,

$$\begin{aligned} P(\epsilon_n(T) > \delta) &= P\left(|Z_m| > \delta\sqrt{n} \text{ for some } m = 1, 2, \ldots, [nT] + 1\right) \\ &\leq ([nT] + 1)(P(|Z_1| > \delta\sqrt{n})) = ([nT] + 1)\mathbb{E}\mathbf{1}_{([Z_1|>\delta\sqrt{n}]} \qquad (17.5) \\ &\leq ([nT] + 1) \cdot \mathbb{E}(\mathbf{1}_{[|Z_1|>\delta\sqrt{n}]}\frac{Z_1^2}{\delta^2 n}) = \frac{[nT] + 1}{\delta^2 n}\mathbb{E}(\mathbf{1}_{[|Z_1|>\delta\sqrt{n}]}Z_1^2) \to 0 \qquad (17.6) \end{aligned}$$

by the dominated convergence theorem. Thus, one arrives at the following useful fact.

Proposition 17.1. If $\mathbb{E}|Z_1|^2 < \infty$, then for any $T > 0$, one has for arbitrary $\delta > 0$,

$$P(\sup_{0\leq t\leq T} |X_t^{(n)} - \tilde{X}_t^{(n)}| > \delta) = P(\epsilon_n(T) > \delta) \to 0 \qquad \text{as } n \to \infty. \tag{17.7}$$

Thus, on any closed and bounded time interval the behaviors of $\{X_t^{(n)} : t \geq 0\}$ and $\{\tilde{X}_t^{(n)} : t \geq 0\}$ are the same in the large-n limit.

Observe that given any finite set of time points $0 < t_1 < t_2 < \cdots < t_k$, the joint distribution of $(X_{t_1}^{(n)}, X_{t_2}^{(n)}, \ldots, X_{t_k}^{(n)})$ converges to the finite-dimensional distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$, where $\{X_t = \sigma B_t : t \geq 0\}$ is a one-dimensional Brownian motion with zero drift and diffusion coefficient σ^2. To see this, note that $X_{t_1}^{(n)}, X_{t_2}^{(n)} - X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)} - X_{t_{k-1}}^{(n)}$ are independent random variables that by the classical central limit theorem converge in distribution to Gaussian random variables with zero means and variance $t_1\sigma^2, (t_2 - t_1)\sigma^2, \ldots, (t_k - t_{k-1})\sigma^2$. That is to say, the joint distribution of $(X_{t_1}^{(n)}, X_{t_2}^{(n)} - X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)} - X_{t_{k-1}}^{(n)})$ converges to that of $(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}})$. By a linear transformation, one gets the desired convergence of finite-dimensional distributions of $\{X_t^{(n)} : t \geq 0\}$.

Many of the events of interest are infinite-dimensional, e.g., $[\max_{0\leq t\leq T} X_t \geq a]$. For convergence of probabilities of such events the convergence of finite-dimensional distributions is inadequate.[1] A metric for $C[0, T]$ can be defined by $d_T(x, y) = \max_{0\le t\le T} |x(t) - y(t)|$, and a metric for $C[0, \infty)$ is $d(x, y) = \sum_{N=1}^{\infty} 2^{-N} \frac{d_N(x,y)}{1+d_N(x,y)}$. Convergence in the latter metric $d(x, y)$ means uniform convergence on compact subintervals $[0, N]$ for all $N \ge 0$. The Borel σ-fields for these metrics coincide with the σ-fields generated by finite-dimensional events.

A precise statement of the *functional central limit theorem* (FCLT) follows. Because the limiting process, namely Brownian motion, is the same for all increments $\{Z_m\}_{m=1}^{\infty}$ as above, the limit Theorem 17.2 is also referred to as the *Invariance Principle*, i.e., invariance with respect to the distribution of the increment process.

Theorem 17.2 *(The Functional Central Limit Theorem)***.** Suppose $\{Z_m : m = 1, 2, \dots\}$ is an i.i.d. sequence with $\mathbb{E}Z_m = 0$ and variance $\sigma^2 > 0$. Then as $n \to \infty$ the stochastic processes $\{\tilde{X}_t^{(n)} : t \ge 0\}$ converge in distribution to a Brownian motion starting at the origin with zero drift and diffusion coefficient σ^2.

There are two distinct types of applications of Theorem 17.2. In the first type it is used to calculate probabilities of infinite-dimensional events associated with Brownian motion by directly computing limits of distributions of functionals of the scaled simple random walks. In the second type it (invariance) is used to calculate asymptotics of a large variety of partial sum processes, since the asymptotic probabilities for these are the same as those of simple random walks. Several such examples are considered in the next two chapters. The following is another useful feature of weak convergence for computations.

Proposition 17.3. Under the conditions of the FCLT, if $g : C[0, \infty) \to \mathbb{R}$ is continuous, then the sequence of real-valued random variables $g(\tilde{X}^{(n)}), n \ge 1$, converges in distribution to $g(X)$.

Proof. For each bounded continuous function f on $\mathbb{R}$ the map $f \circ g$ is also a bounded continuous function on $C[0, \infty)$. Thus the result follows from the meanings of weak convergence on the respective spaces $C[0, \infty)$ and $\mathbb{R}$. ∎

Example 1. Consider the functional $g(\omega) := \max_{0\le t\le 1} \omega(t), \omega \in C[0, \infty)$. As an application of the FCLT one may obtain the following limit distribution: Let $Z_1, Z_2, \dots$ be an i.i.d. sequence of real-valued random variables standardized to have mean zero, variance one. Let $S_n := Z_1 + \cdots + Z_n, n \ge 1$. Then $g(\tilde{X}^{(n)})$, and therefore $\max_{0\le t\le 1} X_t^{(n)}$, converges in distribution to $\max_{0\le t\le 1} B_t$, where $\{B_t : t \ge 0\}$ is standard Brownian motion starting at 0. Thus

$$\lim_{n\to\infty} P(n^{-\frac{1}{2}} \max_{1\le k\le n} S_k \ge a) = \lim_{n\to\infty} P(\max_{0\le t\le 1} X_t^{(n)} \ge a)$$

---

[1]The convergence in distribution of processes with sample paths in metric spaces such as $S = C[0, T]$ and $S = C[0, \infty)$ is fully developed in BCPT pp.135–157.

$$
\begin{aligned}
&= \lim_{n\to\infty} P(g(\tilde{X}^{(n)}) \geq a) \\
&= P(\max_{0\leq t\leq 1} B_t \geq a). \qquad (17.8)
\end{aligned}
$$

For this calculation to be complete one needs to check that the Wiener measure of the boundary of the set $G = \{\omega \in C[0,\infty) : \max_{0\leq t\leq 1}\omega(t) \geq a\}$ is zero. This follows from the fact that $P(\max_{0\leq t\leq 1} B_t = a) = 0$ (see Corollary 7.12). One may note that another point of view is possible in which the FCLT is used to obtain formulae for Brownian motion by making the special choice of simple symmetric random walk for $Z_1, Z_2, \ldots$, do the combinatorics and then pass to the appropriate limit. Both of these perspectives will be illustrated in this chapter.

It is often useful to recognize that it is sufficient that $g : C[0,\infty) \to \mathbb{R}$ be only a.s. continuous with respect to the limiting distribution for the FCLT to apply, i.e., for the convergence of $g(\tilde{X}^{(n)})$ in distribution to $g(X)$. That is,[2]

Proposition 17.4. If $\tilde{X}^{(n)} := \{\tilde{X}_t^{(n)}\}$ converges in distribution to $X := \{X_t : t \geq 0\}$ and if $P(X \in D_g) = 0$, where $D_g = \{\mathbf{x} \in C[0,\infty) : g \text{ is discontinuous at } \mathbf{x}\}$, then $g(\tilde{X}^{(n)})$ converges in distribution to $g(X)$.

Proof. Let Q_n be the distribution of $\tilde{X}^{(n)}$ and Q that of X. Let g be a function on $C[0,\infty)$ into a metric space S. Let $\mu_n = Q_n \circ g^{-1}$, $\mu = Q \circ g^{-1}$. By Alexandrov's theorem [3] it is enough to show that for every closed set F one has $\limsup_n \mu_n(F) \leq \mu(F)$. But for any closed set F, $g^{-1}(F) \subset \overline{g^{-1}(F)} \subset D_g \cup g^{-1}(F)$, where the overbar denotes the closure of the set. By the weak convergence of Q_n, one has $\limsup_n \mu_n(F) = \limsup_n Q_n(g^{-1}F) \leq \limsup_n Q_n(\overline{g^{-1}(F)}) \leq Q(\overline{g^{-1}(F)}) = Q(D_g \cup g^{-1}(F)) \leq \mu(F)$. ∎

Let us now turn to a beautiful result of Skorokhod[4] representing a general random walk (partial sum process) as values of a Brownian motion at a sequence of successive stopping times (with respect to an enlarged filtration). This will be followed by a proof of the functional central limit theorem (invariance principle) based on the Skorokhod embedding representation. Recall that for $c < x < d$, (see Chapter 7, Proposition 7.18),

$$
P(\tau_d^x < \tau_c^x) = \frac{x-c}{d-c}, \qquad (17.9)
$$

where $\tau_a^x := \overline{\tau}_a(B^x) \equiv \inf\{t \geq 0 : B_t^x = a\}$. Also, as calculated in Chapter 13, Example 2,

[2]This is an early result in the theory of weak convergence sometimes referred to as the Mann-Wald theorem.

[3]See BCPT p.137.

[4]Skorokhod (1965).

$$\mathbb{E}(\tau_c^x \wedge \tau_d^x) = (d-x)(x-c). \tag{17.10}$$

Write $\tau_a = \tau_a^0$, $B^0 = B = \{B_t : t \geq 0\}$. Consider now a two-point distribution $F_{u,v}$ with support $\{u, v\}$, $u < 0 < v$, having mean zero. That is, $F_{u,v}(\{u\}) = v/(v-u)$ and $F_{u,v}(\{v\}) = -u/(v-u)$. It follows from (17.9) that with $\tau_{u,v} = \tau_u \wedge \tau_v$, $B_{\tau_{u,v}}$ has distribution $F_{u,v}$ and, in view of (17.10),

$$\mathbb{E}\tau_{u,v} = -uv = |uv|. \tag{17.11}$$

In particular, the random variable $Z := B_{\tau_{u,v}}$ with distribution $F_{u,v}$ is naturally embedded in the Brownian motion. We will see by the theorem below that any given non-degenerate distribution F with mean zero may be similarly embedded by randomizing over such pairs (u, v) to get a random pair (U, V) such that $B_{\tau_{U,V}}$ has distribution F, and $\mathbb{E}\tau_{U,V} = \int_{(-\infty,\infty)} x^2 F(dx)$, the variance of F. Indeed, this is achieved by the distribution γ of (U, V) on $(-\infty, 0) \times (0, \infty)$ given by

$$\gamma(du\,dv) = \theta(v-u)F_-(du)F_+(dv), \tag{17.12}$$

where F_+ and F_- are the restrictions of F to $(0, \infty)$ and $(-\infty, 0)$, respectively. Here θ is the normalizing constant given by

$$1 = \theta\left[\left(\int_{(0,\infty)} vF_+(dv)\right) F_-((-\infty,0)) + \left(\int_{(-\infty,0)} (-u)F_-(du)\right) F_+((0,\infty))\right],$$

or, noting that the two integrals are each equal to $\frac{1}{2}\int_{-\infty}^{\infty} |x|F(dx)$ since the mean of F is zero, one has

$$1/\theta = \left(\frac{1}{2}\int_{-\infty}^{\infty} |x|F(dx)\right)[1 - F(\{0\})]. \tag{17.13}$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which are defined (1) a standard Brownian motion $B \equiv B^0 = \{B_t : t \geq 0\}$, and (2) a sequence of i.i.d. pairs (U_i, V_i) independent of B, with the common distribution γ above. Let $\mathcal{F}_t := \sigma\{B_s : 0 \leq s \leq t\} \vee \sigma\{(U_i, V_i) : i \geq 1\}$, $t \geq 0$. Define the $\{\mathcal{F}_t : t \geq 0\}$-stopping times

$$\begin{aligned} T_0 \equiv 0, \quad T_1 &:= \inf\{t \geq 0 : B_t = U_1 \text{ or } V_1\}, \\ T_{i+1} &:= \inf\{t > T_i : B_t = B_{T_i} + U_{i+1} \text{ or } B_{T_i} + V_{i+1}\}\ (i \geq 1). \end{aligned} \tag{17.14}$$

Theorem 17.5 *(Skorokhod Embedding).* Assume that F has mean zero and finite variance. Then (a) B_{T_1} has distribution F, and $B_{T_{i+1}} - B_{T_i}$ $(i \geq 0)$ are i.i.d. with common distribution F, and (b) $T_{i+1} - T_i$ $(i \geq 0)$ are i.i.d. with

$$\mathbb{E}\,(T_{i+1} - T_i) = \int_{(-\infty,\infty)} x^2 F(dx). \tag{17.15}$$

Proof. (a) Given (U_1, V_1), the conditional probability that $B_{T_1} = V_1$ is $\frac{-U_1}{V_1 - U_1}$. Therefore, for all $x > 0$,

$$\begin{aligned} P\left(B_{T_1} > x\right) &= \theta \int_{\{v>x\}} \int_{(-\infty,0)} \frac{-u}{v-u} \cdot (v-u) F_-(du) F_+(dv) \\ &= \theta \int_{\{v>x\}} \left\{ \int_{(-\infty,0)} (-u) F_-(du) \right\} F_+(dv) \\ &= \int_{\{v>x\}} F_+(dv), \end{aligned} \tag{17.16}$$

since $\int_{(-\infty,0)}(-u)F_-(du) = \frac{1}{2}\int |x| F(dx) = 1/\theta$. Thus the restriction of the distribution of B_{T_1} on $(0,\infty)$ is F_+. Similarly, the restriction of the distribution of B_{T_1} on $(-\infty, 0)$ is F_-. It follows that $P(B_{T_1} = 0) = F(\{0\})$. This shows that B_{T_1} has distribution F. Next, by the strong Markov property, the conditional distribution of $B^+_{T_i} \equiv \{B_{T_i+t} : t \geq 0\}$, given $\mathcal{F}_{T_i}$, is $P_{B_{T_i}}$ (where P_x is the distribution of B^x). Therefore, the conditional distribution of $B^+_{T_i} - B_{T_i} \equiv \{B_{T_i+t} - B_{T_i}; t \geq 0\}$, given $\mathcal{F}_{T_i}$, is P_0. In particular, $Y_i := \{(T_j, B_{T_j}) : 1 \leq j \leq i\}$ and $X^i := B^+_{T_i} - B_{T_i}$ are independent. Since Y_i and X^i are functions of $B \equiv \{B_t : t \geq 0\}$ and $\{(U_j, V_j)$: $1 \leq j \leq i\}$, they are both independent of (U_{i+1}, V_{i+1}). Since $\tau^{(i+1)} := T_{i+1} - T_i$ is the first hitting time of $\{U_{i+1}, V_{i+1}\}$ by X^i, it now follows that (1) $(T_{i+1} - T_i \equiv \tau^{(i+1)}, B_{T_{i+1}} - B_{T_i} \equiv X^i_{\tau^{(i+1)}})$ is independent of $\{(T_j, B_{T_j}) : 1 \leq j \leq i\}$, and (2) $(T_{i+1} - T_i, B_{T_{i+1}} - B_{T_i})$ has the same distribution as (T_1, B_{T_1}).

(b) It remains to prove (17.15). But this follows from (17.11):

$$\begin{aligned} &\mathbb{E}T_1 \\ &= \theta \int_{(0,\infty)} \int_{(-\infty,0)} (-uv)(v-u) F_-(du) F_+(dv) \\ &= \theta \left[\int_{(0,\infty)} v^2 F_+(dv) \cdot \int_{(-\infty,0)} (-u) F_-(du) + \int_{(-\infty,0)} u^2 F_-(du) \cdot \int_{(0,\infty)} v F_+(dv) \right] \\ &= \int_{(0,\infty)} v^2 F_+(dv) + \int_{(-\infty,0)} u^2 F_-(du) = \int_{(-\infty,\infty)} x^2 F(dx). \end{aligned}$$

■

We now present an elegant proof of Donsker's invariance principle, or functional central limit theorem, i.e., Theorem 17.2, using Theorem 17.5. For this, consider a sequence of i.i.d. random variables Z_i $(i \geq 1)$ with common distribution having mean zero and variance 1. Let $S_k = Z_1 + \cdots + Z_k$ $(k \geq 1)$, $S_0 = 0$, and define the polygonal random function $S^{(n)}$ on $[0, 1]$ as follows:

$$S_t^{(n)} := \frac{S_{k-1}}{\sqrt{n}} + n\left(t - \frac{k-1}{n}\right)\frac{S_k - S_{k-1}}{\sqrt{n}}$$
$$\text{for } t \in \left[\tfrac{k-1}{n}, \tfrac{k}{n}\right],\ 1 \le k \le n. \tag{17.17}$$

That is, $S_t^{(n)} = \frac{S_k}{\sqrt{n}}$ at points $t = \frac{k}{n}$ $(0 \le k \le n)$, and $t \mapsto S_t^{(n)}$ is linearly interpolated between the endpoints of each interval $\left[\frac{k-1}{n}, \frac{k}{n}\right]$. In this notation, Theorem 17.2 asserts that $S^{(n)}$ converges in distribution to the standard Brownian motion, as $n \to \infty$.

Proof. Let T_k, $k \ge 1$, be as in Theorem 17.5, defined with respect to a standard Brownian motion $\{B_t : t \ge 0\}$. Then the random walk $\{S_k : k = 0, 1, 2, \dots\}$ has the same distribution as $\{\widetilde{S}_k := B_{T_k} : k = 0, 1, 2, \dots\}$, and therefore, $S^{(n)}$ has the same distribution as $\widetilde{S}^{(n)}$ defined by $\widetilde{S}^{(n)}_{k/n} := n^{-\frac{1}{2}} B_{T_K}$ $(k = 0, 1, \dots, n)$ and with linear interpolation between k/n and $(k+1)/n$ $(k = 0, 1, \dots, n-1)$. Also, define, for each $n = 1, 2, \dots$, the standard Brownian motion $\widetilde{B}_t^{(n)} := n^{-\frac{1}{2}} B_{nt}$, $t \ge 0$. We will show that

$$\max_{0 \le t \le 1} \left| \widetilde{S}_t^{(n)} - \widetilde{B}_t^{(n)} \right| \longrightarrow 0 \quad \text{in probability as } n \to \infty, \tag{17.18}$$

which implies the desired weak convergence. Now

$$\begin{aligned}
&\max_{0 \le t \le 1} \left| \widetilde{S}_t^{(n)} - \widetilde{B}_t^{(n)} \right| \\
&\le n^{-\frac{1}{2}} \max_{1 \le k \le n} \left| B_{T_k} - B_k \right| \\
&\quad + \max_{0 \le k \le n-1} \left\{ \max_{\frac{k}{n} \le t \le \frac{k+1}{n}} \left| \widetilde{S}_t^{(n)} - \widetilde{S}_{k/n}^{(n)} \right| + n^{-\frac{1}{2}} \max_{k \le t \le k+1} |B_t - B_k| \right\} \\
&= I_n^{(1)} + I_n^{(2)} + I_n^{(3)}, \quad \text{say.}
\end{aligned} \tag{17.19}$$

Now, writing $\tilde{Z}_k = \tilde{S}_k - \tilde{S}_{k-1}$, it is simple to check that as $n \to \infty$,

$$I_n^{(2)} \le n^{-\frac{1}{2}} \max\{|\tilde{Z}_k| : 1 \le k \le n\} \to 0 \quad \text{in probability,}$$
$$I_n^{(3)} \le n^{-\frac{1}{2}} \max_{0 \le k \le n-1} \max\{|B_t - B_k| : k \le t \le k+1\} \to 0 \quad \text{in probability,}$$

using the Chebyshev inequality $P(|Z| > \epsilon\sqrt{n}) \le (n\epsilon^2)^{-1}\mathbb{E}Z^2\mathbf{1}_{[|Z|>\epsilon\sqrt{n}]} \to 0$ as $n \to \infty$ if $\mathbb{E}Z^2 < \infty$, whatever be $\epsilon > 0$.

Hence we need to prove, as $n \to \infty$,

$$I_n^{(1)} := n^{-\frac{1}{2}} \max_{1 \le k \le n} \left| B_{T_k} - B_k \right| \longrightarrow 0 \quad \text{in probability.} \tag{17.20}$$

Since $T_n/n \to 1$ a.s., by the SLLN, it follows that (Exercise 8)

$$\varepsilon_n := \max_{1 \le k \le n} \left| \frac{T_k}{n} - \frac{k}{n} \right| \longrightarrow 0 \quad \text{as } n \to \infty \text{ (almost surely).} \tag{17.21}$$

In view of (17.21), there exists for each $\varepsilon > 0$ an integer n_ε such that $P(\varepsilon_n < \varepsilon) > 1 - \varepsilon$ for all $n \ge n_\varepsilon$. Hence with probability greater than $1 - \varepsilon$ one has for all $n \ge n_\varepsilon$ the estimate

$$\begin{aligned} I_n^{(1)} &\le \max_{\substack{|s-t| \le n\varepsilon, \\ 0 \le s,t \le n+n\varepsilon}} n^{-\frac{1}{2}} |B_s - B_t| = \max_{\substack{|s-t| \le n\varepsilon, \\ 0 \le s,t \le n(1+\varepsilon)}} \left| \widetilde{B}^{(n)}_{s/n} - \widetilde{B}^{(n)}_{t/n} \right| \\ &= \max_{\substack{|s'-t'| \le \varepsilon, \\ 0 \le s',t' \le 1+\varepsilon}} \left| \widetilde{B}^{(n)}_{s'} - \widetilde{B}^{(n)}_{t'} \right| \stackrel{d}{=} \max_{\substack{|s'-t'| \le \varepsilon, \\ 0 \le s',t' \le 1+\varepsilon}} |B_{s'} - B_{t'}| \\ &\longrightarrow 0 \quad \text{as } \varepsilon \downarrow 0, \end{aligned}$$

by the continuity of $t \to B_t$. Given $\delta > 0$ one may then choose $\varepsilon = \varepsilon_\delta$ such that for all $n \ge n(\delta) := n_{\varepsilon_\delta}$, $P(I_n^{(1)} > \delta) < \delta$. Hence $I_n^{(1)} \to 0$ in probability. ■

The following modification of Theorem 17.2 will be useful in obtaining the distribution of functionals of Brownian motion with drift parameter μ as limits of asymmetric simple random walk, such as Proposition 7.21 in the preceding chapter. The proof is essentially the same. In preparation, for each $n \ge 1$, let $Z_{k,n}, k = 1, \ldots, n$ be an i.i.d. sequence of ± 1-valued Bernoulli variables with $P(Z_{k,n}) = 1) = 1 - P(Z_{k,n} = -1) = \frac{1}{2} + \frac{\mu}{2\sqrt{n}}, k = 1, \ldots, n$. In particular, $\mathbb{E}Z_{k,n} = \frac{\mu}{\sqrt{n}}$, and $\mathbb{E}(Z_{k,n} - \frac{\mu}{\sqrt{n}})^2 = 1 - \frac{\mu^2}{n}$. Let $S_{k,n} = Z_{1,n} + \cdots + Z_{k,n}$, $1 \le k \le n$, and $S_{0,n} = 0$. Next define define polygonal random function $S^{(n)}$ on $[0, 1]$ by linear interpolation as before:

$$S_t^{(n)} := \frac{S_{k-1,n}}{\sqrt{n}} + n\left(t - \frac{k-1}{n}\right) \frac{S_{k,n} - S_{k-1,n}}{\sqrt{n}}$$
$$\text{for } t \in \left[\tfrac{k-1}{n}, \tfrac{k}{n}\right], \ 1 \le k \le n. \tag{17.22}$$

Corollary 17.6. $S^{(n)}$ converges in distribution to Brownian motion with drift parameter μ as $n \to \infty$

Proof. Consider the centered and rescaled random walk $\overline{S}_{k,n} = (1 - \frac{\mu^2}{n})^{-\frac{1}{2}} \{(Z_{1,n} - \frac{\mu}{\sqrt{n}}) + \cdots + (Z_{k,n} - \frac{\mu}{\sqrt{n}})\} = \sum_{j=1}^k \overline{Z}_{j,n}, k \ge 1, \overline{S}_{0,n} = 0$, where the i.i.d. displacements $\overline{Z}_{j,n} = (1 - \frac{\mu^2}{n})^{-\frac{1}{2}}(Z_{j,n} - \frac{\mu}{\sqrt{n}}), j = 1, \ldots, n$ have mean zero and variance one. Next define define the centered and scaled polygonal random walk $\overline{S}^{(n)}$ on [0, 1] by

$$\overline{S}_t^{(n)} := \frac{\overline{S}_{k-1,n}}{\sqrt{n}} + n\left(t - \frac{k-1}{n}\right)\frac{\overline{S}_{k,n} - \overline{S}_{k-1,n}}{\sqrt{n}}$$
$$\text{for } t \in \left[\tfrac{k-1}{n}, \tfrac{k}{n}\right],\ 1 \le k \le n. \tag{17.23}$$

Then from Skorokhod embedding one has $\overline{S}^{(n)} \Rightarrow B$ as $n \to \infty$. The assertion follows since $S_{k,n} = \sqrt{1 - \frac{\mu^2}{n}}\overline{S}_{k,n} + \frac{\mu}{\sqrt{n}}k$, $k = 1, \ldots n$, and therefore

$$S_t^{(n)} = \sqrt{1 - \frac{\mu^2}{n}}\overline{S}_t^{(n)} + \mu t + \frac{\mu}{n} \Rightarrow B_t + \mu t, \quad 0 \le t \le 1.$$

■

As an application announced in the preceding chapter, we now have

Proof of Proposition 7.21. Consider for each large n, the Bernoulli sequence $\{Z_{m,n} : m = 1, 2, \ldots\}$ defined above. Write $S_{m,n} = Z_{1,n} + \cdots + Z_{m,n}$ for $m \ge 1$, $S_{0,n} = 0$. Then, with $X_t^{(n)} = \frac{S_{[nt],n}}{\sqrt{n}}$, $t \ge 0$,

$$\mathbb{E}X_t^{(n)} = \frac{\mathbb{E}S_{[nt],n}}{\sqrt{n}} = \frac{[nt]\frac{\mu}{\sigma\sqrt{n}}}{\sqrt{n}} \to \frac{t\mu}{\sigma},$$
$$\operatorname{Var} X_t^{(n)} = \frac{[nt]\operatorname{Var} Z_{1,n}}{n} = \frac{[nt]}{n}\left(1 - \left(\frac{\mu}{\sigma\sqrt{n}}\right)^2\right) \to t,$$

the associated polygonal process $\{\tilde{X}_t^{(n)} : t \ge 0\}$ converges in distribution to a Brownian motion with drift μ/σ and diffusion coefficient 1 that starts at the origin (Exercise 17). Let $\{X_t^x : t \ge 0\}$ be a Brownian motion with drift μ and diffusion coefficient σ^2 starting at x. Then $\{W_t = (X_t^x - x)/\sigma : t \ge 0\}$ is a Brownian motion with drift μ/σ and diffusion coefficient 1 that starts at the origin. Hence, by using the second relation of Proposition 2.1 of Chapter 2, one may calculate

$$\begin{aligned}
P(\tau_c^x < \tau_d^x) &= P(\{X_t^x : t \ge 0\} \text{ reaches } c \text{ before } d) \\
&= P\left(\{W_t\}_{t\ge 0} \text{ reaches } \frac{c-x}{\sigma} \text{ before } \frac{d-x}{\sigma}\right) \\
&= \lim_{n\to\infty}(\{S_{m,n} : m = 0, 1, 2, \ldots\} \text{ reaches } c_n \text{ before } d_n) \\
&= \lim_{n\to\infty}\frac{1 - (p_n/q_n)^{\frac{d-x}{\sigma}\sqrt{n}}}{1 - (p_n/q_n)^{\frac{d-x}{\sigma}\sqrt{n} - \frac{c-x}{\sigma}\sqrt{n}}}
\end{aligned}$$

$$= \lim_{n\to\infty} \left\{ 1 - \left(\frac{1+\frac{\mu}{\sigma\sqrt{n}}}{1-\frac{\mu}{\sigma\sqrt{n}}} \right)^{\frac{d-x}{\sigma}\sqrt{n}} \right\} \Bigg/ \left\{ 1 - \left(\frac{1+\frac{\mu}{\sigma\sqrt{n}}}{1-\frac{\mu}{\sigma\sqrt{n}}} \right)^{\frac{d-c}{\sigma}\sqrt{n}} \right\}$$

$$= \left\{ 1 - \frac{\exp\left\{\frac{d-x}{\sigma^2}\mu\right\}}{\exp\left\{-\frac{d-x}{\sigma^2}\mu\right\}} \right\} \Bigg/ \left\{ 1 - \frac{\exp\left\{\frac{d-c}{\sigma^2}\mu\right\}}{\exp\left\{-\frac{d-c}{\sigma^2}\mu\right\}} \right\}.$$

The first asserted probability now follows again providing that one also checks that $P(\partial[\tau_c^x < \tau_d^x]) = 0$. This case is made simpler by the presence of a drift (Exercise 2). If the first relation of Proposition 2.1 Chapter 2 is used instead of the second, then the second probability is calculated similarly. Next, letting $d \uparrow \infty$ in the first relation of the proposition, one obtains the first result and letting $c \downarrow -\infty$ in the second relation one obtains the second assertion. ■

For another application of Skorokhod embedding let us see how to obtain a law of the iterated logarithm (LIL) for sums of i.i.d. random variables using the LIL for Brownian motion (Theorem 7.23).

Theorem 17.7 *(Law of the Iterated Logarithm).* Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $\mathbb{E}X_1 = 0$, $0 < \sigma^2 := \mathbb{E}X_1^2 < \infty$, and let $S_n = X_1 + \cdots + X_n$, $n \geq 1$. Then with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = 1.$$

Proof. By rescaling if necessary, one may take $\sigma^2 = 1$ without loss of generality. In view of Skorokhod embedding one may replace the sequence $\{S_n : n \geq 0\}$ by the embedded random walk $\{\tilde{S}_n = B_{T_n} : n \geq 0\}$. By the SLLN one also has $\frac{T_n}{n} \to 1$ a.s. as $n \to \infty$. In view of the law of the iterated logarithm for Brownian motion, it is then sufficient to check that $\frac{\tilde{S}_{[t]} - B_t}{\sqrt{t \log\log t}} \to 0$ a.s. as $t \to \infty$. From $\frac{T_n}{n} \to 1$ a.s., it follows for given $\epsilon > 0$ that with probability one, $\frac{1}{1+\epsilon} < \frac{T_{[t]}}{t} < 1 + \epsilon$ for all t sufficiently large. Let $t_n = (1+\epsilon)^n$, $n = 1, 2, \ldots$. Then for $t_n \leq t \leq t_{n+1}$, for some $n \geq 1$, one has

$$\begin{aligned}
M_t &:= \max\left\{ |B_s - B_t| : \frac{t}{1+\epsilon} \leq s \leq t(1+\epsilon) \right\} \\
&\leq \max\left\{ |B_s - B_t| : \frac{t}{1+\epsilon} \leq s \leq t \right\} + \max\{|B_s - B_t| : t \leq s \leq t(1+\epsilon)\} \\
&\leq \max\left\{ |B_s - B_{t_n}| : \frac{t_n}{1+\epsilon} \leq s \leq t_{n+1} \right\} + \max\left\{ |B_s - B_{t_n}| : t_n \leq s \leq t_{n+1} \right\} \\
&\leq 2M_{t_n} + 2M_{t_{n+1}}.
\end{aligned}$$

Since $t_{n+2} - t_{n-1} = \gamma t_{n-1} = \frac{\gamma}{1+\epsilon} t_n$, where $\gamma = (1+\epsilon)^3 - 1$, it follows from the scaling property of Brownian motion, using Lévy's Inequality and Feller's tail probability estimate, that

$$
\begin{aligned}
P\left(M_{t_n} > \sqrt{3\frac{\gamma}{1+\epsilon} t_n \log\log t_n}\right) &= P\left(\max_{0\le u\le 1} |B_u| > \sqrt{3\log\log t_n}\right) \\
&\le 4P\left(B_1 \ge \sqrt{3\log\log(t_n)}\right) \\
&\le \frac{4}{\sqrt{3\log\log t_n}} \exp\left(-\frac{3}{2}\log\log t_n\right) \\
&\le c n^{-\frac{3}{2}}
\end{aligned}
$$

for a constant $c > 0$. Summing over n, it follows from the Borel–Cantelli lemma I that with probability one, $M_{t_n} \le \sqrt{3\frac{\gamma}{1+\epsilon} t_n \log\log t_n}$ for all but finitely many n. Since a.s. $\frac{1}{1+\epsilon} < \frac{T_{[t]}}{t} < 1+\epsilon$ for all t sufficiently large, one has that

$$
\limsup_{t\to\infty} \frac{|\tilde{S}_{[t]} - B_t|}{\sqrt{t\log\log t}} \le \sqrt{3\frac{\gamma}{1+\epsilon}}.
$$

Letting $\epsilon \downarrow 0$ one has $\frac{\gamma}{1+\epsilon} \to 0$, establishing the desired result. ■

Remark 17.1. The LIL was first derived[5] for Bernoulli random variables. This was subsequently generalized[6] to bounded, independent (but not i.i.d) random variables. The final formulation[7] was eventually obtained for i.i.d. random variables with finite second moment.

Remark 17.2. The FCLT (Theorem 17.2) is stated for convergence in $S = C[0,\infty)$, when S has the topology of uniform convergence on compacts. One may take the metric to be $\rho(\omega,\omega') = \sum_{k=1}^{\infty} 2^{-k} d_k/(1+d_k)$, where $d_k = \max\{|(\omega(t) - \omega'(t)| : 0 \le t \le k\}$. Hence if $X^{(n)}$ converges to X on $[0,k]$ for every $k > 0$ (in the metric of d_k), then $X^{(n)}$ converges to X on $[0,\infty)$ (in the metric ρ). The following result simply restates this.

Proposition 17.8. Suppose that $X, X^{(n)}, n \ge 1$, are stochastic processes with values in $C[0,\infty)$ for which one has that $\{X_t^{(n)} : 0 \le t \le k\}$ converges in distribution to $\{X_t : 0 \le t \le k\}$ for each $k = 1, 2, \ldots$. Then $X^{(n)}$ converges in distribution to X as processes in $C[0,\infty)$.

[5] Khinchine (1924).

[6] Kolmogoroff (1929).

[7] Hartman and Wintner (1941).

Exercises

1. Fix $t_0 > 0, a, b \in \mathbb{R}, a < b$, and let G denote the finite-dimensional event $G := \{x \in C[0,\infty) : a \le x(t_0) \le b\}$. Identify ∂G and show $W(\partial G) = 0$, where W is Wiener measure.
2. (i) Show that the boundary of the set $F = \{\omega \in C[0,\infty) : \tau_c(\omega) < \tau_d(\omega)|$ has probability zero under Wiener measure W on $\Omega = C[0,\infty)$, i.e., the probability that the standard Brownian motion belongs to ∂F is zero. [*Hint*: Let G be the set of paths that pass below c before reaching d. Then $G \subset F$. Show that G is open under the topology of uniform convergence on compacts. On the other hand, by Blumenthal's zero-one law, $W(F \backslash G) = 0$. If $\omega \notin F$ belongs to the closure of F, then $\omega \in C$, where C comprises all paths ω which neither reach c nor d. But $W(C) = 0$.]
   (ii) Prove (i) for Brownian motion with a nonzero drift. [*Hint*: All one needs to prove in this case, in addition to (i), is that the probability of $F\backslash G$ is zero (for BM under drift). Let $\{B_t : t \ge 0\}$ be a standard BM (with zero drift). Then $\tilde{B}_t =^{dist} tB_{\frac{1}{t}}, t > 0, \tilde{B}_0 = 0$, is also a standard BM. Applying the Law of Iterated Logarithm to $\tilde{B}_t$, as $t \uparrow \infty$, shows that the set $\{0 < t < \epsilon : B_t + \mu t < 0\}$ is non-empty, with probability one, for every $\epsilon > 0$, whatever be μ.]
3. Show that the functions $f : C[0,\infty) \to \mathbb{R}$ defined by
$$f(\omega) = \max_{a \le t \le b} \omega(t), \qquad g(\omega) = \min_{a \le t \le b} \omega(t)$$
   are continuous for the topology of uniform convergence on bounded intervals.
4. Suppose that for each $n = 1, 2, \ldots, \{x_n(t), 0 \le t \le 1\}$, is the deterministic process whose sample path is the continuous function
$$x_n(t) = \begin{cases} nt, & 0 \le t < \frac{1}{n} \\ 2 - nt, & \frac{1}{n} \le t < \frac{2}{n} \\ 0, & \frac{2}{n} \le t \le 1. \end{cases}$$
   (i) Show that the finite-dimensional distributions converge to those of the a.s. identically zero process $\{z(t) : 0 \le t \le 1\}$, i.e., $z(t) \equiv 0, 0 \le t \le 1$.
   (ii) Check that $\max_{0 \le t \le 1} x_n(t)$ does not converge to $\max_{0 \le t \le 1} z(t)$ in distribution.
5. Give an example to demonstrate that it is *not* the case that the FCLT gives convergence of probabilities of all infinite-dimensional events in $C[0,\infty)$. [*Hint*: The polygonal process has finite total variation over $0 \le t \le 1$ with probability 1.]
6. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}X_n = 0$, $\mathrm{Var}\, X_n = \sigma^2 > 0$. Let $S_n = X_1 + \cdots + X_n, n \ge 1, S_0 = 0$. Express each of the random variables defined

below in terms of the rescaled random walk process $\{X_t^{(n)} := n^{-\frac{1}{2}} S_{[nt]} : t \geq 0\}$ and compute their limit distributions in terms of the appropriate random variable associated with Brownian motion having drift 0 and diffusion coefficient $\sigma^2 > 0$.

(i) Fix $\theta > 0$, $Y_n = n^{-\theta/2} \max\{|S_k|^\theta : 1 \leq k \leq n\}$. [*Hint*: Apply Corollary 7.17 to obtain the limit distribution in terms of a series of normal distributions.]

(ii) Let $M_n^{(0)} = n^{-1/2} \max_{k \leq n}(S_k - \frac{k}{n} S_n)$. Compute the distribution of the running maximum of the Brownian bridge (or tied-down Brownian motion) $B_t^* = B_t - tB_1, 0 \leq t \leq 1$, and apply this to the limit distribution of $M_n^{(0)}$ above. [*Hint*: Show that $P(\max_{0 \leq t \leq 1} B_t^* \leq x) = \lim_{\epsilon \downarrow 0} P(\max_{0 \leq t \leq 1} B_t \leq x || B_1| \leq \epsilon)$ and compute the limit directly from the ratio of probabilities. According to Alexandrov's theorem (see BCPT p. 137) it suffices to check $\limsup_{\epsilon \to 0} P(B \in F | -\varepsilon < B_1 < \varepsilon) \leq P(B^* \in F)$ for closed subsets F of $C[0, 1]$. Let ρ denote the uniform metric on $C[0, 1]$. Then, $\sup_{0 \leq t \leq 1} |B_t^* - B_t| = |B_1|$, so $[|B_1| \leq \delta, B \in F] \subset [B^* \subset F_\delta]$ where $F_\delta = \{\omega \in C[0, 1] : \rho(\omega, F) \leq \delta\}$ for $\rho(\omega, F) = \inf\{\rho(\omega, \eta) : \eta \in F\}$. So for $\varepsilon < \delta$, $P(B \in F | -\varepsilon < B_1 < \varepsilon) \leq P(B^* \in F_\delta | -\varepsilon < B_1 < \varepsilon) = P(B^* \in F_\delta)$.]

(iii) $Y_n = n^{-3/2} \sum_{k=1}^n S_k$. Show that the distribution of $\int_0^1 B_t dt$ is Gaussian with mean zero and variance $1/3$, and apply to the limit distribution of Y_n above. [*Hint*: Express as limit of Riemann sum.]

7. Let $\{S_n : n = 0, 1, \dots\}$ denote the simple symmetric random walk starting at 0, and let

$$m_n = \min_{0 \leq k \leq n} S_k, \qquad M_n = \max_{0 \leq k \leq n} S_k, \qquad n = 1, 2 \cdots .$$

Let $\{B_t : t \geq 0\}$ denote a standard Brownian motion starting at zero and let $m = \min_{0 \leq t \leq 1} B_t$, $M = \max_{0 \leq t \leq 1} B_t$. Then, by the FCLT, $n^{-1/2}(m_n, M_n, S_n)$ converges in distribution to (m, M, B_1) since the functional $\omega \to (\min_{0 \leq t \leq 1} \omega_t, \max_{0 \leq t \leq 1} \omega_t, \omega_1)$ is a continuous map of the metric space $C[0, 1]$ into $\mathbb{R}^3$. For notational convenience, let

$$p_n(j) = P(S_n = j), \qquad p_n(u, v, y) = P(u < m_n \leq M_n < v, S_n = y),$$

for integers u, v, y such that $n - 1 \leq u < 0 < v \leq n + 1$, $u < y < v$. Also let $\Phi(a, b) = P(a < Z \leq b)$, where Z has the standard normal distribution. Related results for Brownian motion are also obtained by other methods (e.g., strong Markov property) in later chapters. Show

(i) $p_n(u, v, y) = \sum_{k=-\infty}^{\infty} p_n(y + 2k(v-u)) - \sum_{k=-\infty}^{\infty} p_n(2v - y + 2k(v-u))$.
[*Hint*: Since the random walk is bounded by n in n steps, these are finite sums for $u < v$. To verify this identity first check it for $n = 0$. The only allowable u, v, y are $u = -1$, $v = 1$, $y = 0$. For this case the left side equals one, the first sum on the right has only one nonzero term (for $k = 0$) which equals one, while the second sum has no nonzero term. Then use induction on n together with the identities $p_n(j) = \frac{1}{2}p_{n-1}(j-1) + \frac{1}{2}p_{n-1}(j+1)$ and $p_n(u, v, y) = \frac{1}{2}p_{n-1}(u-1, v-1, y-1) + \frac{1}{2}p_{n-1}(u+1, v+1, y+1)$.]

(ii) For integers, $u < 0 < v$, $u \le y_1 < y_2 \le v$,

$$\begin{aligned} &P(u < m_n < M_n < v, y_1 < S_n < y_2) \\ &\quad = \sum_{k=-\infty}^{\infty} P(y_1 + 2k(v-u) < S_n < y_2 + 2k(v-u))) \\ &\qquad - \sum_{k=-\infty}^{\infty} P(2v - y_2 + 2k(v-u) < S_n < 2v - y_1 + 2k(v-u)). \end{aligned}$$

[*Hint*: Sum over y in (i).]

(iii) For real numbers $u < 0 < v$, $u \le y_1 < y_2 \le v$,

$$\begin{aligned} &P(u < m \le M < v, y_1 < B_1 < y_2) \\ &\quad = \sum_{k=-\infty}^{\infty} \Phi(y_1 + 2k(v-u), y_2 + 2k(v-u)) \\ &\qquad - \sum_{k=-\infty}^{\infty} \Phi(2v - y_2 + 2k(v-u), 2v - y_1 + 2k(v-u)). \end{aligned}$$

[*Hint*: Respectively substitute the integers $[v\sqrt{n}]$, $-[-u\sqrt{n}]$, $[y_1\sqrt{n}]$, $-[-y_2\sqrt{n}]$ into (ii) ([ ] denoting the integer part function). Use Scheffé's theorem[8] to justify the interchange of limit with summation over k.]

(iv) $P(M < v, y_1 < B_1 < y_2) = \Phi(y_1, y_2) - \Phi(2v - y_2, 2v - y_1)$.
[*Hint*: Take $u = -n - 1$ in (iv) and then pass to the limit.]

(v) $P(u < m \le M < v) = \sum_{k=-\infty}^{\infty} (-1)^k \Phi(u + 2k(v-u), v + 2k(v-u))$.
[*Hint*: Take $y_1 = u$, $y_2 = v$ in (v).]

[8] See BCPT p.14.

(vi) $P(\max_{0\leq t\leq 1} |B_t| < v) = \sum_{k=-\infty}^{\infty} (-1)^k \Phi((2k-1)v, (2k+1)v)$. [*Hint*: Take $u = -v$ in (vii).]

(vii) Explain how the above extends to $m_t := \min\{B_s : s \leq t\}$ and $M_t := \max\{B_s : s \leq t\}$.[*Hint*: Rescale $\{B_{ut} : 0 \leq u \leq 1\}$.]

Chapter 18
ArcSine Law Asymptotics

Suppose two players A and B are engaged in independent repeated plays of a fair game in which each player wins or loses one unit with equal probability. The implicit symmetry of this scenario results in the counterintuitive phenomena that in a long series of plays it is not unlikely that one of the players will remain on the winning side while the other player loses for more than half of the series. This chapter derives the distribution of (a) the last time in $2m$ steps that a simple symmetric random walk visits zero in a finite interval, (b) the time spent on the positive side in a finite interval, and (c) the time of the last zero in a finite interval and the arcsine limit distribution for corresponding functionals of Brownian motion. The reference to *first*, *second*, and *third* arcsine laws largely follows nomenclature of Feller (1968/1971) commonly cited in the probability literature, although they are not derived in that order here, the *first* being due to Lévy. Apart from its aid in illustrating an important nuance for decision makers when dealing with random phenomena, the arcsine law involves a rather non-intuitive distribution of natural functionals of the random walk and Brownian motion. The asymptotic results for random walk are obtained by an application of the local limit theorem from Chapter 16. Although the functional central limit theorem of the previous Chapter 17 can also be applied, it is not required beyond identifying the random walk limits with corresponding functionals of Brownian motion.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_18

Consider a *simple symmetric random walk* $\{S_m\}_{m=0}^{\infty}$ starting at zero, $S_m = X_1 + \cdots + X_m$, $(m \geq 1)$, $S_0 = 0$. The first problem for this chapter is to calculate the distribution of the last visit to zero by $S_0, S_1, \ldots, S_{2n}$. For this we consider the probability that the number b of $+1$'s exceeds the number a of -1's until time N and with a given positive value of the excess $b-a$ at time N. Notice that $N = b+a$ in this instance.

Lemma 1. Let a and b be two integers, $0 \leq a < b$. Then

$$P(S_1 > 0, S_2 > 0, \ldots, S_{a+b-1} > 0, S_{a+b} = b-a)$$
$$= \left[\binom{a+b-1}{b-1} - \binom{a+b-1}{b}\right]\left(\frac{1}{2}\right)^{a+b} = \binom{a+b}{b}\frac{b-a}{a+b}\left(\frac{1}{2}\right)^{a+b} \tag{18.1}$$

Proof. Each of the $\binom{a+b}{b}$ paths from $(0,0)$ to $(a+b, b-a)$ has probability $(\frac{1}{2})^{a+b}$. We seek the number M of those for which $S_1 = 1, S_2 > 0, S_3 > 0, \ldots, S_{a+b-1} > 0, S_{a+b} = b-a$. Now the paths from $(1,1)$ to $(a+b, b-a)$ that cross or touch zero (the horizontal axis) are in one-to-one correspondence with the set of all paths that go from $(1,-1)$ to $(a+b, b-a)$. This correspondence is set up by reflecting each path of the last type about zero (i.e., about the horizontal time axis) up to the first time after time zero that zero is reached and leaving the path from then on unchanged. The reflected path leads from $(1,1)$ to $(a+b, b-a)$ and crosses or touches zero. Conversely, any path leading from $(1,1)$ to $(a+b, b-a)$ that crosses or touches zero, when reflected in the same manner, yields a path from $(1,-1)$ to $(a+b, b-a)$. But the number of all paths from $(1,-1)$ to $(a+b, b-a)$ is simply $\binom{a+b-1}{b}$, since it requires b plus 1's and $a-1$ minus 1's. Hence

$$M = \binom{a+b-1}{b-1} - \binom{a+b-1}{b},$$

since there are altogether $\binom{a+b-1}{b-a}$ paths from $(1,1)$ to $(a+b, b-a)$. Now a straightforward simplification yields

$$M = \binom{a+b}{b}\frac{b-a}{a+b}.$$

■

Lemma 2. For the simple symmetric random walk starting at zero, we have

$$P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2n} = 0) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}. \tag{18.2}$$

Proof. By symmetry, the leftmost side of (18.2) equals, using Lemma 1,

$$
\begin{aligned}
&2P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0)\\
&\quad = 2\sum_{r=1}^{n} P(S_1 > 0, S_2 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r)\\
&\quad = 2\sum_{r=1}^{n} \left[\binom{2n-1}{n+r-1} - \binom{2n-1}{n+r}\right]\left(\frac{1}{2}\right)^{2n}\\
&\quad = 2\binom{2n-1}{n}\left(\frac{1}{2}\right)^{2n} = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = P(S_{2n} = 0),
\end{aligned}
$$

where we have adopted the convention that $\binom{2n-1}{2n} = 0$ in writing the middle equality. ■

Theorem 18.1. Let $\Gamma^{(m)} = \max\{j : 0 \le j \le m, S_j = 0\}$. Then

$$
\begin{aligned}
&P(\Gamma^{(2n)} = 2k)\\
&\quad = P(S_{2k} = 0)P(S_{2n-2k} = 0)\\
&\quad = \binom{2k}{k}\left(\frac{1}{2}\right)^{2k}\binom{2n-2k}{n-k}\left(\frac{1}{2}\right)^{2n-2k}\\
&\quad = \frac{(2k)!(2n-2k)!}{(k!)^2((n-k)!)^2}\left(\frac{1}{2}\right)^{2n} \qquad \text{for } k = 0, 1, 2, \ldots, n.
\end{aligned}
\tag{18.3}
$$

Proof. By Lemma 2,

$$
\begin{aligned}
P(\Gamma^{(2n)} = 2k) &= P(S_{2k} = 0, S_{2k+1} \neq 0, S_{2k+2} \neq 0, \ldots, S_{2n} \neq 0)\\
&= P(S_{2k} = 0)P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-2k} \neq 0)\\
&= P(S_{2k} = 0)P(S_{2n-2k} = 0).
\end{aligned}
$$

■

The following symmetry relation is a corollary of Theorem 18.1:

$$
P(\Gamma^{(2n)} = 2k) = P(\Gamma^{(2n)} = 2n - 2k) \qquad \text{for all } k = 0, 1, \ldots, n. \tag{18.4}
$$

Theorem 18.2. Let $\{Z_1, Z_2, \ldots\}$ be a sequence of i.i.d. random variables such that $\mathbb{E}Z_1 = 0$, $\mathbb{E}Z_1^2 = 1$. Then, defining $\gamma^{(n)} = \frac{1}{n}\Gamma^{(n)}$, one has

$$
\lim_{n\to\infty} P(\gamma^{(n)} \le x) = \frac{2}{\pi}\sin^{-1}\sqrt{x}. \tag{18.5}
$$

Proof. We will use the local limit theorem from Chapter 16 for the sequence $X_n := \gamma^{(2n)} \equiv \frac{1}{2n}\Gamma^{(2n)}$, with $x_i^{(n)} = \frac{2i}{2n} = \frac{i}{n}$, $0 \le i \le n$, $p_i^{(n)} = P(X_n = x_i^{(n)}) = P(\Gamma^{(2n)} = 2i)$ as given by (18.3), and $A_i^{(n)} = (\frac{i}{n}, \frac{i+1}{n}]$ $(i = 1, 2, \dots, n-1)$, $A_0^{(n)} = [0, \frac{1}{n}]$. Then $|A_i^{(n)}| = \frac{1}{n}$. For $i \to \infty$ and $n - i \to \infty$ as $n \to \infty$, one has, by Stirling's approximation,

$$
\begin{aligned}
p_i^{(n)} &= \frac{(2i)!(2n-2i)!}{(i!)^2((n-i)!)^2} 2^{-2n} \\
&= \frac{\sqrt{2\pi} e^{-2i}(2i)^{2i+\frac{1}{2}} \sqrt{2\pi} e^{-(2n-2i)}(2n-2i)^{2n-2i+\frac{1}{2}}}{\left(\sqrt{2\pi} e^{-i} i^{i+\frac{1}{2}} \sqrt{2\pi} e^{-(n-i)}(n-i)^{n-i+\frac{1}{2}}\right)^2} \cdot 2^{-2n}(1+o(1)) \\
&= \frac{1}{\pi\sqrt{i(n-i)}}(1+o(1)).
\end{aligned}
$$

Fix $x \in (0, 1)$. Then $x \in A_i^{(n)}$ with $i = i(n, x)$, where $nx - 1 \le i(n, x) < nx$ for all $n > \frac{1}{x}$. Hence, with $i = i(n, x)$,

$$
f_n(x) := \frac{p_i^{(n)}}{|A_i^{(n)}|} = n\left(\frac{1}{\pi\sqrt{i(n-i)}}(1+o(1))\right) \to \frac{1}{\pi\sqrt{x(1-x)}} = f(x).
$$

By the local limit theorem (Proposition 16.1), $X_n \equiv \gamma^{(2n)}$ converges in distribution to the distribution whose density is $f(x)$ depicted in Figure 18.1. ∎

Remark 18.1. Paul Lévy[1] had already discovered the (first) arcsine law for Brownian motion without appeal to random walk limits. Feller's asymptotic arcsine law for random walk could also be viewed as a consequence of Lévy's arcsine law and the invariance principle. On the other hand, Lévy's arcsine law for Brownian motion follows from Feller's via the (same) functional central limit theorem. So this provides a good illustration of the dual use of the functional central limit theorem for those familiar with Chapter 17. The reference to *first*, *second*, and *third* arcsine laws largely follows nomenclature of Feller occasionally cited in the probability literature, although they are not derived in that order here.

The next lemma is a useful tool for identifying continuity sets for weak convergence, see Exercises 1 and 2.

Lemma 3. Let $\Omega = C[0, 1]$, and let Q be a probability measure on its Borel σ-field. If the distribution function $F(x) = Q(\{\omega \in \Omega : X(\omega) \le x\})$ of an upper semicontinuous random variable X on Ω is continuous at $x = t_0$, then $A = \{\omega \in \Omega : X(\omega) \le t_0\}$ is a Q-continuity set.

[1] Lévy (1939).

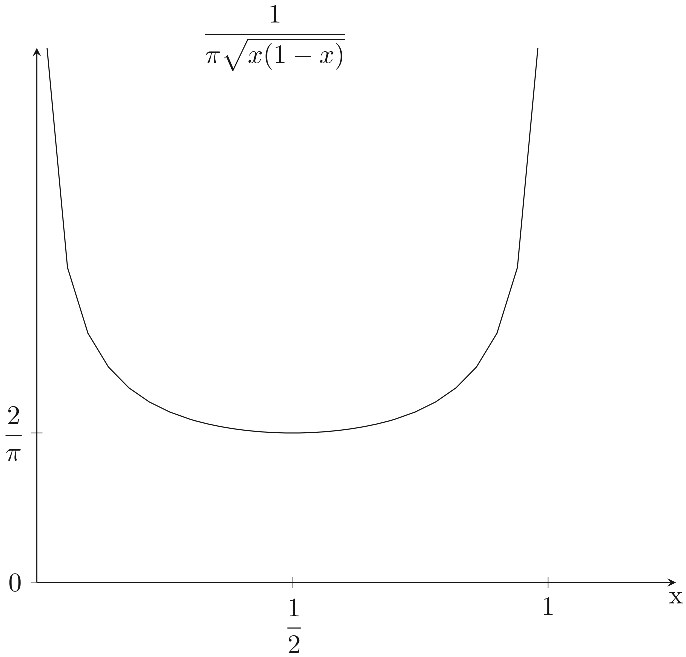


Fig. 18.1 Arcsine Law (pdf)

Proof. From upper-semicontinuity, $U = \{\omega \in \Omega : X(\omega) < t_0\} \subset A$ is open. So $\partial A \subset \overline{A} \backslash U \subset \{\omega \in C[0,1] : X(\omega) = t_0\}$. Thus, $Q(\partial A) \leq P(X(\omega) = t_0) = 0$, making A a Q-continuity set. ■

Remark 18.2. Consider a simple symmetric random walk $\{S_m\}_{m=0}^{\infty}$ and the polygonal processes $\{\tilde{X}_t^{(n)} : t \geq 0\}$ $(n \geq 1)$ associated with it (see (17.4)). Then

$$\begin{aligned}\gamma^{(n)} &:= \sup\{t : 0 \leq t \leq 1, \tilde{X}_t^{(n)} = 0\} \\ &= \frac{1}{n}\sup\{m : 0 \leq m \leq n, S_m = 0\} = \frac{1}{n}\Gamma^{(n)}.\end{aligned}$$

The following (invariance) corollary can be obtained by an application of the FCLT (see Exercise 6).

Corollary 18.3 ***(The Second ArcSine Law).*** Let $\{B_t : t \geq 0\}$ be a standard Brownian motion starting at zero. Let $\gamma = \sup\{t : 0 \leq t \leq 1, B_t = 0\}$. Then γ has the probability density function

$$f(x) = \frac{1}{\pi(x(1-x))^{1/2}}, \quad 0 < x < 1, \tag{18.6}$$

and corresponding distribution function

$$P(\gamma \leq x) = \int_0^x f(y)dy = \frac{2}{\pi} \sin^{-1} \sqrt{x}, \quad 0 \leq x \leq 1. \tag{18.7}$$

One may also consider the time spent above zero by the polygonal process $\{\tilde{X}_t^{(2n)} : 0 \leq t \leq 1\}$. It is simple to see that this time equals $\frac{1}{2n} U_{2n}$, where for mathematical convenience, $U_{2n} = \#\{k \leq 2n : S_{k-1} \text{ or } S_k > 0\}$ counts the number of (interpolated) path segments connecting $(k-1, S_{k-1})$ to (k, S_k) that lie on the positive side of the horizontal, i.e., either $S_{k-1} = 0, S_k > 0$, or $S_{k-1} > 0, S_k = 0$, or both are positive. By an induction argument, one can show that U_{2n} has the same distribution as $\Gamma^{(2n)}$ as follows.

Lemma 4. For the simple symmetric random walk, U_{2n} and $\Gamma^{(2n)}$ have the same distribution.

Proof. We first prove that $P(U_{2n} = 0) = P(\Gamma^{(2n)} = 0)$. For this,

$$\begin{aligned}
&P(U_{2n} = 0) \\
&\quad = P(S_j \leq 0, 1 \leq j \leq 2n) = P(S_j \geq 0, 1 \leq j \leq 2n) \\
&\quad = 2P(X_{2n+1} = 1, S_j \geq 0, 1 \leq j \leq 2n) \\
&\quad = 2P(X_{2n+1} = 1, S_j + X_{2n+1} \geq 1, 1 \leq j \leq 2n) \\
&\quad = 2P(S_j \geq 1, 1 \leq j \leq 2n+1) \\
&\quad = 2P(S_j > 0, 1 \leq j \leq 2n+1),
\end{aligned}$$

where the second to last equality follows by relabeling i.i.d. increments as $X'_1 = X_{2n+1}, X'_j = X_{j-1}, 2 \leq j \leq 2n$. Thus, $P(U_{2n} = 0) = 2\sum_{m=1}^{n+1} P(S_j > 0, 1 \leq j \leq 2n, S_{2n+1} = 2m-1) = 2\binom{2n}{n}2^{-(2n+1)} = \binom{2n}{n}2^{-2n}$, by the telescoping sum on the left side of Lemma 1, with $a+b = 2n+1, a-b = 2m-1$, i.e., $b = n+m, m = 1, \ldots, n+1$. Thus, using Theorem 18.1, $P(U_{2n} = 0) = \binom{2n}{n}2^{-n} = P(\Gamma^{(2n)} = 0)$. Also $P(U_{2n} = 2n) = P(S_j \geq 0, 1 \leq j \leq 2n) = P(S_j \leq 0, 1 \leq j \leq 2n) = P(U_{2n} = 0) = P(\Gamma^{(2n)} = 0) = P(\Gamma^{(2n)} = 2n)$. Indeed, by reflection symmetry, $P(U_{2n} = 2m) = P(U_{2n} = 2n-2m)$, for $0 \leq m \leq n$. For $1 \leq k \leq n-1$, consider the paths corresponding to the event $[U_{2n} = 2k]$. Either a path has positive segments up to the time $2j$ of the first return to zero and has $2k-2j$ positive excursions in the interval $[2j, 2n]$, or the path is negative until the first return to zero at time $2j$ and has $2k$ positive excursions in the interval $[2j, 2n]$. Summing the probabilities of these events, one has upon conditioning these events on $[X_1 = 1]$ and $[X_1 = -1]$, respectively,

$$P(U_{2n} = 2k) = \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(U_{2n-2j} = 2k - 2j) + \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(U_{2n-2j} = 2k), \tag{18.8}$$

where $\tau = \min\{j \geq 1 : S_j = 0\}$. Make the induction hypothesis $U_{2m} =^{\text{dist}} \Gamma^{(2m)}$ for $m \leq n-1$, and substitute into (18.8) to get

$$\begin{aligned}
&P(U_{2n} = 2k)\\
&= \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(\Gamma^{(2n-2j)} = 2k - 2j) + \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(\Gamma^{(2n-2j)} = 2k)\\
&= \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(S_{2k-2j} = 0)P(S_{2n-2k} = 0)\\
&\quad + \frac{1}{2}\sum_{j=1}^{k} P(\tau = 2j)P(S_{2k} = 0)P(S_{2n-2j-2k} = 0)\\
&= \frac{1}{2}P(S_{2n-2k} = 0)\sum_{j=1}^{k} P(\tau = 2j)P(S_{2k-2j} = 0)\\
&\quad + \frac{1}{2}P(S_{2k} = 0)\sum_{j=1}^{k} P(\tau = 2j)P(S_{2n-2j-2k} = 0)\\
&= \frac{1}{2}P(S_{2n-2k} = 0)P(S_{2k} = 0) + \frac{1}{2}P(S_{2k} = 0)P(S_{2n-2k} = 0)\\
&= P(S_{2k} = 0)P(S_{2n-2k} = 0) = P(\Gamma^{(2n)} = 2k),
\end{aligned}$$

where the second to the last line follows from the strong Markov property. This completes the induction. ■

In view of Theorem 18.2 and Lemma 4, one obtains the following corollary.

Corollary 18.4.

$$\lim_{n\to\infty} P(U_{2n} \leq t) = \frac{2}{\pi}\sin^{-1}(\sqrt{t}). \tag{18.9}$$

The connection to Brownian motion is made using the functional central limit theorem to obtain the following.

***Corollary 18.5** (The First ArcSine Law).* Let U be the Lebesgue measure of the set $\{t \le 1 : B_t \ge 0\}$, where $\{B_t : t \ge 0\}$ is a standard Brownian motion starting at 0. Then,

$$P(U \le t) = \frac{2}{\pi} \sin^{-1}(\sqrt{t}). \tag{18.10}$$

The final arcsine law concerns the time at which the random walk $S_0, S_1, \ldots, S_n$ attains its maximum value, i.e., the argmax of $i \to S_i, i \le n$. Since the argmax for random walk may be a *set* of points, consider the point V_n of the *first* maximum. That is, for even $k = 2m$, say,

$$[V_{2n} = k] = [S_0 < S_k, \ldots, S_{k-1} < S_k] \cap [S_{k+1} \le S_k, \ldots, S_{2n} \le S_k]. \tag{18.11}$$

In preparation we record some basic random identities that are obtained by symmetries, reflections, and translations of random walk paths. First by simple symmetry, one has the following relationships (Exercise 3):

$$P(S_1 \le 0, \ldots, S_{2n} \le 0) = P(S_1 \ge 0, \ldots, S_{2n} \ge 0). \tag{18.12}$$

$$P(S_1 < 0, \ldots, S_{2n} < 0) = P(S_1 > 0, \ldots, S_{2n} > 0). \tag{18.13}$$

$$P(S_1 \ne 0, \ldots, S_{2n} \ne 0) = 2P(S_1 < 0, \ldots, S_{2n} < 0). \tag{18.14}$$

Moreover, since every strictly positive path emanating from (0, 0) must pass through (1, 1),

$$P(S_1 > 0, \ldots, S_{2n} > 0) = \frac{1}{2} P(S_2 \ge 1, \ldots, S_{2n} \ge 1) = \frac{1}{2} P(S_1 \ge 0, \ldots, S_{2n-1} \ge 0), \tag{18.15}$$

where the last equality is the result of a 1–1 unit coordinate shift map on the possible paths. But since $2n - 1$ is odd, $[S_1 \ge 0, \ldots, S_{2n-1} \ge 0] = [S_1 \ge 0, \ldots, S_{2n-1} \ge 1] = [S_1 \ge 0, \ldots, S_{2n-1} \ge 0, S_{2n} \ge 0]$, so that $P(S_1 \ge 0, \ldots, S_{2n-1} \ge 0) = P(S_1 \ge 0, \ldots, S_{2n-1} \ge 1) = P(S_1 \ge 0, \ldots, S_{2n} \ge 0)$. That is, using this with (18.15),

$$P(S_1 \ge 0, \ldots, S_{2n} \ge 0) = 2P(S_1 > 0, \ldots, S_{2n} > 0). \tag{18.16}$$

Similarly, but with a little more cleverness, one may construct a one-to-one map between the indicated sets of paths for the following equivalence[2] linking each of the above probabilities (18.12)–(18.16) to $P(S_{2n} = 0)$.

[2]This construction is due to Edward Nelson according to Feller (1968), p. 96.

Lemma 5.

$$P(S_1 \geq 0, \ldots, S_{2n} \geq 0) = P(S_{2n} = 0).$$

Proof. Consider a possible simple random walk path (k, s_k), $k = 0, 1, \ldots, 2n$ from $(0, 0)$ $(2n, 0)$, i.e., any path from the origin for which $s_{2n} = 0$. To follow the proof, it may be helpful to draw figures to depict the indicated transformations. Suppose that the first (absolute) minimum point is $(k', -m')$, $m' \geq 0$. Reflect the part of the path from $(0, 0)$ to $(k', -m')$ by 180° along the vertical line $k = k'$ to a path section from $(k', -m')$ to $(2k', 0)$, and translate this reflected section by $(2n - k', m')$ to start at $(2n, 0)$ and end at $(2n + k', m')$. The final step is to render $(k', -m')$ as the origin of a new coordinate system by translating the entire new path by $(-k', m')$. The resulting path then extends from the origin to $(2n, 2m')$, and all ordinates lie strictly above or on the horizontal axis. Note that if $m' = 0$, then $k' = 0$, and these transformations leave the path fixed. Moreover, this transformation from a path for which $s_0 = 0, s_{2n} = 0$ to the path with $s_0 = 0, s_1 \geq 0, \ldots, s_{2n} \geq 0$ is invertible, i.e., paths $s_0 = 0, s_1 \geq 0, \ldots, s_{2n} \geq 0$ are mapped to paths with $s_{2n} = 0$ by reversing the transformations (Exercise 4). In particular, the proof is complete since it is shown that the respective numbers of paths are the same, and all the paths are equally probable. ■

The proof of the next lemma is interesting in its own right, as it employs Feller's notion of *path duality*, wherein one observes that reversing the order of the displacements of a finite path segment of length n is a one-to-one transformation between sets of paths of length n. Equivalently, one rotates the given path through 180^o about its right endpoint and translates the resulting path to make the right endpoint the starting point (origin).

Lemma 6. Let $k = 2m$ or $k = 2m + 1$ according to whether the first maximum occurs at an even or odd time. If $0 < k < 2n$,

$$P(V_{2n} = k) = \frac{1}{2} P(S_{2m} = 0) P(S_{2n-2m} = 0),$$

and for $m = 0$, $m = n$, respectively, one has

$$P(V_{2n} = 0) = P(S_{2n} = 0),$$
$$P(V_{2n} = 2n) = \frac{1}{2} P(S_{2n} = 0).$$

Proof. For path length $k = 2m \geq 2$, define the dual paths by $S_j^* = X_1^* + \cdot + X_j^* = S_{2m} - S_j$, $j = 0, 1, \ldots, 2m$, where $X_1^* = X_{2m}, \ldots, X_{2m}^* = X_1$. Then the events $[S_{2m} > S_j, j = 0, \ldots, 2m - 1]$ and $[S_j^* > 0, j = 1, \ldots, 2m]$ are dual and, therefore, have the same probability $\frac{1}{2} P(S_j^* \geq 0, j = 0, \ldots, 2m) = \frac{1}{2} P(S_{2m} = 0)$, by (18.16) and Proposition 5. The event $[S_{k+1} \leq S_k, \ldots, S_{2n} \leq S_k]$ in (18.11) is the event that in a random walk path of length $2n - k$, the walk remains non-negative at each step, independently of the event $[S_{2m} > S_j, j = 0, \ldots, 2m-1]$, and is an event with probability $P(S_{2n-2m} = 0)$. The case of odd $k = 2m+1$ is left to the reader. In the case $k = 0$, the path remains negative for the duration $2n$, and for $k = 2n$, the dual path must remain strictly positive. ■

The proof of the following now follows the same lines as that for Theorem 18.2 and is left as Exercise 5

Theorem 18.6.

$$\lim_{n\to\infty} P(V_{2n} \le t) = \frac{2}{\pi} \sin^{-1}(\sqrt{t}).$$

Corollary 18.7 *(The Third ArcSine Law).* Let $M_t = \max\{B_s : 0 \le s \le t\}, t \ge 0$, be the running maximum of standard Brownian motion starting at 0, and let $V = \inf\{t \in [0,1] : B_t = M_1\}$ be the argmax on $[0, 1]$. Then,

$$P(V \le t) = \frac{2}{\pi} \sin^{-1}(\sqrt{t}). \tag{18.17}$$

Exercises

Throughout the exercises below, $\{S_n\}_{n=0}^{\infty}$ denotes the simple symmetric random walk starting at 0.

1. Let $C[0, \infty)$ be given the topology of uniform convergence on compacts. Show that
   (i) $\overline{\tau}_a(f)$ defined by (7.21) is lower semicontinuous on $C[0,\infty)$ into $[0,\infty]$, and
   (ii) $\tau_a := \overline{\tau}_a(B)$ is a $\{\mathcal{G}_t\}$-stopping time, where $\{\mathcal{G}_t : t \ge 0\}$ is the Brownian filtration, i.e., $\mathcal{G}_t := \sigma\{B_s : 0 \le s \le t\}$.
2. Show how (18.7) follows by combining the convergence in distribution of $X_n = \gamma^{(2n)}$ to the distribution with density $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$, $0 < x < 1$, i.e., local limit theorem, with the FCLT. [*Hint*: Use Exercise 1 and Lemma 3.]
3. Establish the equivalent probabilities in (18.12)–(18.14). [*Hint*: Construct the appropriate bijections between the indicated sets of random walk paths.]
4. Give an algorithm for inverting Nelson's construction described in the proof of Lemma 5. [*Hint*: If $s_0 = 0, s_1 \ge 0, \ldots, s_{2n} = 0$, then the path is fixed. Consider the case $s_{2n} = 2m' \ge 2$. Reflect the part of the path from (k', m') to $(2n, 2k')$ on the vertical line $k = k'$ corresponding to the last time $s_{k'} = m'$, and translate this reflected part by $(-k', -m')$. Translate this entire concatenated path by $(2n - k', -m')$ to complete the inversion.]
5. Give the proof of Theorem 18.6. [*Hint*: The proof follows that of Theorem 18.2.]

6. Prove Corollary 18.3.
7. Prove Corollary 18.5.
8. Prove a version of Lemma 6 when V_n is defined to be the time of the last maximum, rather than the first. [*Hint*: For $k = 0$ and $k = 2n$, the probabilities are interchanged; otherwise, write $k = 2m$ and $k = 2m - 1$ for the even and odd cases.]
9. Show the following for $r \neq 0$:

   (i) $P(S_1 \neq 0, S_2 \neq 0, \cdots, S_{2n-1} \neq 0, S_{2n} = 2r) = \binom{2n}{n+r}\frac{|r|}{n}2^{-2n}$.

   (ii) $P(\Gamma^{(2n)} = 2k, S_{2n} = 2r) = \binom{2k}{k}\binom{2n-2k}{n-k+r}\frac{|r|}{n-k}2^{-2n}$
   $= P(S_{2k} = 0)\frac{|r|}{n-k}P(S_{2n-2k} = 2r)$.

   (iii) Show that the joint distribution of (γ, B_1) has pdf $\frac{1}{2\pi}\frac{|x|}{t^{\frac{1}{2}}(1-t)^{\frac{3}{2}}}e^{-\frac{x^2}{2(1-t)}}$, $0 < t < 1, x \in \mathbb{R}$. [*Hint*: Use the multidimensional version of the local limit theorem; see Remark 16.3 and the local (central) limit theorem (Proposition 16.7) for simple symmetric random walk.]

10. Prove Corollary 18.7. [*Hint*: Let $U_n' = \#\{j \leq n : S_j > 0\}$, and check that $P(\tau^{(2n)} = 2k) = P(U_{2n}' = 2k)$ for $k = 0, 1, \ldots, n$ by expressing the respective events in terms of the random walk sums, noting independence of the segments of the walk (possibly viewed as a reverse path) before and after $2k$. In the first case, it will readily follow from this that for $1 \leq k \leq n$ $P(\tau^{(2n)} = 2k) = \frac{1}{2}P(S_{2k} = 0)P(\tau^{(2n-2k)} = 0), 1 \leq k \leq n$, while in the second, one also sees that $P(U_{2n}' = 2k) = \frac{1}{2}P(S_{2k} = 0)P(\tau^{(2n-2k)} = 0)$. The case $k = 0$ is immediate. Note that the difference between the functionals U_n and U_n' of the polygonal walk $\tilde{X}^{(n)}$ vanishes in the limit.]
11. Let $\{B_t : t \geq 0\}$ be standard Brownian motion starting at 0. Let $s < t$. Show that the probability that $\{B_t : t \geq 0\}$ has at least one zero in (s, t) is given by $(2/\pi)\cos^{-1}(s/t)^{1/2}$. [*Hint*: Let $\rho(x) = P(\{B_t : t \geq 0\}$ has at least one zero in $(s, t)|B_s = x)$. Then for $x > 0$,

$$\rho(x) = P\left(\min_{s\leq r\leq t} B_r \leq 0|B_s = x\right) = P\left(\max_{s\leq r\leq t} B_r \geq 0|B_s = -x\right)$$
$$= P\left(\max_{s\leq r\leq t} B_r \geq x|B_s = 0\right) = P(\tau_x \leq t - s).$$

Likewise for $x < 0$, $\rho(x) = P(\tau_{-x} \leq t - s)$. Thus $P(\{B_t : t \geq 0\}$ has at least one zero in $(s, t)) = \mathbb{E}\rho(B_s) = \int_0^\infty \rho(x)(\frac{2}{\pi s})^{1/2}e^{-\frac{1}{2s}x^2}dx$.]

Chapter 19
Brownian Motion on the Half-Line: Absorption and Reflection

Two important Markov processes derived from Brownian motion starting from a point $x \geq 0$ are (i) Brownian motion absorbed at zero and (ii) Brownian motion reflected at zero. The precise definitions of these two processes are given and their structure delineated, including the Markov property and the computation of the transition probabilities.

Let $B = \{B_t : t \geq 0\}$ denote a one-dimensional standard Brownian motion on $(-\infty, \infty)$, defined on a probability space $(\Omega, \mathcal{F}, P)$. We will write $B^{(x)} = \{B_t^{(x)} : t \geq 0\}$ for the standard Brownian motion starting at x and sometimes use P_x and $\mathbb{E}_x$ for probabilities and expected values associated with $B^{(x)}$.

There are two special Markov processes derived from $B^{(x)}$, $x \geq 0$, which we wish to single out in this chapter. Specifically these processes arise by imposing either an *absorbing* or a *reflecting* boundary, respectively, at $x = 0$.

Definition 19.1. Let $\tau_0 = \inf\{t > 0 : B_t^{(x)} = 0\}$, where $B^{(x)}$ is Brownian motion starting at $x \geq 0$. The stochastic process $\overline{B}^{(x)} = \{\overline{B}_t^{(x)} : t \geq 0\}$ with state space $[0, \infty)$ given by

$$\overline{B}_t^{(x)} := \begin{cases} B_t^{(x)} & \text{if } t \leq \tau_0 \\ 0 & \text{if } t \geq \tau_0 \end{cases} \equiv B_{t \wedge \tau_0}^{(x)}, \quad t \geq 0,$$

is referred to as (standard) *Brownian motion* on $[0, \infty)$ starting from $x \geq 0$ with absorbing boundary at 0.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_19

Remark 19.1. Let $\tau_0 : C[0,\infty) \to [0,\infty]$ be defined by $\tau_0(\omega) = \inf\{t \geq 0 : \omega(t) = 0\}$. Then $\tau_0 = \tau_0(B^{(x)})$ above depends on the initial point $x \geq 0$.

It is easy to check that in the case $x = 0$, one has $\tau_0 = 0$ with probability one and therefore $\overline{B}_t^{(0)} = 0, t \geq 0$, almost surely (Exercise 2). Notice that since $P(\overline{B}_t^{(x)} = 0) = P(\tau_0 \leq t) > 0$, the transition probabilities will not be absolutely continuous at the boundary 0. On the other hand, for a Borel set $A \subset (0,\infty)$, one has $P(\overline{B}_t^{(x)} \in A) = P(B_t^{(x)} \in A, \tau_0 > t) \leq P(B_t^{(x)} \in A)$ and, therefore, for $t > 0$, the distribution of $\overline{B}_t^{(x)}$ restricted to the positive half-line $(0,\infty)$ inherits absolute continuity from the distribution of $B_t^{(x)}$. The density is computed in the proof of the following result.

Proposition 19.1. The (standard) *Brownian motion absorbed at* 0 starting from $x \geq 0$, $\overline{B}^{(x)}$, is a Markov process with continuous sample paths on the state space $[0,\infty)$ having homogeneous transition probabilities given by

$$\overline{p}(t;x,dy) = \frac{1}{\sqrt{2\pi t}}(e^{-\frac{(x-y)^2}{2t}} - e^{-\frac{(x+y)^2}{2t}})dy, \quad x, y > 0, t > 0,$$
$$\overline{p}(t;x,\{0\}) = P(\tau_0 \leq t), \quad x \geq 0, t > 0. \tag{19.1}$$

Moreover,

$$\mathbb{E}\overline{B}_t^{(x)} = x, \quad t \geq 0.$$

Proof. Continuity of the sample paths follows immediately from that of the Brownian motion. Let A be a Borel subset of $(0,\infty)$. Then for all $x > 0$, introducing the *shifted process* $\{(B_s^{(x)+})_t \equiv B_{s+t}^{(x)}\, t \geq 0\}$ and using the Markov property of $B^{(x)}$, one gets *on the event* $[\tau_0 > s]$

$$\begin{aligned}
&P(\overline{B}_{s+t}^{(x)} \in A, \tau_0 > s \mid \sigma\{B_u^{(x)}\, 0 \leq u \leq s\}) \\
&\quad = P(B_{s+t}^{(x)} \in A, \tau_0(B^{(x)}) > s+t \mid \sigma\{B_u^{(x)}\, 0 \leq u \leq s\}) \\
&\quad = \mathbf{1}_{[\tau_0(B^{(x)})>s]} P(B_t^{(y)} \in A, \tau_0(B^{(y)}) > t)\big|_{y=B_s^{(x)}} \\
&\quad = \mathbf{1}_{[\tau_0(B^{(x)})>s]} P(\overline{B}_t^{(y)} \in A)\big|_{y=\overline{B}_s^{(x)}} \\
&\quad = P(\overline{B}_t^{(y)} \in A)\big|_{y=B_s^{(x)}}.
\end{aligned} \tag{19.2}$$

The second to third lines in (19.2) use the Markov property of $B^{(x)}$. Note that we have proved the Markov property with respect to the larger filtration $\{\sigma(B_u^{(x)}) : 0 \leq u \leq s\}, s \geq 0$ (Exercise 7).

For $A = \{0\}$, the Markov property may be checked by taking $A = (0, \infty)$ and by complementation. This establishes the Markov property of $\overline{B}^{(x)}$. To compute the transition probabilities, let $s, t \geq 0$, and $x > 0$, and apply the reflection principle for Brownian motion as follows for Borel $A \subset (0, \infty)$:

$$\begin{aligned}\overline{p}(t; x, A) &= P(\overline{B}_t^{(x)} \in A)\\ &= P(B_t^{(x)} \in A, \tau_0 > t | \overline{B}_0 = x)\\ &= P(B_t^{(x)} \in A) - P(B_t^{(x)} \in A, \tau_0 \leq t)\\ &= P(B_t^{(x)} \in A) - P(B_t^{(x)} \in -A),\end{aligned} \tag{19.3}$$

where $-A := \{-y : y \in A\}$. The asserted formula for the density on the positive half-line now follows by a change of variable in $P(B_t^{(x)} \in -A) = \int_{-A} p(t; x, y)dy = \int_A p(t; x, -y)dy$. Finally, $\mathbb{E}\overline{B}_t^{(x)} = x$ follows by the indicated integration $\int_{(0,\infty)} y\overline{p}(t; x, y)dy$ and is left as an exercise. ■

Remark 19.2. The same proof as above can be applied to the absorption of any one-dimensional Markov process having continuous sample paths.

One may note from the above proof that the absolutely continuous part of the transition probability $\overline{p}(t; x, dy)$, i.e.,

$$p^{(0)}(t; x, y) := \frac{1}{\sqrt{2\pi t}}(e^{-\frac{(x-y)^2}{2t}} - e^{-\frac{(x+y)^2}{2t}})dy, \;\; x, y \geq 0, t > 0, \tag{19.4}$$

is the (defective) density of "the process viewed prior to the time τ_0 it reaches 0."

For the case of reflection at $x = 0$, we will exploit the symmetries of the Brownian motion transition probability density $p(t; x, y)$; namely,

$$p(t; x, y) = p(t; -x, -y). \tag{19.5}$$

To simplify notation, we will now write P_x and $\mathbb{E}_x$ to indicate probabilities and expected values associated with the process starting at x.

Definition 19.2. For $x \geq 0$, the stochastic process $\tilde{B} = \{\tilde{B}_t : t \geq 0\}$ started at x with state space $[0, \infty)$ defined by

$$\tilde{B}_t := |B_t|, \quad t \geq 0,$$

is referred to as (standard) *Brownian motion* on $[0, \infty)$ starting from $x \geq 0$ with reflecting boundary at 0.

Proposition 19.2. The (standard) *Brownian motion reflected at* 0 starting from $x \geq 0$, $\tilde{B}$, is a Markov process with continuous sample paths on the state space $[0, \infty)$ having absolutely continuous homogeneous transition probabilities with density given by

$$\tilde{p}(t; x, y) = \frac{1}{\sqrt{2\pi t}}(e^{-\frac{(x-y)^2}{2t}} + e^{-\frac{(x+y)^2}{2t}}), \quad x, y \geq 0, t > 0.$$

Proof. Continuity of the sample paths follows from those of Brownian motion. In order to show that the process $\{Y_t := |B_t| : t \geq 0\}$ is a Markov process on $S = [0, \infty)$, consider an arbitrary real-valued bounded (Borel measurable or continuous) function g on $[0, \infty)$, and write $h(x) = g(|x|)$, $x \in \mathbb{R}$. One has

$$\begin{aligned}
&\mathbb{E}_x \left(g(|B_{s+t}|) \mid \sigma\{B_u\ 0 \leq u \leq s\}\right) \\
&= \mathbb{E}_x \left[\mathbb{E}\left(g(|B_{s+t}|) \mid \sigma\{B_u\ 0 \leq u \leq s\}\right)\right] \\
&= \mathbb{E}_x \left[\mathbb{E}_{B_s}\left(h(B_{s+t}) \mid \sigma\{B_u\ 0 \leq u \leq s\}\right)\right] \\
&= \mathbb{E}_x \left[\left(\int_{-\infty}^{\infty} h(y) p(t; x', y) dy\right)_{x'=B_s}\right] \\
&= \mathbb{E}_x \left[\left(\int_0^{\infty} g(z) \left(p(t; x', z) + p(t; x', -z)\right) dz\right)_{x'=B_s}\right] \\
&= \mathbb{E}_x \left[\left(\int_0^{\infty} g(z) \left(p(t; x', z) + p(t; -x', z)\right) dz\right)_{x'=B_s}\right] \\
&= \int_0^{\infty} g(z) \tilde{p}(t; |B_s|, z) dz, \qquad (19.6)
\end{aligned}$$

where

$$\tilde{p}(t; x, y) = p(t; x, y) + p(t; -x, y). \qquad (19.7)$$

This proves that $\{|B_t| : t \geq 0\}$ is a time-homogeneous Markov process with the transition probability density $\tilde{p}$, with respect to the filtration $\sigma(B_u : 0 \leq u \leq s)$, $s \geq 0$, larger than the standard filtration $\sigma(|B_u| : 0 \leq u \leq s)$, $s \geq 0$ (Exercise 1). ■

Recalling Corollary 7.11, it is possible to obtain a remarkable alternative representation of reflecting Brownian motion due to Paul Lévy as follows.

Theorem 19.3 *(Lévy's Representation of Reflecting Brownian Motion).* Let $B^{(x)}$ denote a standard Brownian motion starting at $x \geq 0$, and let $\tau_0 := \inf\{t : B_t^{(x)} = 0\}$. Define $R_t = B_t^{(x)}, 0 \leq t < \tau_0$, and $R_t = M_t^{(0)} - B_t^{(x)}, t \geq \tau_0$, where $M_t^{(0)} = \max\{B_s^{(x)} : \tau_0 \leq s \leq t\}$. Then the two stochastic processes R and $|B^{(x)}|$ have the same distribution.

Proof. Note that regardless of the starting point $x \geq 0$, $R_t = |B_t|$ for $t \leq \tau_0$, $B = B^{(0)}$. In particular, by the strong Markov property for Brownian motion, it suffices to prove the theorem for the case $x = 0$. To show that R and $|B|$ have the same distribution, we will show that R is a Markov process with the same transition probabilities as $|B|$ computed above. Since $\sigma\{R_s : s \le t\} \subset \mathcal{F}_t = \sigma\{B_s : s \le t\}$, it suffices to establish the Markov property relative to the past σ-fields $\mathcal{F}_t$, $t \ge 0$. For this, one has, for $r \ge 0$, $s \le t$,

$$\begin{aligned}
&P_0(R_t \le r|\mathcal{F}_s)\\
&\quad = P_0(M_s^{(0)} - B_t \le r, \max_{u \le t-s}(B_s^+)_u - B_t \le r|\mathcal{F}_s)\\
&\quad = P_0(R_s - (B_t - B_s) \le r, \max_{u \le t-s}\{(B_s^+)_u - B_s\} - (B_t - B_s) \le r|\mathcal{F}_s).\\
&\quad = P_0(z - B_{t-s} \le r, M_{t-s} \le B_{t-s} + r)\big|_{z=R_s},
\end{aligned} \tag{19.8}$$

since $M^{(0)} = M$ is the running max for B when $x = 0$, R_s is $\mathcal{F}_s$-measurable and is independent of $B_t - B_s$ and $B_s^+ - B_s$ and, jointly, the latter two are distributed as (B_{t-s}, M_{t-s}). Therefore,

$$P_0(R_s \le r|\mathcal{F}_s) = P_0(B_{t-s} \ge z - r, M_{t-s} \le B_{t-s} + r)|_{z=R_s}. \tag{19.9}$$

Now the joint density of M_{t-s} and B_{t-s} is given at $M_{t-s} = a$, $B_{t-s} = b$ by (see Corollary 7.12)

$$f(t-s; a, b) = -\frac{\partial}{\partial a}\frac{\partial}{\partial b}(1 - \Phi_{t-s}(2a - b)) = -\frac{\partial}{\partial a}\varphi_{t-s}(2a - b), \tag{19.10}$$

where Φ_{t-s} is the distribution function of B_{t-s}, i.e., normal with mean zero and variance $t - s$, and φ_{t-s} is its density. Hence

$$\begin{aligned}
&P_0(B_{t-s} \ge z-r, M_{t-s} \le B_{t-s}+r)|_{z=R_s}\\
&\quad = \int_{z-r}^{\infty}\{\int_b^{b+r} -\frac{\partial}{\partial a}\varphi_{t-s}(2a-b)da\}db|_{z=R_s}\\
&\quad = \int_{z-r}^{\infty}\{\varphi_{t-s}(b)-\varphi_{t-s}(b+2r)\}db|_{z=R_s}\\
&\quad =\{P_0(B_{t-s} > z-r)-P_0(B_{t-s} \ge z+r)\}|_{z=R_s}\\
&\quad =P_0(z-r \le B_{t-s} \le z+r)|_{z=R_s}=P_0(|B_{t-s}-z| \le r)|_{z=R_s}.
\end{aligned} \tag{19.11}$$

Thus the reflecting Brownian motion and $M_t - B_t$, $t \ge 0$, have the same transition probability. ■

Remark 19.3. Another way to view Lévy's representation for the process starting from zero is $R_t = \tilde{B}_t + M_t^{(0)}$, where $\tilde{B} = -B$ is a Brownian motion and $M_t^{(0)}$ is thought of as a "reflective forcing" that steers it away from the origin. This perspective was successfully developed into a dynamic representation by Skorokhod via stochastic differential equations.

A more comprehensive theory of Markov processes on domains with boundaries can be developed in a theory for diffusions.

Exercises

1. Let $\{X_t : t \geq 0\}$ be a Brownian motion starting at 0 with zero drift and diffusion coefficient $\sigma^2 > 0$. Define $Y_t = |X_t|, t \geq 0$.

   (i) Calculate $\mathbb{E}Y_t$, Var Y_t.
   (ii) Does $\{Y_t : t \geq 0\}$ have stationary increments?
   (iii) Does $\{Y_t : t \geq 0\}$ have uncorrelated increments? [*Hint*: Use the Markov property to calculate $\mathbb{E}|Y_s Y_t|$ for $s < t$. Assume $\mathbb{E}Y_s(Y_t - Y_s) = \mathbb{E}Y_s\mathbb{E}(Y_t - Y_s)$ and obtain contradiction.]
   (iv) Is $\{Y_t : t \geq 0\}$ a process with independent increments?

2. (i) Show that for standard Brownian motion starting from $x = 0$, one has $\tau_0 = 0$ with probability one and therefore $\overline{B}_t^{(x)} = 0, t \geq 0$, almost surely. [*Hint*: Note that the hitting time of 0 starting from x is the hitting time of $-x$ starting from 0, and use Proposition 16.3 in Chapter 16 to compute $P_x(\tau_0 > \epsilon)$ $(\epsilon > 0)$ in the limit as $x \downarrow 0$ using the dominated convergence theorem.]
   (ii) Use the transition probabilities to compute $\mathbb{E}_x\overline{B}_t^{(x)} = x$ for Brownian motion starting at x with absorption at zero.
3. Let $m(t, x) = \mathbb{E}_x\overline{B}_t^{(x)} = \int_{(0,\infty)} y\overline{p}(t, x, y)dy$, and show that $\frac{\partial m}{\partial t} = \frac{1}{2}\frac{\partial^2 m}{\partial x^2}$ together with initial condition $m(0, x) = x$ and boundary condition $m(t, 0) = 0, t \geq 0$. Use this for an alternative approach to show that $m(t, x) = x$.
4. Use (19.4) to obtain the distribution of the first passage time to zero starting from $x > 0$ for standard Brownian motion.
5. (i) Show that the location of the maximum of $B_t, 0 \leq t \leq 1$ is a.s. unique in $(0, 1)$. [*Hint*: Let $M_s^- = \max_{0\leq u\leq s} B_u$, $M_t^+ = \max_{t\leq u\leq 1} B_u$, $s < t$. Then argue that $P(M_s^- = M_t^+) = 0$ since $[M_s^- = M_t^+] = [B_t - B_s = (M_s^- B_s) - (M_t^+ - B_t)]$, and $B_t - B_s$ has an absolutely continuous distribution, independent of $\sigma(M_s^- B_s, M_t^+ - B_t)$.]
   (ii) Use Theorem 19.3 to show the equivalence of the arcsine law for the argmax of Brownian motion with that of the time of the last zero. [*Hint*: The zeroes of B and $|B|$ are the same.]
6. The goal is to give an alternative derivation of the arcsine law for the last zero γ before $t = 1$ of Brownian motion given in Chapter 18, Theorem 18.3, and hence a limit formula for the random walk via the invariance principle, using strong Markov property methods of the type employed in Chapter 7. Complete the following steps. Let $\tau_x = \inf\{t > 0 : B_t^{(x)} = x\}$.

   (i) Show that $P(B_t^{(x)} \in dy, \tau_0 > t) = \frac{1}{\sqrt{2\pi t}}\{e^{\frac{-(x-y)^2}{2t}} - e^{\frac{-(x+y)^2}{2t}}\}dy, \quad x, y, t > 0.$

(ii) For $0 < s < 1$, $x, y > 0$, show $P(\gamma \leq s, |B_s| \in dx, |B_1| \in dy)$ $= \sqrt{\frac{2}{\pi s}}e^{-\frac{x^2}{2s}}\{\frac{1}{\sqrt{2\pi(1-s)}}e^{-\frac{(x-y)^2}{2(1-s)}} - e^{-\frac{(x+y)^2}{2(1-s)}}\}dydx$. [*Hint*: Express as $P(|B_s| \in dx, |B_u| \neq 0, s \leq u \leq 1, |B_1| \in dy) = P(|B_s| \in dx)P_x(B_{1-s} \in dy, \tau_0 > 1-s)$.]

(iii) Show that $P(\gamma \leq s, |B_1| \in dy) = \frac{2}{\pi}e^{-\frac{y^2}{2}}\{\int_0^{\sqrt{\frac{s}{(1-s)}}y} e^{-\frac{u^2}{2}}du\}dy$. [*Hint*: Integrate the difference terms of (ii) with respect to x by completing the square and a change of variable.]

(iv) Show that $P(\gamma \in ds, |B_1| \in dy) = \frac{y}{\pi\sqrt{s(1-s)^3}}e^{-\frac{y^2}{2(1-s)}}dsdy$.

(v) Complete the derivation of (18.7) in Theorem 18.3 by the indicated integrations.

7. Show the Markov property of a process X_t, $t \in \mathbb{Z}_+$, or $t \in [0, \infty)$, with respect to a filtration $\mathcal{G}_t$, $t \geq 0$, larger than $\mathcal{F}_t = \sigma(X_u : 0 \leq u \leq t)$, $t \geq 0$, implies the usual Markov property with respect to $\mathcal{F}_t$, $t \geq 0$.

Chapter 20
The Brownian Bridge

The Brownian bridge, or tied-down Brownian motion, is derived from the standard Brownian motion on [0, 1] started at zero by constraining it to return to zero at time $t = 1$. A precise definition is provided and its (Gaussian) distribution is computed. The Brownian bridge arises in a wide variety of contexts. An application is given to a derivation of the Kolmogorov–Smirnov statistic in non-parametric statistics in this chapter. An application to the Hurst statistic in special topics Chapter 27, to mention a few.

Certain functionals of Brownian motion paths occur naturally and are of interest in their own right.

Definition 20.1. Let $\{B_t : t \geq 0\}$ be a standard Brownian motion starting at zero. The stochastic process $\{B_t^* : 0 \leq t \leq 1\}$ defined by

$$B_t^* := B_t - tB_1, \quad 0 \leq t \leq 1, \tag{20.1}$$

is called the *Brownian bridge*; see Figure 20.1.

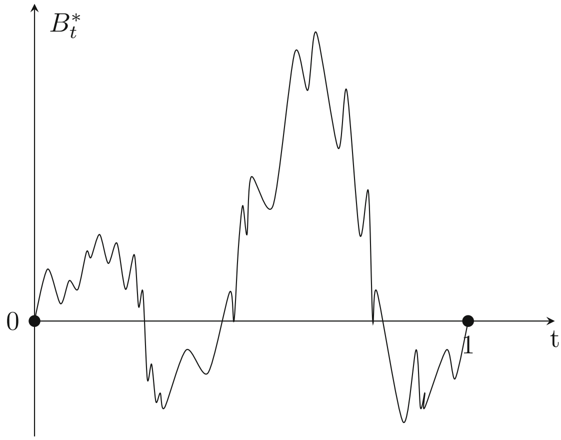
Observe that $B_t^* = B_t - tB_1$ vanishes for $t = 0$ and $t = 1$. Another name for Brownian bridge is the *tied-down Brownian motion*.

Proposition 20.1 *(Structure of the Brownian Bridge).* $\{B_t^* : 0 \leq t \leq 1\}$ is a Gaussian process and has a.s. continuous sample paths from Brownian motion, with $\mathbb{E}B_t^* = 0$, and $\text{Cov}(B_s^*, B_t^*) = s(1-t), 0 \leq s \leq t \leq 1$.

Proof. Since $\{B_t : t \geq 0\}$ is a Gaussian process with independent increments, it is simple to check that the finite dimensional distributions of the Brownian bridge are

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_20

Fig. 20.1 Brownian bridge



also Gaussian since $(B^*_{t_1}, \ldots, B^*_{t_k})$ is a linear transformation of the jointly Gaussian vector $(B_{t_1}, \ldots, B_{t_k}, B_1)$ for arbitrary fixed $0 < t_1 < \cdots, t_k \leq 1, k \geq 1$. Continuity of paths and the zero mean are directly inherited from the corresponding properties of the Brownian motion. Similarly,

$$\begin{aligned}\mathrm{Cov}(B^*_s, B^*_t) &= \mathrm{Cov}(B_s, B_t) - t\,\mathrm{Cov}(B_s, B_1) - s\,\mathrm{Cov}(B_t, B_1) \\ &\quad + st\,\mathrm{Cov}(B_1, B_1) \\ &= s - ts - st + st = s(1-t), \qquad \text{for } s \leq t.\end{aligned}$$

■

From Proposition 20.1 one can also explicitly write down the joint Gaussian density of $(B^*_{t_1}, B^*_{t_2}, \ldots, B^*_{t_k})$ for arbitrary $0 < t_1 < t_2 < \cdots < t_k < 1$ (Exercise 1).

Example 1 *(Application to Non-parametric Statistics)*. The remainder of this chapter is devoted to an important application of the Brownian bridge that arises in the asymptotic (large sample) theory of statistics. To explain this application, let us consider a sequence of real-valued i.i.d. random variables $Y_1, Y_2, \ldots,$ having a (common) distribution function F. The nth *empirical distribution* is the discrete probability distribution on the line assigning a probability $1/n$ to each of the n values $Y_1, Y_2, \ldots, Y_n$. The corresponding distribution function F_n is called the (nth) *empirical distribution function*,

$$F_n(t) = \frac{1}{n}\#\{j : 1 \leq j \leq n,\ Y_j \leq t\} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{[Y_j \leq t]}, -\infty < t < \infty, \tag{20.2}$$

where $\#A$ denotes the cardinality of the set A. Suppose $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ is the ordering of the first n observations. Note that $\{F_n(t) : t \geq 0\}$ is for each n a stochastic process, referred to as an *empirical process*. Also, $\mathbb{E}\mathbf{1}_{[Y_j \le t]} = F(t)$ and, for $t_1 \le t_2$,

$$\operatorname{Cov}\left(\mathbf{1}_{[Y_j \le t_1]}, \mathbf{1}_{[Y_k \le t_2]}\right) \begin{cases} 0, & \text{if } j \neq k \\ F(t_1)(1 - F(t_2)), & \text{if } j = k. \end{cases} \tag{20.3}$$

It follows from the central limit theorem that

$$n^{-1/2}\left(\sum_{j=1}^{n} \mathbf{1}_{[Y_j \le t]} - nF(t)\right) = \sqrt{n}(F_n(t) - F(t))$$

is asymptotically (as $n \to \infty$) Gaussian with mean zero and variance $F(t)(1 - F(t))$. For $t_1 < t_2 < \cdots < t_k$, the multidimensional central limit theorem applied to the i.i.d. sequence of k-dimensional random vectors $(\mathbf{1}_{[Y_j \le t_1]}, \mathbf{1}_{[Y_j \le t_2]}, \ldots, \mathbf{1}_{[Y_j \le t_k]})$ shows that $(\sqrt{n}(F_n(t_1) - F(t_1)), \sqrt{n}(F_n(t_2) - F(t_2)), \ldots, \sqrt{n}(F_n(t_k) - F(t_k)))$ is asymptotically (k-dimensional) Gaussian with zero mean and dispersion matrix $\sum = ((\sigma_{ij}))$, where

$$\sigma_{ij} = \operatorname{Cov}(\mathbf{1}_{[Y_j \le t_i]}, \mathbf{1}_{[Y_j \le t_j]}) = F(t_i)(1 - F(t_j)), \qquad \text{for } t_i \le t_j. \tag{20.4}$$

In the special case of observations from the uniform distribution on [0, 1], one has

$$F(t) = t, \quad 0 \le t \le 1, \tag{20.5}$$

so that the finite dimensional distributions of the stochastic process $\{\sqrt{n}(F_n(t) - t) : 0 \le t \le 1\}$ converge to those of the Brownian bridge as $n \to \infty$. As in the case of the functional central limit theorem (Theorem 17.2), probabilities of many infinite-dimensional events of interest also converge to those of the Brownian bridge. In this example we will restrict our considerations to the particular functional $\sup_{0 \le t \le 1} |\sqrt{n}(F_n(t) - t)|$.

The precise result that we will prove here is as follows. The symbol $\stackrel{d}{=}$ below denotes equality in distribution.

Proposition 20.2. Let $Y_1, Y_2, \ldots$ be i.i.d. uniform on [0, 1] and let, for each $n \ge 1$, $\{F_n(t) : 0 \le t \le 1\}$ be the corresponding empirical process based on $Y_1, \ldots, Y_n$. Then the statistic $D_n := \sqrt{n} \sup_{0 \le t \le 1} |F_n(t) - t|$ converges in distribution to $\sup_{0 \le t \le 1} |B_t^*|$ as $n \to \infty$.

Proof. We will make use of the fact that the distribution of the order statistic $(Y_{(1)}, \ldots, Y_{(n)})$ of n i.i.d. random variables $Y_1, Y_2, \ldots$ from the uniform distribution on [0, 1] is the same as the distribution of the ratios $(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}})$, where $S_k = T_1 + \cdots + T_k$, $k \ge 1$, and $T_1, T_2, \ldots$ is an i.i.d. sequence of (mean one) exponentially distributed random variables.

To show this note that (i) the joint density of $(T_1, T_2, \ldots, T_{n+1})$ is $f_1(t_1, t_2, \ldots, t_{n+1}) = e^{-\sum_{i=1}^{n+1} t_i}\mathbf{1}_{(0,\infty)^{n+1}}(t_1, \ldots, t_{n+1})$, (ii) by a linear transformation of Jacobian one, the joint density of $(S_1, S_2, \ldots, S_{n+1})$ is therefore seen to be $f_2(s_1, s_2, \ldots, s_{n+1}) = e^{-s_{n+1}}\mathbf{1}_{R_{n+1}}(s_1, \ldots, s_{n+1})$ where R_{n+1} is the set $R_{n+1} := \{(s_1, \ldots, s_{n+1}) : 0 < s_1 < s_2 < \cdots < s_{n+1} < \infty\}$, from which it follows that the conditional density of $(S_1, \ldots, S_n)$, given $S_{n+1} = s$, is $f_3(s_1, \ldots, s_n \mid s) = \frac{1}{n!s^n}\mathbf{1}_{R_{n+1}}(s_1, s_2, \ldots, s_n, s)$, and, finally, (iii) the conditional density of $(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}})$, given $S_{n+1} = s$, is $f_4(u_1, u_2, \ldots, u_n \mid s) = \frac{1}{n!}\mathbf{1}_{R_{n+1}}(u_1, u_2, \ldots, u_n, 1)$. Since f_4 is independent of s, it is also the (unconditional or marginal) density of $(\frac{S_1}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}})$. Clearly, f_4 is also the joint density of $(Y_{(1)}, \ldots, Y_{(n)})$.

In the notation introduced above, now consider the following trick: Since $F_n(t) = k/n$ for $Y_{(k)} \le t < Y_{(k+1)}$ and $F_n(Y_{(k+1)}) = (k+1)/n$, $\sup\{|F_n(t) - t| : Y_{(k)} \le t < Y_{(k+1)}\} = \max\{|Y_{(k)} - \frac{k}{n}|, |Y_{(k+1)} - \frac{k}{n}|\}$ (Exercise 2),

$$
\begin{aligned}
D_n &:= \sqrt{n} \sup_{0 \le t \le 1} |F_n(t) - t| = \sqrt{n} \max_{k \le n} \left| Y_{(k)} - \frac{k}{n} \right| + \mathrm{O}(n^{-\frac{1}{2}}) \\
&\stackrel{d}{=} \sqrt{n} \max_{k \le n} \left| \frac{S_k}{S_{n+1}} - \frac{k}{n} \right| + \mathrm{O}(n^{-\frac{1}{2}}) \\
&= \frac{n}{S_{n+1}} \max_{k \le n} \left| \frac{S_k - k}{\sqrt{n}} - \frac{k}{n} \frac{S_{n+1} - n}{\sqrt{n}} \right| + \mathrm{O}(n^{-\frac{1}{2}}) \\
&\approx \frac{n}{S_{n+1}} \sup_{0 \le t \le 1} |X_t^{(n)} - tX_1^{(n)}|,
\end{aligned} \tag{20.6}
$$

where $X_t^{(n)} = S_{[nt]}/\sqrt{n}$, and $\approx$ indicates that the difference between its two sides goes to zero in probability as $n \to \infty$. By the SLLN, $n/(S_{n+1}) \to 1$ a.s. as $n \to \infty$. The result now follows from the FCLT (Theorem 17.2), and the definition of Brownian bridge. ■

Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence having a (common) distribution function F that is continuous on the real number line. Note that in the case that F is strictly increasing on an interval (a, b) with $F(a) = 0$, $F(b) = 1$, one has for $0 < t < 1$,

$$
P(F(Y_k) \le t) = P(Y_k \le F^{-1}(t)) = F(F^{-1}(t)) = t, \tag{20.7}
$$

so the sequence $U_1 = F(Y_1)$, $U_2 = F(Y_2)$, $\ldots$ is i.i.d. uniform on $[0, 1]$ (Exercise 3). Let F_n be the empirical distribution function of $Y_1, \ldots, Y_n$, and G_n that of $U_1, \ldots, U_n$. Then, since the proportion of Y_k's, $1 \le k \le n$, that do not exceed t coincides with the proportion of U_k's, $1 \le k \le n$, that do not exceed $F(t)$, we have

$$
\sqrt{n}[F_n(t) - F(t)] = \sqrt{n}\,[G_n(F(t)) - F(t)], \quad a \le t \le b. \tag{20.8}
$$

If $a = -\infty$ ($b = +\infty$), the index set $[a, b]$ for the process is to exclude a (b).

It now follows from (20.8) that the *Kolmogorov–Smirnov statistic* defined by

$$D_n := \sup\{\sqrt{n}|F_n(t) - F(t)| : a \le t \le b\} \tag{20.9}$$

satisfies

$$\begin{aligned} D_n &:= \sup_{a \le t \le b} \sqrt{n}|F_n(t) - F(t)| = \sup_{a \le t \le b} \sqrt{n}\,|G_n(F(t)) - F(t)| \\ &= \sup_{0 \le t \le 1} \sqrt{n}\,|G_n(t) - t|\,. \end{aligned} \tag{20.10}$$

Thus Proposition 20.2 has the following basic consequence.

Corollary 20.3. The distribution of D_n is the same (namely that obtained under the uniform distribution) for all continuous F. In particular, regardless of such F, D_n converges in distribution as $n \to \infty$ to

$$D := \sup_{0 \le t \le 1} |B_t^*|. \tag{20.11}$$

Moreover,

$$P(D > d) = 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2d^2}, \quad d > 0. \tag{20.12}$$

Proof. The main part of the proof precedes the statement. The calculation of the distribution of D is outlined in Exercise 6(iii). ■

Remark 20.1. The common distribution of D_n may be tabulated for small and moderately large values of n. These results are often used to test the *statistical hypothesis* that observations $Y_1, Y_2, \dots, Y_n$ are from a specified distribution with a continuous distribution function F. If the observed value, say d, of D_n is so large that the probability (approximated by (20.12) for large n) is very small for a value of D_n as large as or larger than d to occur (under the assumption that $Y_1, \dots, Y_n$ do come from F), then the hypothesis is rejected.

In closing, note that by the strong law of large numbers, $F_n(t) \to F(t)$ as $n \to \infty$, *with probability one*. From Propositions 20.2 and (20.10), it follows that

$$\sup_{-\infty < t < \infty} |F_n(t) - F(t)| \to 0 \qquad \text{in probability as } n \to \infty. \tag{20.13}$$

In fact, it is possible to show that the uniform convergence in (20.13) is also *almost sure* (Exercise 7). This stronger result is known as the *Glivenko–Cantelli lemma.*

Exercises

1. Compute the joint density of $(B^*_{t_1}, B^*_{t_2}, \ldots, B^*_{t_k})$.
2. Show that $\sqrt{n} \sup_{0\le t\le 1} |F_n(t) - t| = \sqrt{n} \max_{k\le n} \left| Y_{(k)} - \frac{k}{n} \right| + O(n^{-\frac{1}{2}})$.
3. Suppose that F is an arbitrary distribution function (not necessarily continuous). Define an inverse to F as $F^{-1}(u) = \inf\{x : F(x) > u\}$. Show that if U is uniform on [0, 1] then $X = F^{-1}(U)$ has a distribution function F.
4. Let $\{B_t : t \ge 0\}$ be standard Brownian motion starting at 0 and let $B^*_t = B_t - tB_1, 0 \le t \le 1$.

   (i) Show that $\{B^*_t : 0 \le t \le 1\}$ is independent of B_1.

   (ii) Give a construction of the standard Brownian motion on [0, 1] from the Brownian bridge and an independent Gaussian random variable. [*Hint*: Use (i).]

5. Show that (i) $B_t = (1+t)B^*_{\frac{t}{1+t}}, t \ge 0$ is distributed as the standard Brownian motion starting at zero, and (ii) $B^*_t = (1-t)B_{\frac{t}{1-t}}, 0 \le t < 1$, $B^*_1 = 0$, is distributed as Brownian bridge.
6. Let $\{B_t : t \ge 0\}$ be a standard Brownian motion starting at 0 and let $\{B^*_t : 0 \le t \le 1\}$ be the Brownian bridge.

   (i) Show that for time points $0 \le t_1 < t_2 < \cdots < t_k \le 1$,

$$\lim_{\varepsilon\to 0} P(B_{t_i} \le x_i, i = 1, 2, \cdots, k \mid -\varepsilon < B_1 < \varepsilon)$$
$$= P(B^*_{t_i} \le x_i, i = 1, \cdots, k).$$

   Likewise, for conditioning on $B_1 \in D_\epsilon = [0, \varepsilon)$ or $D_\varepsilon = [-\epsilon, 0)$ the limit is unchanged. [*Hint*: Check $B^*_{t_1}, \ldots, B^*_{t_k}$ and B_1 are uncorrelated and hence independent. So for fixed Borel $G \subset \mathbb{R}$, $P([B^* \in F] \cap [B_1 \in G]) = P(B^* \in F)P(B_1 \in G)$ holds for a collection of sets containing the finite dimensional events, the latter comprising a $\pi-$system. Use the $\pi-\lambda$ theorem to conclude this equality for Borel subsets of $C[0, 1]$.]

   (ii) Show that the process $\{B^*_t : 0 \le t \le 1\}$ is the weak limiting distribution in (i) as $\varepsilon \to 0$. [*Hint*: According to Alexandrov's theorem[1] it suffices to check $\limsup_{\epsilon\to 0} P(B \in F| -\varepsilon < B_1 < \varepsilon) \le P(B^* \in F)$ for closed subsets F of $C[0, 1]$. Let ρ denote the uniform metric on $C[0, 1]$. Then, $\sup_{0\le t\le 1} |B^*_t - B_t| = |B_1|$, so $[|B_1| \le \delta, B \in F] \subset [B^* \subset F_\delta]$ where $F_\delta = \{\omega \in C[0, 1] : \rho(\omega, F) \le \delta\}$ for $\rho(\omega, F) = \inf\{\rho(\omega, \eta) : \eta \in F\}$. So for $\varepsilon < \delta$, $P(B \in F| -\varepsilon < B_1 < \varepsilon) \le P(B^* \in F_\delta| -\varepsilon < B_1 < \varepsilon) = P(B^* \in F_\delta)$.]

[1] See BCPT p.137.

(iii) Show that for $m^* = \inf_{0\le t\le 1} B_t^*$, $M^* = \sup_{0\le t\le 1} B_t^*$, $u < 0 < v$,

$$P(u < m^* \le M^* \le v)$$
$$= \sum_{k=-\infty}^{\infty} \exp\{-2k^2(v-u)^2\} - \sum_{k=-\infty}^{\infty} \exp\{-2[v + k(v-u)]^2\}.$$

[*Hint*: Express as a limit of the ratio of probabilities as in (i) and use Exercise 17(iii) of Chapter 17. Also, $\Phi(x, x+\varepsilon) = \varepsilon/(2\pi)^{1/2}\exp(-x^2/2) + o(1)$ as $\varepsilon \to 0$.]

(iv) Prove

$$P\left(\sup_{0\le t\le 1} |B_t^*| \le y\right) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2y^2}, \qquad y > 0.$$

[*Hint*: Take $u = -v$ in (iii).]

(v) $P(M^* < v) = 1 - e^{-2v^2}$, $v > 0$. [*Hint*: Use Exercise 17(iv) of Chapter 17 for the ratio of probabilities described in (i).]

7. Prove the *Glivenko–Cantelli Lemma* by justifying the following steps:

(i) For each t, the event $[F_n(t) \to F(t)]$ has probability one.
(ii) For each t, the event $[F_n(t^-) \to F(t^-)]$ has probability one.
(iii) Let $\tau(y) = \inf\{t : F(t) \ge y\}$, $0 < y < 1$. Then $F(\tau(y)^-) \le y \le F(\tau(y))$.
(iv) Let $D_{m,n} = \max_{1\le k\le m}\{|F_n(\tau(k/m)) - F(\tau(k/m))|, |F_n(\tau(k/m)^-) - F(\tau(k/m)^-)|\}$. Then, by considering the cases $\tau\left(\frac{k-1}{m}\right) \le t < \tau\left(\frac{k}{m}\right)$, $t < \tau\left(\frac{1}{m}\right)$ or if $t \ge \tau(1)$, show that $\sup_t |F_n(t) - F(t)| \le D_{m,n} + \frac{1}{m}$.[*Hint*: Check that both $F_n(t) - F(t) \le D_{m,n} + \frac{1}{m}$ and $F_n(t) - F(t) \ge -D_{m,n} - \frac{1}{m}$ by using monotonicity, followed by adding and subtracting appropriate terms.]
(v) $C = \bigcup_{m=1}^{\infty}\bigcup_{k=1}^{m}[F_n(\tau(k/m)) \not\to (F(\tau(k/m))] \cup [F_n(\tau(k/m)^-) \not\to F(\tau(k/m)^-)]$
has probability zero, and for $\omega \in C^c$ and each in $m \ge 1$ $D_{m,n}(\omega) \to 0$ as $n \to \infty$.
(vi) $\sup_t |F_n(t,\omega) - F(t)| \to 0$ as $n \to \infty$ for $\omega \in C^c$.

8. (*The Gnedenko–Korolyuk Formula*) Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be two independent i.i.d. random samples with continuous distribution functions F and G, respectively. To test the *null hypothesis* that both samples are from the same population (i.e., $F = G$), let F_n and G_n be the respective *empirical distribution functions* and consider the statistic $D_n = \sup_x |F_n(x) - G_n(x)|$. Under the null hypothesis, $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are $2n$ i.i.d. random variables with

the common distribution F. Verify that under the null hypothesis one has the following:

(i) The distribution of D_n does not depend on F and can be explicitly calculated according to the formula

$$P\left(D_n \le \frac{m}{n}\right) = P\left(\max_{0\le k\le 2n} |S_k^*| \le m\right),$$

where S_k^* is a simple symmetric random walk starting at 0 and tied down at 0 at $k = 2n$, i.e., random walk bridge. [*Hint*: By arranging $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ in increasing order as $Z_{(1)} < Z_{(2)} < \cdots < Z_{(2n)}$, each $Z_{(j)}$ may be regarded as $+1$-type if $Z_{(j)} = X_i$ for some i, or of -1-type otherwise, i.e., ± 1-types given by $\theta_j = \mathbf{1}[Z_{(j)} \text{ is type}+1] - \mathbf{1}[Z_{(j)} \text{ is type}-1]$, $j = 1, \ldots, 2n$. By symmetry, the types θ_j comprise a Bernoulli ± 1-valued exchangeable (symmetrically dependent) sequence for which all $\binom{2n}{n}$ arrangements are equally likely, and such that $S_k^* := \sum_{j=1}^k \theta_j$ is a simple symmetric walk tied down at $S_0^* = S_{2n}^* = 0$. The key observation is that for any value of S_j^* one has $2n|F_n(x) - G_n(x)| = |S_j^*|$ for $Z_{(j)} \le x < Z_{(j+1)}$. Thus $2nD_n = 2n \sup_x |F_n(x) - G_n(x)| = \max_{j\le 2n} |S_j^*|$.]

(ii) Find the analytic expression for the probability in (i). [*Hint*: Use Corollary 3.4 of Chapter 3 with $a = b = m$.]

(iii) Calculate the large-sample theory (i.e., asymptotic as $n \to \infty$) limit distribution of $\sqrt{n}D_n$. See Exercise 6(iii).

(iv) Show

$$P\left(\sup_x (F_n(x) - G_n(x)) < \frac{m}{n}\right) = 1 - \frac{\binom{2n}{n+m}}{\binom{2n}{n}}, \qquad m = 1, \ldots, n.$$

[*Hint*: Only one absorbing barrier occurs in the random walk approach.]

Chapter 21
Special Topic: Branching Random Walk, Polymers, and Multiplicative Cascades

The proof of the Kesten–Stigum theorem presented in Chapter 14 involved the application of size-bias methods to branching processes. Related techniques apply to the analysis of a natural class of multiplicative cascades, random polymer models, and to branching random walks. In this chapter we will introduce these three classes of models and provide some basic results for each; specifically a proof of the Kahane–Peyriére theorem[1] for multiplicative cascades based on distinguished path analysis (size-biasing), the infinite volume limit at critical strong disorder in Bolthausen's[2] conception of weak and strong disorder for tree polymers, and, lastly, for the Biggins–Kingman–Hammersley theorem[3] the calculation of the speed of the leftmost particle in branching random walk. The first example further illustrates the distinguished path analysis, the second introduces the derivative martingale, and the third introduces the many-to-one lemma.

Apart from their common underlying indexing by trees, a common feature shared by the models analyzed in this chapter is that their structure is governed by the occurrence of *large deviation* events. Size-bias, or tilting, of probabilities so that deviant events become the average is a standard tool of large deviation theory[4] extending at least back to Cramer and Chernoff, with important refinements of

[1] Kahane and Peyrere (1976).

[2] Bolthausen (1989, 1991)

[3] Biggins (1976); Kingman (1975); Hammersley (1974).

[4] See BCPT pp. 94–100, for an exposition to a variety of standard "concentration inequalities"derived along these lines.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_21

Bahadur and Ranga Rao, Varadhan and others. When applied in the context of branching it has become known as *distinguished path or spine decomposition* analysis.[5]

The following example provides a simple example of a multiplicative cascade for which non-degeneracy is readily obtained by previous theory.

***Example 1** (Multiplicative Cascades).* Statistical theories of turbulent fluids[6] lead to models for the distribution of energy dissipation by random measures which are characteristically highly variable (singular) with regions of spatial intermittency. A class of models which has its origins in Kolmogorov's statistical turbulence theory depicts the redistribution of energy as the result of a cascade of energies from large to small scales. Inspired by earlier ideas of Richardson (1922, page 66), Kolmogorov (1941, 1962) imagined a large scale stirring of the fluid which would result in the splitting off of smaller scale eddies which, in turn, would continue to split off at still smaller scales. Briefly, in the initial formulation Kolmogorov used (deterministic) scaling analysis[7] to argue that within a certain range of spatial scales where dissipation could be ignored, referred to as the *inertial range*, moments of velocity increments over lengths ℓ would scale with exponent $C_p\ell^{\frac{p}{3}}$. However, this was met with the historically famous objection by physicist Lev Landau that the intermittency in velocity fluctuations on these moments had not been taken into account by Kolmogorov. This led to the Kolmogorov (1962) and Obukhov (1962) refined hypothesis in which the lognormal distribution is explicitly incorporated into the cascade and yielding velocity moments having scaling exponent that can be expressed as $\frac{p}{3} + \frac{\mu}{18}(3p - p^2)$, where the parameter μ is sometimes referred to as the *intermittency correction*. One may view these scaling exponents as the slopes of plots of the logarithm of absolute p-th order moments of the velocity displacements versus logarithm of the lengths of the displacement; see (21.17). The linearity is associated with a form of multi-scale invariance.

This, in turn, led to the development of statistical cascade models of turbulence by mathematicians such as Mandelbrot (1974a, b), Kahane (1974), Kahane and Peyrere (1976) considered in the present treatment.

Remark 21.1. In a more recent development, She and Levesque (1994) formulated a (deterministic) second order, non-homogeneous linear difference equation for

---

[5] An early development in the analysis of the fine scale structure of non-degenerate cascades by size-bias change of measure by Peyrière (1974), also motivated Waymire and Williams (1994) to investigate the use of a distinguished path technique for proving the existence of non-degenerate limiting cascades. Lyons et al. (1995) applied these change of measure techniques to provide the distinguished path proof of the Kesten–Stigum theorem given in Chapter 14.

[6] There is also a rather large literature on the spatial structure of precipitation fields based on log-log plots of moments of rainfall of order h as a function of spatial scale that lead naturally to multiplicative cascade models, e.g., see Gupta and Waymire (1993); Salas et al. (2017).

[7] The article by Frisch (1991) is among the most readable accounts for mathematicians of the most essential aspects of the Kolmogorov statistical theory.

the velocity moments. This was subsequently shown by Dubrulle (1994), and independently She and Waymire (1995), to correspond to a statistical cascade model in which the lognormal distribution of Kolmogorov (1962) and Obukhov (1962) is replaced by the logPoison distribution.[8] This yields velocity moments with scaling exponent in the form $\alpha p + \lambda(1 - \beta^{\frac{p}{3}})$ with $\alpha = 1/9, \beta = 2/3, \lambda = 2$.

Let us begin with the simplest of such multiplicative cascade models. Consider a unit square $U = [0, 1)^2$ over which an amount ϵ_0 is distributed uniformly. Regard ϵ_0 as a (constant) density of a measure $\mu_0(d\mathbf{x}) = \epsilon_0 d\mathbf{x}$ on U. Now partition the unit square into $b = 4$ subsquares $\Delta_{(i,j)} = [i/2, (i+1)/2) \times [j/2, (j+1)/2), i, j = 0, 1$, and redistribute ϵ_0 over these subsquares by i.i.d. non-negative random factors $W_{ij}, i, j = 0, 1$, referred to as *cascade generators* (or *weights*) distributed as W, say, with $\mathbb{E}W_{ij} = \mathbb{E}W = 1$, i.e., *conservation on average*.[9] Then one obtains a new (random) measure on U given by

$$\mu_1(d\mathbf{x}) = \epsilon_0 \sum_{i,j\in\{0,1\}} W_{ij} \mathbf{1}_{\Delta_{ij}}(\mathbf{x}) d\mathbf{x}. \tag{21.1}$$

This process may be iterated by similarly partitioning each Δ_{ij} into four subregions and redistributing the respective amounts $\epsilon_0 W_{ij}$ over Δ_{ij} by i.i.d. random factors distributed as W, to obtain a sequence of random measures $\mu_0, \mu_1, \mu_2, \dots$. The *binomial cascade* with parameter $p \in (0, 1)$ is defined by taking

$$W = \begin{cases} \frac{1}{p}, & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

At the n-th scale there are 4^n pixels Δ of area $1/4^n$ each and having (random) measure 0 if a factor of 0 occurs in the cascade path defining Δ or $p^{-n} \times 4^{-n}$ if there are no 0's. Thus the total measures $\mu_n(U)$ take the form

$$Z_n = \mu_n(U) = (\frac{1}{p})^n \epsilon_0 (\frac{1}{4})^n X_n = \epsilon_0 \frac{X_n}{(4p)^n}, \quad n = 0, 1, 2, \dots, \tag{21.2}$$

where X_n counts the number of *nonzero* offspring at the n-th generation; i.e., at each generation there are $b = 4$ trials, each one of which has probability p of a nonzero cascade factor, independently of the other three. Thus $\{X_n : n = 0, 1, 2, \dots\}$ is a Bienaymé–Galton–Watson simple branching process starting from a single

[8]Some early laboratory investigations into the scope, validity, and experimental/statistical challenges can be found in Chavarria et al. (1995), Politano and Pouquet (1995), Benzi et al. (1996), Molchan (1997), Ossiander and Waymire (2002), Budaev (2008), as well as in the more recent Zhao et al. (2021).

[9]In the context of statistical turbulence models, the lack of sample pathwise energy conservation is sometimes argued from the perspective that hot wire measurements involve velocities in one-dimensional projections of a three-dimensional velocity field.

progenitor and having binomial offspring distribution with parameters $b = 4$, p, i.e., $f(k) = \binom{4}{k} p^k (1-p)^{4-k}$, $k = 0, 1, 2, 3, 4$. In general the product along any path $\prod_{j=1}^{n} W_{\gamma|j} \to 0$ a.s. as $n \to \infty$ for i.i.d. mean one non-negative random variables distributed as W (Exercise 1). However, there are uncountably many paths in the limit. Thus one may expect that if the "amount of branching"as determined by the partitioning parameter $b = 4$ is "suitably large"relative to the multiplicative factors W, then there will be a non-trivial limiting cascade measure with positive probability. A straightforward application of the super-criticality criteria for survival of a branching process (Theorem 9.1), together with the Kesten–Stigum theorem (Theorem 14.2), yields that the fine scale limit $\lim_{n\to\infty} Z_n$ defined by (21.2) is positive with positive probability if and only if $4p > 1$. The general theorem[10] is as follows. For a natural number $b \geq 2$, referred to as the *branching parameter*, let

$$\mathbb{T} = \cup_{n=0}^{\infty} \{0, 1, \ldots, b-1\}^n, \quad \partial\mathbb{T} = \{0, 1, \ldots, b-1\}^{\infty}, \tag{21.3}$$

where $\{0, 1, \ldots, b-1\}^0 = \{\emptyset\}$ is distinguished as a *root* vertex of height $|\emptyset| = 0$. Let $\lambda(dt) = \delta_{\frac{1}{b}}^{\infty}(dt)$ denote the product measure; i.e., λ may be viewed as the distribution of an i.i.d. sequence of uniformly distributed random variables on $\{0, 1, \ldots, b-1\}$ or, equivalently, as the Haar measure on $\partial\mathbb{T}$ viewed as a compact abelian group for coordinate-wise addition mod b and the product topology. For $v = (v_1, \ldots, v_k) \in \mathbb{T}$, write $|v| = k$ to denote the (genealogical) height of v, with $|\emptyset| = 0$. Let $W_\emptyset = 1$, and let $\{W_v : \emptyset \neq v \in \mathbb{T}\}$ be a family of mean one, i.i.d. positive random variables on a probability space $(\Omega, \mathcal{F}, P)$ for weighted b-ary trees. While the offspring distribution is deterministic, the weights are random variables. It is convenient to take the canonical product space model on $\Omega = [0, \infty)^{\mathbb{T}}$ for i.i.d. non-negative random variables indexed by the tree $\mathbb{T}$.

Define a sequence $\mu_n, n \geq 1$, of random measures on the Borel σ-field $\mathcal{B}$ of $\partial\mathbb{T}$ for the product topology, via the specification that $\mu_n << \lambda$ with Radon-Nikodym derivative on $\mathcal{F}_n \otimes \mathcal{B}$, where $\mathcal{F}_n = \sigma(W_v : |v| \leq n)$, given by

$$\frac{d\mu_n}{d\lambda}(t) = Q_n(t) = \prod_{j=1}^{n} W_{t|j}, \quad t \in \partial\mathbb{T}. \tag{21.4}$$

It is sometimes convenient to regard the cascade as a random measure on the unit square or unit interval partitioned using the parameter b; see Figure 21.1 for a depiction of equal subdivisions ($b = 2$) of the unit interval.

Theorem 21.1 *(Kahane–Peyriére).* The vague limit $\mu_\infty = \lim_n \mu_n$ exists almost surely. Moreover $\mathbb{E}\mu_\infty(\partial\mathbb{T}) > 0$ if and only if $\mathbb{E}_P W \ln W < \ln b$.

---

[10]This theorem has been refined from a number of perspectives, including general offspring distributions and statistical dependence between cascade generators, e.g., see Peyriére (1977); Burd and Waymire (2000); Waymire and Williams (1996).

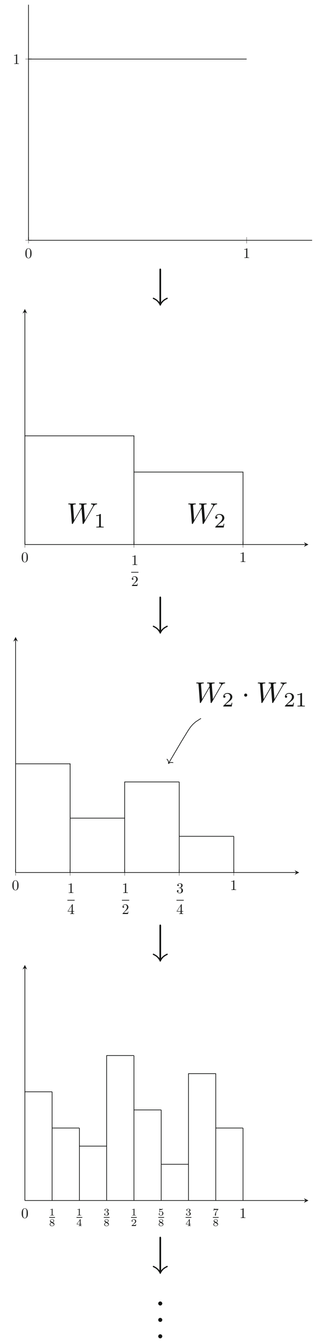


Fig. 21.1 Realization as a One-Dimensional Cascade Density on [0, 1)

Proof. The proof is in two parts. In the first part we establish the a.s. existence of the vague limit measure μ_∞. The second part is devoted to the non-triviality of the limit. For a.s. existence, let $f \in C(\partial\mathbb{T})$ and define

$$M_n(f) = \int_{\partial\mathbb{T}} f(t)\mu_n(dt), \quad n \geq 1.$$

One can see that $M_n(f)$ is a martingale by first noting that for each $s \in \partial\mathbb{T}$, the independence and mean one property yield

$$\mathbb{E}(\prod_{j=1}^{n+1} W_{s|j}|\mathcal{F}_n) = \prod_{j=1}^{n} W_{s|j}, \tag{21.5}$$

i.e., the product along a fixed path is a martingale. Thus, integrating (21.5) over the paths s with respect to $\lambda(ds)$, it follows that

$$\begin{aligned}\mathbb{E}(M_{n+1}(f)|\mathcal{F}_n) &= \int_{\partial\mathbb{T}} f(s)\mathbb{E}(\prod_{j=1}^{n+1} W_{s|j}|\mathcal{F}_n)\lambda(ds)\\ &= \int_{\partial\mathbb{T}} f(s)\prod_{j=1}^{n} W_{s|j}\lambda(ds)\\ &= M_n(f). \end{aligned}\tag{21.6}$$

So $M_n(f), n \geq 1$, is a martingale. Note that this includes the total mass $\mu_n(\partial\mathbb{T}) = M_n(1), n \geq 1$. In addition $\mathbb{E}|M_n(f)| \leq ||f||_\infty$. Thus, $M_n(f), n \geq 1$, is a bounded martingale. By the martingale convergence theorem, $M_\infty(f) = \lim_{n\to\infty} M_n(f)$ exists a.s. and in L^1 for each $f \in C(\partial\mathbb{T})$. Restricting to a countable dense set $\mathcal{D}$ of $f \in C(\partial\mathbb{T})$, including $f \equiv 1$, one obtains almost surely a densely defined bounded, positive linear functional $f \to M_\infty(f)$; i.e., removing a set Ω_0 of probability zero, $M_n(f+g) = M_n(f)+M_n(g), n \geq 1, f, g \in \mathcal{D}$, on $\Omega\backslash\Omega_0$, and therefore in the limit as $n \to \infty$. Extend M_∞ on $\Omega \setminus \Omega_0$ to a bounded linear functional on $C(\partial\mathbb{T})$ and apply the Riesz representation theorem to obtain the a.s. defined random measure $\mu_\infty(dt)$. To check that $\mu_\infty(dt)$ is the vague limit measure[11] for $f \in C(\partial\mathbb{T})$, and $\epsilon > 0$, there is an $f_\epsilon \in \mathcal{D}$ such that $||f - f_\epsilon||_\infty < \epsilon$. Thus for $\omega \in \Omega \setminus \Omega_0$,

$$\begin{aligned}&\limsup_{n,m} |\int_{\partial\mathbb{T}} f(t)\mu_n(dt,\omega) - \int_{\partial\mathbb{T}} f(t)\mu_m(dt,\omega)|\\ &\leq \epsilon \limsup_{n,m}\{\mu_n(\omega,\partial\mathbb{T}) + \mu_m(\omega,\partial\mathbb{T})\}\end{aligned}$$

[11] A stronger version of this result for all bounded measurable functions f was obtained by Kahane (1989). Also see Waymire and Williams (1995).

$$+ \|f_\epsilon\|_\infty \limsup_{n,m} |\mu_n(\omega, \partial\mathbb{T}) - \mu_m(\omega, \partial\mathbb{T})|$$
$$= \epsilon B(\omega), \tag{21.7}$$

where $B(\omega) = \limsup_{n,m}(\mu_n(\omega, \partial\mathbb{T}) + \mu_m(\omega, \partial\mathbb{T})) < \infty$ since $1 \in \mathcal{D}$. Letting $\epsilon \downarrow 0$, one sees that the sequence is Cauchy for all $f \in C(\partial\mathbb{T})$ with probability one.

The second part of the proof is to obtain the non-degeneracy criteria for μ_∞. For this, the distinguished path method may now be applied as follows: Similarly to the proof of the Kesten–Stigum theorem, extend the cascade model (weighted b-ary tree) on (Ω, P) to a model on the space of weighted b-ary trees ω with distinguished path weights along uniformly selected paths $t = \gamma$, such that the weights at vertices $v \in \gamma$ along the path are i.i.d. with a (mean) size-biased probability distribution $q(dx) = xP(W \in dx)$ on $[0, \infty)$, while those of the path are i.i.d., independent of weights along the path, having the given mean one distribution $p(dx) = P(W \in dx)$ on $[0, \infty)$. Specifically define a probability $\mathcal{Q}$ on the space of trees with distinguished path weights, for the filtration $\mathcal{F}_n$, by

$$\int_{\Omega\times\partial\mathbb{T}} g(\omega, t)\mathcal{Q}(d\omega, dt) = \mathbb{E}_P \int_{\Omega\times\partial\mathbb{T}} g(\omega, t)\mu_n(dt) = \mathbb{E}_P \int_{\partial\mathbb{T}} g(\omega, t)Q_n(t)\lambda(dt), \tag{21.8}$$

for bounded, $\mathcal{F}_n \times \mathcal{B}$-measurable g on $\Omega \times \partial\mathbb{T}$. Then, reversing the order of integration via Fubini–Tonelli, one may write

$$\mathcal{Q}(d\omega \times dt) = P_t(d\omega)\lambda(dt), \tag{21.9}$$

where $P_t << P$ on $\mathcal{F}_n$, with

$$P_t(d\omega) = Q_n(t)P(d\omega). \tag{21.10}$$

One may easily check (Exercise 2) for arbitrary fixed $t \in \partial\mathbb{T}$, the weights $W_{t|j}$, $j = 1, 2, \ldots$ are i.i.d. with size-bias distribution defined by $\mathbb{E}_{P_t} g(W_{t|j}) = \mathbb{E}_P W_{t|j} g(W_{t|j})$ under P_t, for arbitrary bounded measurable function g, while off the t-path, they are i.i.d. with $\mathbb{E}_{P_t} g(W_v) = \mathbb{E}_P g(W_v)$, $v \neq t|j$, for any j. In particular $P_t(W_{t|j} = 0) = 0$ for any path t. Moreover, the marginal projection of $\mathcal{Q}$ onto the space Ω of weighted b-ary trees may be expressed by integrating over all uniformly selected paths as

$$P^*(d\omega) := \mathcal{Q} \otimes \pi_\Omega^{-1}(d\omega) = \int_{\partial\mathbb{T}} P_t(d\omega)\lambda(dt). \tag{21.11}$$

As in the proof of the Kesten–Stigum theorem, first departure bounds are used to determine conditions for $P^* << P$ vs $P^* \perp P$ for the Lebesgue decomposition given by (Fig. 21.2)

$$P^*(d\omega) = \mu_\infty(\partial\mathbb{T})\mathbf{1}[\mu_\infty(\partial\mathbb{T}) < \infty]P(d\omega) + \mathbf{1}[\mu_\infty(\partial\mathbb{T}) = \infty)]P^*(d\omega). \tag{21.12}$$

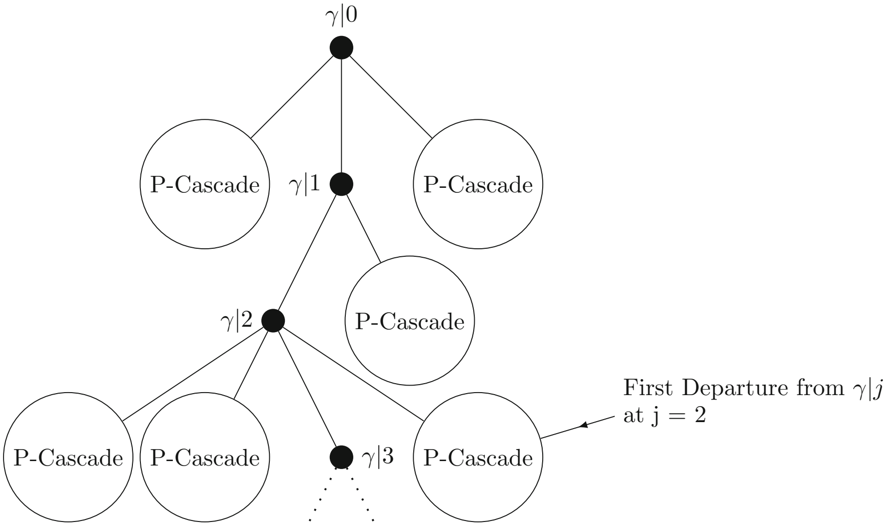


Fig. 21.2 First Departure Cascade Bounds

Specifically, one has *first departure bounds* with respect to an arbitrary fixed path $t = \gamma$, given by factoring out the common part and denoting concatenation of paths by $*$,

$$\begin{aligned}
&b^{-n}\prod_{j=1}^{n} W_{\gamma|j}\\
&\leq \mu_n(\partial\mathbb{T})\\
&\leq \prod_{j=1}^{n} W_{\gamma|j} b^{-n} + \sum_{j=1}^{n-1}\prod_{i=1}^{j} W_{\gamma|i} b^{-j} \sum_{|v|=n-j, v|1\neq\gamma|(j+1)} \prod_{i=1}^{n-j} W_{\gamma|j*v|i} b^{-(n-j)}, \qquad (21.13)
\end{aligned}$$

where

$$M_n = \sum_{j=1}^{n-1} b^{-(n-j)} \sum_{|v|=n-j, v|1\neq\gamma|(j+1)} \prod_{i=1}^{n-j} W_{\gamma|j*v|i} \qquad (21.14)$$

is a bounded non-negative submartingale under conditioning along the path $\sigma(W_{\gamma|j} : j = 0, 1, 2, \dots)$, with respect to the filtration

$$\mathcal{F}_n = \sigma\{W_v, |v| \leq n\}. \qquad (21.15)$$

Bounding the common part in the first departure inequality one also has

$$b^{-n}\prod_{j=1}^{n}W_{\gamma|j}\le\mu_n(\partial\mathbb{T})\le\prod_{j=1}^{n}W_{\gamma|j}b^{-n}+(\sup_{j\le n}b^{-j}\prod_{i=1}^{j}W_{\gamma|i})M_n. \tag{21.16}$$

For the *necessity* of $\mathbb{E}_P W\ln W < \ln b$ for non-triviality of μ_∞, suppose that $\mathbb{E}_P W\ln W \ge \ln b$. First consider the case $\mathbb{E}_P W\ln W > \ln b$ and $P_t(W = b) < 1$. For arbitrary fixed path weights along $t = \gamma$, it follows from the first departure lower bound and strong law of large numbers that P_γ-a.s., $\mu_n(\partial\mathbb{T}) \ge \exp\{n(\sum_{j=1}^{n}\frac{\ln W_{\gamma|j}}{n} - \ln b)\} \to \infty$. In the case $\mathbb{E}_P W\ln W = \ln b$ and $P_t(W = b) < 1$, one has P_t-a.s. $\limsup_{n\to\infty}\sum_{j=0}^{n}\{\ln W_{t|j} - \ln b\} = \infty$ by Chung–Fuchs[12] recurrence criteria for random walk along the t-path. Since the limit $\mu_\infty(\partial\mathbb{T})$ exists, this is enough to assert that in either case, $P_t(\mu_\infty(\partial\mathbb{T}) = \infty) = 1$, and by integrating out paths t, the triviality of μ_∞ follows from the Lebesgue decomposition, i.e., $P(\mu_\infty(\partial\mathbb{T}) = 0) = 1$. Finally the special case $P_t(W = b) = 1$, i.e., $P(W = b) = \frac{1}{b} = 1 - P(W = 0)$ follows by branching process extinction of a critical binomial offspring distribution with parameters b, $p = \frac{1}{b}$, as treated at the outset. To see that $\mathbb{E}_P W\ln W < \ln b$ is also *sufficient* for a non-trivial limit measure, fix a path $t = \gamma \in \partial\mathbb{T}$. It again follows from the strong law of large numbers that P_γ-a.s., $(b^{-j}\prod_{i=1}^{j}W_{\gamma|i})^{\frac{1}{j}} \to \frac{1}{b}e^{\mathbb{E}_P W\ln W} < 1$. Using the first departure upper bound one has that P_γ-a.s., $\mu_\infty(\partial\mathbb{T}) < \infty$. Integrating out $t = \gamma$, it follows that $\mu_\infty(\partial\mathbb{T}) < \infty$ P^*-a.s. Thus, by Lebesgue decomposition $\mathbb{E}_P\mu_\infty(\partial\mathbb{T}) = 1$. ■

Note that in the above example with $b = 4$, and Bernoulli $W = \frac{1}{p}$ with probability p, one has $\mathbb{E}_P W\ln W = p\frac{1}{p}\ln\frac{1}{p} < \ln 4$, if and only if $4p > 1$, as previously shown using the usual non-extinction condition for branching processes (also see Exercise 3).

Remark 21.2. Physics aside, let us close with a brief consideration of the $\log - \log$ linearity of moments versus length scale $\ell = b^{-n} = |\Delta_n|$, or $n = \frac{\log\ell}{\log b^{-1}}$, under conditions for which a nonzero limiting cascade measure and indicated moments exist.[13] Namely,

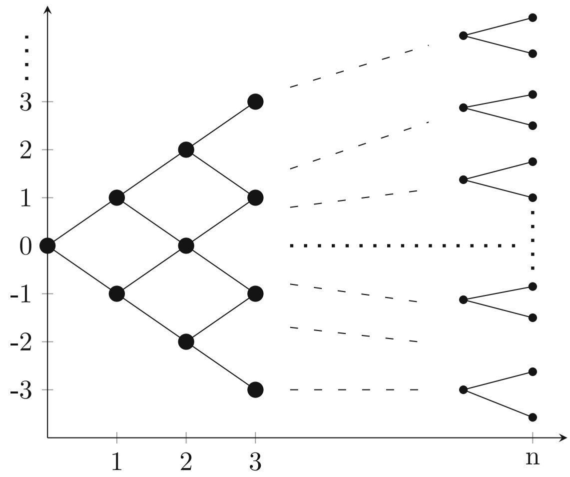
$$\begin{aligned}\log\mu_\infty^p(\Delta_n) &= (\log\mathbb{E}W^p - p\log b)n + \log\mathbb{E}Z_\infty^p\\ &= c(p\log b - \log\mathbb{E}W^p)\log\ell + \log\mathbb{E}Z_\infty^p,\end{aligned} \tag{21.17}$$

where $c = 1/\log b > 0$; i.e., one has $\log - \log$ linearity with slope given by $m(p) = c(p\log b - \log\mathbb{E}W^p)$.

[12] See BCPT p.123, 134.

[13] See Guivarc'h (1990), Barral and Jin (2014) in this regard.

Fig. 21.3 2^n Polygonal Paths



***Example 2** (Tree Polymer).* This class of models builds on the notion of multiplicative cascades with generators typically expressed as $W = 2e^{-Y}$, $b = 2$, and with $\mathbb{E}e^{-Y} = 1/2$; i.e., $\{Y_v : \emptyset \neq v \in \mathbb{T}\}$ is a binary tree-indexed family of i.i.d. real-valued random variables with $\mathbb{E}e^{-Y} = 1/2$, $Y_\emptyset = 0$. In this case the normalization of μ_n to a probability measure on $\partial\mathbb{T} = \{1, 2\}^\infty$ is given by the so-called *partition function*

$$Z_n = \sum_{|t|=n} \prod_{j=1}^{n} e^{-Y_{t|j}} = \sum_{|t|=n} e^{-H_n(t)}, \tag{21.18}$$

where $H_n(t) = \sum_{j=0}^{n} Y_{t|j}$, $t \in \partial\mathbb{T}$ denotes a tree path (Fig. 21.3).

Tree polymers may be viewed as random probability distributions of the polygonal paths $j \to t|j$, $j = 0, 1, \ldots$, for $t \in \partial\mathbb{T}$ selected according to the normalized probability formally given by

$$p_\infty(dt) = \lim_{n\to\infty} Z_n^{-1} \mu_n(dt). \tag{21.19}$$

The nature of this limit presents an interesting problem in itself. In the case that $Z_\infty = \lim_{n\to\infty} Z_n > 0$, referred to as *weak disorder*, $p_\infty(dt)$ is merely the cascade measure $\mu_\infty(dt)$ normalized to a probability by Z_∞. The disorder is said to be *strong disorder* if $Z_\infty = 0$ a.s. Complete determination of the limit (21.19) in the case of strong disorder is outside the scope of this text.[14] However, it hinges

[14]The existence of $p_\infty(dt)$ as a weak limit in probability in the critical strong disorder case $\mathbb{E}_P W \ln W = 2$ was proven by Johnson and Waymire (2011), and in Barral et al. (2014), while the non-critical strong disorder case $\mathbb{E}_P W \ln W > \ln 2$ is covered in Barral et al. (2012); and by Dey and Waymire (2015).

on the introduction of another important martingale that arises in a wide variety of branching contexts, namely the *derivative martingale*[15] D_∞ defined as the a.s. (positive) limit of the following martingale[16] In the case of *critical* strong disorder, i.e., $\mathbb{E}_P W \ln W = \ln 2$, also referred to as the *boundary case*, the derivative martingale D_∞ is determined by an a.s. limit of the sequence

$$D_n = \sum_{|t|=n} H_n(t)e^{-H_n(t)}, \quad n = 1, 2, \dots. \tag{21.20}$$

$D_n, n \geq 1$ is easily checked to be a (signed) martingale (Proposition 21.2 below). Under the additional moment assumption, $\mathbb{E}Y^2e^{-Y} < \infty$, the a.s. limit D_∞ can be shown to exist,[18] however, this requires a deeper analysis involving renewal theory and ladder random variables (presented in Chapter 25) than is possible from a simple application of the martingale convergence theorem (see Exercise 11). On the other hand, assuming it exists, one can show $D_\infty \geq 0$ a.s. (Proposition 21.3 below). In fact, the event $[D_\infty > 0]$ is an inherited event, and hence has probability zero or one (recall Proposition 12.6) of Chapter 12).

Remark 21.3. For the origins of the weak and strong disorder nomenclature, note that in the case $Y = 0$ a.s., i.e., no disorder, the polygonal paths are simple symmetric random walk paths. In the context of polymers, one views the randomness of their distributions as the result of impurities governing displacements. If the disorder is sufficiently weak (relative to the branching rate), then the usual limit theorems, e.g., law of large numbers, central limit theorem, are expected to apply as they would in the absence of disorder. However, if the impurities are sufficiently strong, then new limit theorems can be expected. With this jargon the condition $\mathbb{E}_P W \ln W < \ln b$ for existence of a non-trivial limit given by the Kahane–Peyriére theorem is the requirement that the branching rate be sufficiently large relative to the disorder.

Remark 21.4. From a perspective of fixed points to a special class of random iterations, it is noteworthy that the distributions of both Z_∞ and D_∞ provide fixed point solutions of the so-called *smoothing transformation*[19] (see Exercise 7).

In preparation for the next theorem note that the *critical strong disorder* conditions $\mathbb{E}W = 1$ and $\mathbb{E}_P W \ln W = \ln 2$ for the multiplicative cascade model correspond to $\mathbb{E}e^{-Y} = 1/2$ and $\mathbb{E}Ye^{-Y} = 0$, respectively, in the context of the tree polymer model.

[15] Biggins and Kyprianou (2004).

[16] See Kyprianou (1998), for historic background on the derivative martingale in the broader context of branching Brownian motion and branching random walk. In the particular case of branching Brownian motion and the Fisher-KPP equation[17] (see Exercise 10, Chap 28) Lalley and Selker (1987), prove it's a.s. existence without referring to it by name.

[17] This equation gets its name from Fisher (1937) and, independently, Kolmogorov et al. (1937).

[18] See Chen (2015) for the complete result.

[19] Holley and Liggett (1981); Durrett and Liggett (1983).

Theorem 21.2. Assume $\mathbb{E}e^{-Y} = 1/2$ and $\mathbb{E}Ye^{-Y} = 0$. Then $Z_n, n \geq 1$ is a positive martingale, and $D_n, n \geq 1$, is a martingale, with respect to $\mathcal{F}_n = \sigma\{Y_v : |v| \leq n\}$, $n \geq 1$.

Proof. Since one may express the partition function as a total cascade mass for $W = 2e^{-Y}$, $b = 2$, namely $Z_n = \mu_n(\partial\mathbb{T}) \equiv M_n(1)$, the martingale property follows immediately from the first part of the proof of Theorem 21.1. One may argue similarly for the martingale property of D_n. Namely, it is easy to check from the independence and the critical strong disorder (or boundary) conditions $\mathbb{E}e^{-Y} = 1/2$, $\mathbb{E}Ye^{-Y} = 0$, that for each fixed path $t \in \partial\mathbb{T}$,

$$\begin{aligned}
&\mathbb{E}(2^{n+1}\sum_{i=1}^{n+1} Y_{t|i} \prod_{j=1}^{n+1} e^{-Y_{t|j}} |\mathcal{F}_n) \\
&= 2^{n+1}\sum_{i=1}^{n} Y_{t|i} \prod_{j=1}^{n} e^{-Y_{t|j}} \frac{1}{2} + 2^{n+1}\prod_{j=1}^{n} e^{-Y_{t|j}} \mathbb{E}(Y_{t|n+1}e^{-Y_{t|n+1}}|\mathcal{F}_n) \\
&= 2^{n}\sum_{i=1}^{n} Y_{t|i} \prod_{j=1}^{n} e^{-Y_{t|j}}. \qquad (21.21)
\end{aligned}$$

That is, $2^n \sum_{i=1}^n Y_{t|i} \prod_{j=1}^n e^{-Y_{t|j}}$ is a martingale. Thus, as was done for Z_n, integrating (21.21) over the paths with respect to $\lambda(dt)$, and letting

$$\Delta_n(t_1,\dots,t_n)=\{s \in \partial\mathbb{T} : s_j=t_j, 1 \leq j \leq n\}, \ (t_1,\dots,t_n) \in \{0,1,\dots,b-1\}^n,$$

yields the martingale property of $D_n, n \geq 1$, since

$$\begin{aligned}
\int_{\partial\mathbb{T}} \sum_{i=1}^{n} Y_{s|i} \prod_{j=1}^{n} e^{-Y_{s|j}} \lambda(ds) &= 2^n \int_{\cup_{|t|=n}\Delta_n(t)} \sum_{i=1}^{n} Y_{s|i} \prod_{j=1}^{n} e^{-Y_{s|j}} \lambda(ds) \\
&= 2^n \sum_{|t|=n}\sum_{i=0}^{n} Y_{t|i} \prod_{j=1}^{n} e^{-Y_{t|j}} 2^{-n} = D_n. \quad (21.22)
\end{aligned}$$

■

Proposition 21.3. Assume $\mathbb{E}e^{-Y} = \frac{1}{2}$, and $\mathbb{E}Ye^{-Y} = 0$. Then one has that $\lim_{n\to\infty} \inf_{|t|=n} \sum_{j=1}^n Y_{t|j} = \infty$ a.s. In particular $D_n \geq 0$ a.s. for all n sufficiently large.

Proof. Let $Z_n = \sum_{|t|=n} e^{-\sum_{j=1}^n Y_{t|j}}$. Since $Z_n, n \geq 0$, is a non-negative martingale one has $0 \leq Z_\infty = \lim_{n\to\infty} Z_n$ exists a.s.. Moreover, $\mathbb{E}Z_\infty \leq 1$ by Fatou's lemma. Let

$$L = \limsup_{n\to\infty} e^{-\inf_{|t|=n}\sum_{j=1}^n Y_{t|j}}$$

$$\leq \limsup_{n\to\infty} \sum_{|t|=n} e^{-\sum_{j=1}^{n} Y_{t|j}}$$
$$= \limsup_{n\to\infty} Z_n = Z_\infty.$$

In particular $\mathbb{E}L \leq 1$. Now observe that L satisfies the stochastic recursion (see Exercise 12)

$$L =^{\text{dist}} L_1 e^{-Y_1} \vee L_2 e^{-Y_2}, \tag{21.23}$$

where L_1, L_2 are independent, identically distributed as L and independent of Y_1, Y_2. So, from (21.23), one has $\mathbb{E}L \leq \mathbb{E}(L_1 e^{-Y_1} + L_2 e^{-Y_2}) = \mathbb{E}L$, with strict inequality unless $L = 0$ a.s. Thus, $L = 0$ a.s. This proves the assertion that $\lim_{n\to\infty} \inf_{|t|=n} \sum_{j=1}^{n} Y_{t|j} = \infty$ a.s. From here one sees that $D_n \geq (\inf_{|s|=n} \sum_{j=1}^{n} Y_{s|j}) \sum_{|t|=n} \prod_{i=1}^{n} e^{-Y_{t|i}} > 0$ a.s. for all n sufficiently large. ■

Corollary 21.4. $[\liminf_{n\to\infty} D_n > 0]$ has probability zero or one.

Proof. In view of Proposition 12.6 of Chapter 12, it suffices to show that $[\liminf_{n\to\infty} D_n = 0]$ is an inherited event. That is, it occurs for all finite subtrees, and if this event occurs for the infinite tree rooted at $< 1 >$, then it occurs for any infinite subtree rooted at an offspring of $< 1 >$. To see inheritance, fix arbitrary $v \in \cup_{\ell=0}^{\infty}\{1,2\}^{\ell}$, say $|v| = k$, and for $n > k$, let

$$D_{n,v} = \sum_{|t|=n, t\geq v} \sum_{j=k+1}^{n} Y_{t|j} e^{-\sum_{i=k+1}^{n} Y_{t|i}},$$

where $t \geq v$ means that $t|k = v$. Then, using the previous theorem, one has a.s. $\liminf_{n\to\infty} D_{n,v} \geq 0$. Now, for $|v| = 1$,

$$\begin{aligned} D_n &\geq \sum_{|t|=n, t\geq v} H_n(t) e^{-H_n(t)} \\ &= e^{-H_1(v)} D_{n,v} + H_1(v) e^{-H_1(v)} \sum_{|t|=n, t\geq v} e^{-\sum_{j=2}^{n} Y_{t|j}}. \end{aligned} \tag{21.24}$$

Since this is a case of critical strong disorder, the liminf for the second term on the right is a.s. zero. Thus

$$\liminf_{n\to\infty} D_n \geq e^{-H_1(v)} \liminf_{n\to\infty} D_{n,v} \geq 0. \tag{21.25}$$

In particular, if $\liminf_{n\to\infty} D_n = 0$, then $\liminf_{n\to\infty} D_{n,v} = 0$; i.e., the event is inherited. ■

As already noted, the derivative martingale has proven to be a powerful tool for the analysis of polymers and branching random walks in the case of strong disorder. For example, it can be shown that in the critical strong disorder case there is a sequence of positive constants[20] $a_n = n^{-\frac{1}{2}}, n \geq 1$, and $Z_n/a_n D_n \to c > 0$ in probability (but not a.s.) as $n \to \infty$ Using this one may re-scale to show[21] that $Z_n^{-1}\mu_n(dt)$ converges weakly in probability to $p_\infty(dt)$, where

$$p_\infty(\Delta_m(v)) = \frac{D_\infty(v)\prod_{j=1}^m Y_{v|j}}{\sum_{|u|=m} D_\infty(u)\prod_{j=1}^m Y_{u|j}}, \tag{21.26}$$

where $\Delta_m(v) = \{t \in \{1,2\}^\infty : t|m = v|m\}$, $v \in \cup_{m=0}^\infty \{1,2\}^m$, and $D_\infty(u)$ denotes the derivative martingale for the subtree rooted at u.

Example 3 *(Branching Random Walk)***.** Let us consider a branching random walk on the real-number line with initial ancestor at $Y_0 = 0$, having two children[22] in the first generation displaced by i.i.d. amounts X_1, X_2. Each generation repeats this birth-displacement process. The second generation consists of four random walkers at positions $Y_{i,j} = X_i + X_{i,j}, i, j = 1, 2$, for i.i.d. displacements $X_1, X_2, X_{1,1}, X_{1,2}, X_{2,1}, X_{2,2}$. The construction proceeds recursively. Let $\{Y_v : |v| = n\}$ be the locations of the 2^n walkers at the nth generation. $Y_0 = 0, Y_1 = X_1, Y_2 = X_2$,

$$Y_v = \sum_{j=0}^n X_{v|j}, \quad v \in \{1,2\}^n, n = 1, 2, \ldots,$$

where $v|0 = 0, v|j = (v_1, \ldots, v_j), j \leq n$, for $v = (v_1, v_2, \ldots, v_n) \in \{1,2\}^n$. Also the genealogical length of v is denoted by $|v| = n$. To single out a fixed but arbitrary path let $\mathbf{1}_j = (1, 1, \ldots, 1)$ be a leftmost path of length j, and $S_n = \sum_{j=0}^n X_{\mathbf{1}_j}, S_0 = 0$.

The following general lemma[23] summarizes a size-biasing tool often employed for such branching random walk computations.

Lemma 1 *(Many-To-One-Formula)***.** Fix arbitrary $\lambda > 0$ such that $\psi(\lambda) = \ln \mathbb{E}\sum_{|v|=1} e^{-\lambda Y_v} < \infty$. Then for any non-negative Borel measurable function g,

[20] See Aidekon and Shi (2014).

[21] See Johnson and Waymire (2011).

[22] The example treated here illustrates the theory initiated by Hammersley, Kingman, and Biggins, but is less comprehensive. See Biggins (2010), for a historical and more comprehensive review of these developments, as well as more recent results.

[23] This simple formula and its generalizations arise in a variety of "branching"contexts, but its origins appear to be generally unattributed.

$$\mathbb{E}\sum_{|v|=n} g(Y_v) = \mathbb{E}\{e^{\lambda\hat{S}_n+n\psi(\lambda)}g(\hat{S}_n)\},$$

where $\hat{S}_0 = 0$, $\{\hat{S}_n : n = 0, 1, \ldots\}$ is a random walk with the size-biased distribution of i.i.d. displacements defined by the specifications that:

$$\mathbb{E}f(\hat{S}_1) = \mathbb{E}\sum_{|v|=1} \frac{e^{-\lambda Y_v}}{\mathbb{E}\sum_{|u|=1} e^{-\lambda Y_u}} f(Y_v), \tag{21.27}$$

for all non-negative Borel measurable functions f.

Proof. The proof is by induction. For $n = 1$, this is the definition of the size-bias distribution defining $\hat{S}_1$. Specifically,

$$\begin{aligned}
&\mathbb{E}\sum_{|v|=1} g(Y_v)\\
&= \mathbb{E}\big(g(X_1) + g(X_2)\big)\\
&= \mathbb{E}\Big(\frac{e^{-\lambda X_1}}{\mathbb{E}(e^{-\lambda X_1} + e^{-\lambda X_2})} e^{\lambda X_1+\psi(\lambda)} g(X_1) + \frac{e^{-\lambda X_2}}{\mathbb{E}(e^{-\lambda X_1} + e^{-\lambda X_2})} e^{\lambda X_2+\psi(\lambda)} g(X_2)\Big)\\
&= \mathbb{E}e^{\lambda\hat{S}_1+\psi(\lambda)} g(\hat{S}_1).
\end{aligned}$$

Assume the formula holds for n and condition on the first generation of branching to get, using the substitution formula for conditional expectation,

$$\begin{aligned}
\mathbb{E}\sum_{|v|=n+1} g(Y_v) &= \mathbb{E}\sum_{|u|=n} \{g(X_1 + Y_{1u}) + g(X_2 + Y_{2u})\}\\
&= \mathbb{E}e^{\lambda\hat{S}_n+n\psi(\lambda)}[g(X_1 + \hat{S}_n) + g(X_2 + \hat{S}_n)]\\
&= \mathbb{E}e^{\lambda\hat{S}_{n+1}+(n+1)\psi(\lambda)} g(\hat{X}_1 + \hat{S}_n)\\
&= \mathbb{E}e^{\lambda\hat{S}_{n+1}+(n+1)\psi(\lambda)} g(\hat{S}_{n+1}),
\end{aligned}$$

where $\hat{X}_1$ is independent of $\hat{S}_n$ with the distribution defined by (21.27). ■

Remark 21.5. One may notice that in the present framework $\psi(\lambda) = \ln\mathbb{E}\sum_{|v|=1} e^{-\lambda Y_v} = \ln 2\mathbb{E}e^{-\lambda X_1} < \infty$. However, the pair X_1 and X_2 need not be i.i.d. for the many-to-one lemma. The branching random walk framework may be generalized accordingly to the case of i.i.d. displacement vectors distributed as

(X_1, X_2). Also the binary branching random walk is readily generalizable to any supercritical branching process in place of the binary tree.[24]

Notice that using the right-side of the many-to-one formula with $g(x) = e^{-\lambda x}$, one has after a tiny bit of preparatory algebra

$$\mathbb{E}\hat{S}_1 = \mathbb{E}\Big(\frac{X_1 e^{-\lambda X_1}}{\mathbb{E}(e^{-\lambda X_1} + e^{-\lambda X_2})} + \frac{X_2 e^{-\lambda X_2}}{\mathbb{E}(e^{-\lambda X_1} + e^{-\lambda X_2})}\Big) = -\psi'(\lambda). \tag{21.28}$$

As a warm-up, using Jensen's inequality and bounding the maximum by the sum, followed by the many-to-one formula, one has

$$\begin{aligned}\mathbb{E}\big(-\min_{|v|=n} \lambda Y_v\big) &= \mathbb{E} \ln e^{\max(-\lambda Y_v)} \\ &\le \ln \mathbb{E} e^{\max_{|v|=n}(-\lambda Y_v)} \\ &\le \ln \mathbb{E} \sum_{|v|=n} e^{-\lambda Y_v} \\ &= \ln e^{n\psi(\lambda)} = n\psi(\lambda). \end{aligned} \tag{21.29}$$

Thus, for any $\lambda > 0$,

$$\frac{1}{n}\mathbb{E} \min_{|v|=n} Y_v \ge -\frac{\psi(\lambda)}{\lambda},$$

and therefore, for every n,

$$\frac{1}{n}\mathbb{E} \min_{|v|=n} Y_v \ge \sup_{\lambda>0} \frac{-\psi(\lambda)}{\lambda} = -\inf_{\lambda>0} \frac{\psi(\lambda)}{\lambda}. \tag{21.30}$$

In fact, a more judicious application of the many-to-one formula provides a tool for the speed of the leftmost (or right-most) particle in the branching random walk defined as follows.

Definition 21.1. The speeds (rates) of the left and right-most particles of the branching random walk are given by the almost sure limits

$$r = \lim_{n\to\infty} \min_{|v|=n} \frac{Y_v}{|v|}, \qquad R = \lim_{n\to\infty} \max_{|v|=n} \frac{Y_v}{|v|}, \tag{21.31}$$

respectively, provided the limits exist.

[24]The monograph by Shi (2012) features the utility of distinguished path analysis for branching random walk in this more general framework.

The point of the following result is the *computation* of the extremal particle speeds using the size-biasing reflected in the many-to-one formula under the assumption that the speeds exist as almost sure and L^1-limits. This is a bit weaker statement than also showing the existence[25] of the particle speed limits.

Theorem 21.5 *(Biggins–Kingman–Hammersley).* Let

$$\psi(\lambda) = \ln \mathbb{E}(e^{-\lambda X_1} + e^{-\lambda X_2}) = \ln(2\mathbb{E}e^{-\lambda X_1}).$$

Assume $\psi(\lambda) < \infty$ for some $\lambda > 0$. Also assume that the limit defining the speed r of the leftmost particle exists almost surely and in L^1. Then

$$\lim_{n\to\infty} \min_{|v|=n} \frac{Y_v}{|v|} = \gamma := -\inf_{\lambda>0} \frac{\psi(\lambda)}{\lambda}$$

.

Proof. That γ is a lower bound on the speed r follows from the calculation (21.30) under the hypothesis of the theorem. For the reverse inequality it suffices to check that there is an $N = N(\epsilon)$ such that

$$\lim_{k\to\infty} \frac{1}{kN} \min_{|v|=kN} Y_v \le \gamma + \epsilon$$

for all sufficiently small $\epsilon > 0$. Construct a branching random walk $\tilde{T}$ such that the first generation of $\tilde{T}$ is all v in the Nth generation of T such that $Y_v \le (\gamma + \epsilon)N$. More generally, for $n \ge 1$, if v is in the nth generation of $\tilde{T}$, then its offspring in the $(n+1)$st generation of $\tilde{T}$ consists of u in the $(n+1)N$th generation of T that are descendants of v in T, and

$$Y_v - Y_u \le (\gamma + \epsilon)N.$$

With t to be determined, using the many-to-one formula,

$$\begin{aligned} \mu_{\tilde{T}} &= \mathbb{E} \sum_{|v|=N} \mathbf{1}[Y_v \le (\gamma+\epsilon)N] \\ &= \mathbb{E}e^{\lambda \hat{S}_N + N\psi(\lambda)} \mathbf{1}[\hat{S}_N \le (\gamma+\epsilon)N]. \end{aligned} \tag{21.32}$$

Now choose $\lambda > 0$ such that

$$\frac{\psi(\lambda)}{\lambda} > \psi'(\lambda) > \inf_{s>0} \frac{\psi(s)}{s} - \epsilon = -(\gamma + \epsilon).$$

[25] See Shi (2012)

Namely, take $\lambda < \inf\{s > 0 : \frac{\psi(s)}{s} = \inf_{r>0} \frac{\psi(r)}{r}\}$. Then

$$\mathbb{E}\hat{S}_1 = -\psi'(\lambda) < \gamma + \epsilon.$$

Now choose $a \in (\psi'(\lambda), \frac{\psi(\lambda)}{\lambda})$ to get

$$\begin{aligned} \mu_{\tilde{T}} &= \mathbb{E}e^{\lambda \hat{S}_N + N\psi(\lambda)} \mathbf{1}[\hat{S}_N \leq (\gamma + \epsilon)N] \\ &\geq e^{(-a\lambda + \psi(\lambda))N} P(-aN \leq \hat{S}_N \leq (\gamma + \epsilon)N). \end{aligned} \tag{21.33}$$

Then,

$$\mathbb{E}\hat{S}_1 = -\psi'(\lambda) \in (-a, \gamma + \epsilon).$$

So, in the limit $N \to \infty$,

$$P(-aN \leq \hat{S}_N \leq (\gamma + \epsilon)N) \to 1, \tag{21.34}$$

and

$$e^{(-a + \frac{\psi(\lambda)}{\lambda})N} \to \infty.$$

In particular there is an N sufficiently large that $\mu_{\tilde{T}} > 1$. By super-criticality of the branching process, it will survive to every generation. Thus, in view of the bound (21.34), under the hypothesis of the theorem one has

$$\lim_{k \to \infty} \frac{1}{kN} \min_{|v|=kN} Y_v \leq \gamma + \epsilon.$$

■

Exercises

1. Show that apart from the trivial case of random variables with a.s. constant value one, the infinite product of i.i.d. mean one non-negative random variables is a.s. zero. [*Hint*: First consider the case when the value 0 has positive probability, otherwise take logarithms and apply the strong law of large numbers, followed by strict Jensen inequality.]
2. Verify that for arbitrary fixed $t \in \mathbb{T}$, the weights $W_{t|j}, j = 1, 2, \ldots$ are i.i.d. with size-bias distribution $\mathbb{E}_{P_t} g(W_{t|j}) = \mathbb{E}_P W_{t|j} g(W_{t|j})$ under P_t, for arbitrary bounded measurable function g, while off the t-path, they are i.i.d. with $\mathbb{E}_{P_t} g(W_v) = \mathbb{E}_P g(W_v)$, $v \neq t|j$, for any j.

3. Consider the Bienaymé–Galton–Watson branching process $\{X_n : n \geq 0\}$, $X_0 = 1$, with binomial offspring distribution with parameters b, p. (i) Show that the size-bias distribution is binomial with parameters $b-1, p$. (ii) Consider the multiplicative cascade with Bernoulli distributed weights $W = \frac{1}{p}$ with probability p, else $W = 0$. Show that $W = \frac{1}{p}$ a.s. under mean size-bias of the distribution of W. (iii) Show that the submartingale M_n coincides in the respective first departure bounds (21.13) and in Proposition 14.4 for the Kesten–Stigum theorem.
4. Assuming the limit exists almost surely and in L^1, determine the speed R of the right-most particle. [*Hint*: Consider the leftmost particle for the branching random walk with $-Y_v$ in place of Y_v.]
5. Determine the speeds of the left and right-most particles in a Gaussian branching random walk with mean zero and variance $\sigma^2 > 0$.
6. Consider the multiplicative cascade model in which the cascade generators are W are uniformly distributed on [0, 2]. Show that the (i) cascade survives almost surely and (ii) $Z_\infty = \mu_\infty(\partial\mathbb{T})$ has a Gamma distribution. (iii) Extend (ii) this to mean one Beta distributed generators on [0, 2].
7.
   (i) Show that $Z_\infty = \mu_\infty(\partial\mathbb{T})$ is a non-negative solution (fixed point) of $Z_\infty =^{dist} \frac{1}{b}\sum_{j=0}^{b-1} W_j Z_\infty^{(j)}$, where $Z_\infty^{(j)}, j = 0, 1, \ldots, b-1$, is i.i.d., independent of $W_0, \ldots, W_{b-1}$, and distributed as Z_∞.
   (ii) Assuming that the almost sure limit exists, show that the derivative martingale D_∞ is another solution in the context of random polymers.
8. (*Biggin's theorem*[26]) Consider the branching random walk, and define $\varphi(\lambda) = \mathbb{E}\sum_{|v|=1} e^{-\lambda Y_v}$. Also assume $\varphi(a) < \infty$ for some $a \in \mathbb{R}$. (i) Show that $Z_n(a) = \varphi^{-n}(a)\sum_{|v|=n} e^{-aY_v}, n = 1, 2, \ldots$ is a positive martingale with almost sure limit $Z_\infty(a)$. (ii) Assume further that $\varphi'(a) := \mathbb{E}\sum_{|v|=1} Y_v e^{-aY_v}$ exists and is finite. Use the distinguished path analysis for multiplicative cascades to show that $\mathbb{E}Z_\infty(a) = 1$ if and only if $\mathbb{E}\sum_{|v|=1} e^{-aY_v}\ln^+(\sum_{|u|=1} e^{-aY_u}) < \infty$ and $a\varphi'(a) < \varphi(a)\ln\varphi(a)$.
9. (*General Many-to-one Formula*) Show that the many-to-one formula extends as follows: For any non-negative Borel measurable function g, $\mathbb{E}\sum_{|v|=n} g(Y_1, Y_2, \ldots, Y_v) = \mathbb{E}\{e^{\lambda\hat{S}_n + n\psi(\lambda)} g(\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n)\}$, where $\hat{S}_0 = 0$, $\{\hat{S}_n : n = 0, 1, \ldots\}$ is a random walk with the size-biased distribution of i.i.d. displacements defined by the specifications (21.27).
10. Assuming critical strong disorder, show that $\mathbb{E}D_n = 0$, where $D_n, n \geq 1$, is the derivative martingale.
11. Consider the tree polymer in which Y is normally distributed with mean μ and variance σ^2. (i) Show critical strong disorder if and only if $\mu = \sigma^2$. (ii) Assum-

[26] While this result is proven for more generally defined branching random walks in Biggins (1977), the proof suggested here using distinguished path analysis is due to Lyons (1997).

ing critical strong disorder show that $\mathbb{E}\sum_{|t|=n} H_n(t)^+ e^{-H_n(t)} = c\sqrt{n}$. [*Hint*: Condition the j-th term of $H_n(t)^+ = \sum_{j=1}^n Y_{t|j}\mathbf{1}[H_n(t) > 0]\prod_{i=1}^n e^{-Y_{t|i}}$ on $Y_{t|j}$. Perform the indicated integrations, including an application of Fubini theorem in the final expected value.]

12. Define for $n \geq 1$, $L_n = e^{-\inf_{|t|=n}\sum_{j=1}^n Y_{t|j}}$, and show that $L_{n+1} = e^{-Y_1}L_n^{(1)} \vee e^{-Y_2}L_n^{(2)}$, where $L_n^(1), L_n^{(2)}$ are independent of Y_1, Y_2, as well as mutually independent and distributed as L_n. Show that (21.23) follows from this.
13. Let μ denote the mean displacement for the branching random walk. (a) Compute the large deviation rate for 2^n independent random walkers: That is, show that $I_1(x) = \ln 2 - I_0(x) = \lim_{n\to\infty}\frac{1}{n}\ln P(\max_{1\leq j\leq 2^n}\sum_{i=1}^n X_i^{(j)} > nx)$, for $x > R + \mu$ in the case that $X_i^{(j)}$ are i.i.d. distributed as X_v, where $I_0(x)$ is the large deviation rate for a single random walk.[*Hint*: The upper bound on $P(\max_{1\leq j\leq 2^n}\sum_{i=1}^n X_i^{(j)} > nx)$ is straightforward by sub-additivity of probability and the Cramér-Chernoff large deviation theorem for a single random walker, (see BCPT, p. 94). For the lower bound, use the inequality $1-(1-s)^t \geq 1-e^{-st} \geq st(1-st), s, t > 0$, and the Cramér-Chernoff large deviation theorem.] (b) Show[27] that the large deviation rate $I_2(x) \leq \ln 2 - I_0(x) = I_1(x)$, $x > R + \mu$, for the corresponding branching random walk. [*Hint*: Use sub-additivity of the probability to see the same $I_1(x)$ as an upper bound.]
14. (*Gantert-H ofelsauerit Stochastic Order Lemma*) (a) Suppose that $\{U_i\}_{i\geq 1}$ and $\{V_i\}_{i\geq 1}$ are two independent sequences of random variables, and the $V_i, i \geq 1$, are identically distributed. Show for any $k \geq 1, x \in \mathbb{R}$, $\max_{1\leq i\leq k}(U_i + V_i)$ is stochastically smaller than $\max_{1\leq i\leq k}(U_i + V_1)$; i.e., $P(\max_{1\leq i\leq k}(U_i + V_i) \leq x) \leq P(\max_{1\leq i\leq k}(U_i + V_1) \leq x)$ for all x. (b) Show that the right-most particle in the branching random walk at time n is stochastically smaller than the right-most particle among 2^n i.i.d. random walkers with the same displacement distribution.[*Hint*: Use induction as follows: Consider the right-most position of the branching random walk at $n+1$ generation, and use the induction hypothesis followed by the result (a).]

[27] In Gantert and Hofelsauer (2019) it is shown using the stochastic order in Exercise 14 that the two large deviation rates coincide for $x > R + \mu$, and more, where μ is the mean displacement.

Chapter 22
Special Topic: Bienaymé–Galton–Watson Simple Branching Process and Excursions

The tree contours and heights are identified as two natural discrete parameter stochastic processes associated with the branching process introduced in Chapter 14 as a probability distribution on a metric space of family trees. Analysis of contour paths in the special case of critical (shifted) geometric offspring distributions leads naturally to consideration of continuous parameter processes in terms of Brownian motion excursions.

For the development given here we continue to use the more detailed description of branching processes as probability distributions on the space of trees introduced in Chapter 14. It is sufficient to consider the case $X_0 = 1$ of a single progenitor $\langle 1\rangle$. Denote the family tree by $\tau = \tau(\langle 1\rangle)$. Suppose that τ is a finite tree rooted at $\langle 1\rangle$. The vertices $\langle 1v\rangle \in \tau$ are referred to as the *progeny* of $\langle 1\rangle$, and the *total progeny* n of $\langle 1\rangle$ is denoted $\|\tau\| = n$. In particular $\|\tau\| = n$ implies a total of $n+1$ vertices in τ since the root $\langle 1\rangle$ is excluded in this definition of total progeny. Loosely speaking, by traversing the contour of τ one obtains a polygonal path $(0, s_0 = 1)$, $(1, s_1), \ldots, (2n+1, s_{2n+1} = 0)$ starting at $s_0 = 1$ and reaching 0 for the first time in $2n+1$ time steps; see Figure 22.1. Recall that each offspring vertex of the tree may be uniquely associated with its parental edge, where we complete this picture by assigning a "ghost" edge to $\langle 1\rangle$. The definition of the associated contour path can be made more precise by an inductive definition as follows; also see Figure 22.1.

Definition 22.1. Let τ be a finite tree rooted at $\langle 1\rangle$ with total progeny (excludes the root) $\|\tau\| = n, n \geq 0$. If $\tau = \{\langle 1\rangle, \langle 11\rangle, \ldots, \langle 11...1\rangle\}$, then we say that τ is *pure trunk* and define its associated *contour path* $s = \{(j, s_j) : j = 0, 1, \ldots, 2n+1\}$ by

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_22

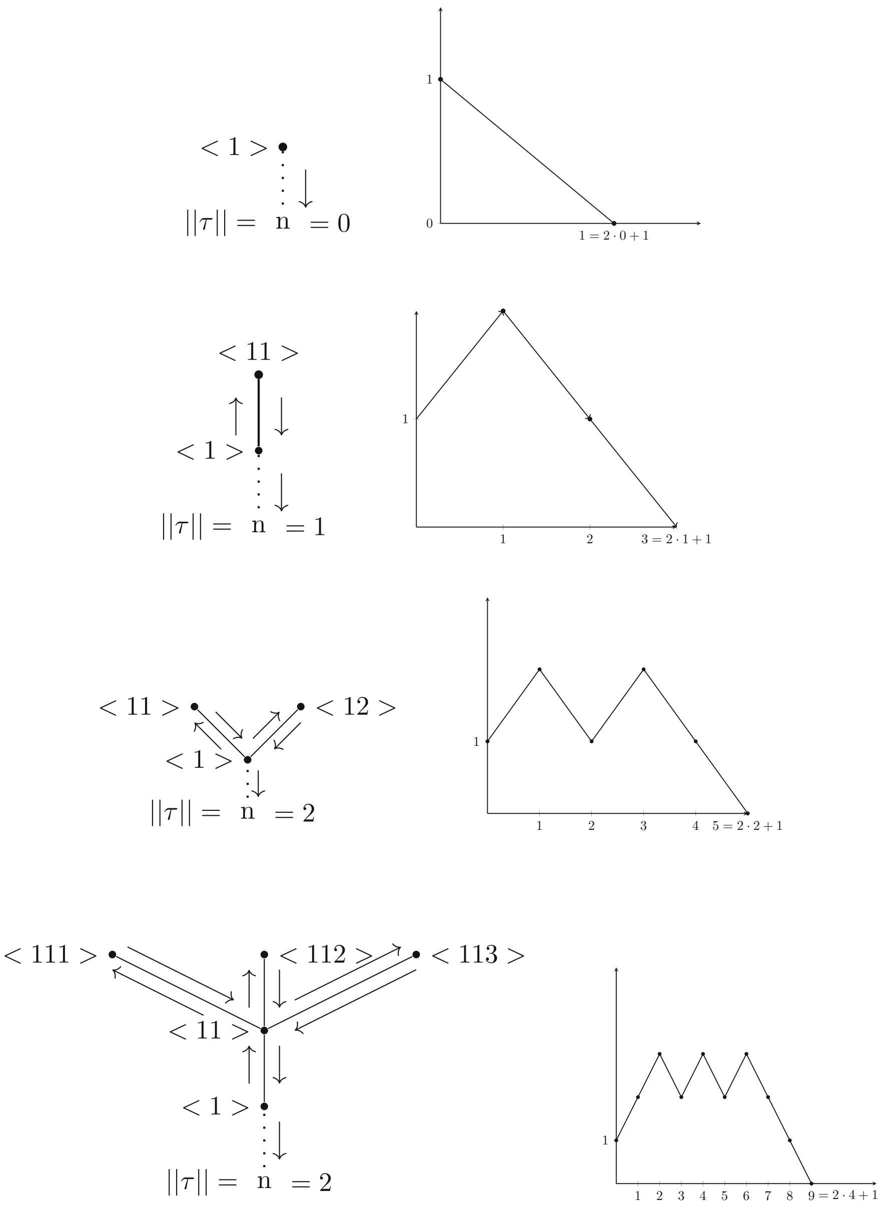


Fig. 22.1 Tree Contour

$$s_j := \begin{cases} j+1 & \text{if } j = 0, \dots, n \\ 2n+1-j, & \text{if } j = n+1, \dots, 2n+1. \end{cases} \tag{22.1}$$

If $\tau = \tau(\langle 1 \rangle)$ is not a pure trunk, then there is a unique k_0 such that the k_0-tuple $v^{(0)} = \langle 1, \dots, 1 \rangle \in \tau$, $\langle v^{(0)} 1 \rangle$, $\langle v^{(0)} 2 \rangle \in \tau$, and if $u = \langle u_1, \dots, u_j \rangle \in \tau$ with $j \leq k_0$, then u is the j-tuple vertex $\langle 1, \dots, 1 \rangle$. By induction on the total progeny there are unique contour paths $s^{(0)}, s^{(1)}, s^{(2)}$ associated with the pure trunk $\tau^{(0)}$ rooted at $\langle 1 \rangle$, and the subtrees $\tau^{(i)}, i = 1, 2$, rooted at $v^{(0)}$ such that $\langle v^{(0)} i \rangle \in \tau^{(i)}, i = 1, 2$. Define the *contour path* $s = \{(j, s_j) : j = 0, 1, \dots, 2n+1\}$ associated with τ by

$$s_j = \begin{cases} s_j^{(0)} & \text{for } j \leq k_0 - 1 \\ k_0 - 1 + s_{j-k_0}^{(1)} & \text{for } j = k_0, \dots, k_0 + j^{(1)} \\ k_0 - 1 + s_{j-k_0-j^{(1)}}^{(2)} & \text{for } j = k_0 + j^{(1)}, \dots, k_0 + j^{(1)} + j^{(2)} \\ s_{j-k_0-j^{(1)}-j^{(2)}}^{(0)} & \text{for } j = k_0 + j^{(1)} + j^{(2)}, \dots, 2n+1, \end{cases} \tag{22.2}$$

where $j^{(i)} = 2\|\tau^{(i)}\|, i = 1, 2$.

Proposition 22.1. For each $n \geq 0$ the set Ω_n consisting of all trees τ with single progenitor and total progeny $\|\tau\| = n$ is in one-to-one correspondence with the set $\mathcal{P}_n$ of all polygonal paths $s = \{(0, s_0), (1, s_1), \dots, (2n+1, s_{2n+1})\}$ such that $s_0 = 1, s_{2n+1} = 0, s_j > 0, |s_{j+1} - s_j| = 1, 1 \leq j \leq 2n$.

Proof. It is sufficient to show that the contour path map defined in Definition 22.1 is invertible. Again one may proceed inductively, noting that the result is obvious for $n = 0$ and $n = 1$. Let $n \geq 2$ and suppose one is given such a polygonal path $s = \{(j, s_j) : j = 0, 1, \dots, 2n+1\}, n \geq 0$, such that $s_0 = 1, s_{2n+1} = 0$. If s has a unique local maximum at $k_0 \in \{0, \dots, 2n\}$, which is necessarily the global maximum, then the associated tree τ is a pure trunk with $\|\tau\| = k_0$ progeny. Otherwise s must have at least two local maxima. Letting m_1 denote the location of the first such local maxima and m_2 that of the last local maxima in $\{0, 1, \dots, 2n+1\}$, the function s achieves at least one global minimum value between these points. Define

$$k_0 := \min\{s_j : m_1 \leq j \leq m_2\} - 1. \tag{22.3}$$

Then $k_0 \geq 0$ provides the total progeny in the pure trunk subtree $\tau^{(0)}$ rooted at $\langle 1 \rangle$. By induction, the polygonal path $\{(j - k_0, s_j - k_0) : j = k_0, \dots 2n - k_0\}$ defines a tree τ' which may be rooted at the k_0-tuple vertex $\langle 11 \dots 1 \rangle$ to obtain the associated tree having contour path s. ■

The probabilities of the contour paths associated with a Bienaymé–Galton–Watson distribution of random trees depend on the (conditional) probabilities with which trees τ occur having n total progeny. As will be shown in the next proposition, there is a unique class of offspring distributions, namely the *shifted geometric* *distribution*, for which these conditional probabilities are the same (uniform) over $\tau \in \Omega_n$, or equivalently $s \in \mathcal{P}_n$.

Proposition 22.2. Consider the Bienaymé–Galton–Watson probability P concentrated on the space Ω of trees with a single progenitor and defined by a positive offspring distribution $f(k) > 0, k = 0, 1, 2, \ldots$. Let $\Omega_n := \{\tau \in \Omega : \|\tau\| = n\}, n = 0, 1, 2 \ldots$. Then (i) $P(\Omega_n) > 0$ for each $n \geq 0$, and (ii)$P(\{\tau\}|\Omega_n)$ is constant (uniform) for $\tau \in \Omega_n$ for each $n \geq 0$, if and only if $f(k) = \theta^k \cdot (1-\theta), k = 0, 1, 2, \ldots$, for some $0 < \theta < 1$. In this case,

$$P(\{\tau\}|\Omega_n) = \frac{2n+1}{\binom{2n+1}{n}}, \quad \tau \in \Omega_n.$$

Proof. Since $f(k) > 0$ for all k, each $\tau \in \Omega_n$ has positive probability, $n = 1, 2, \ldots$. The total number of trees in Ω_n is $\frac{1}{2n+1}\binom{2n+1}{n}$ as may be counted for the corresponding contour paths using the reflection principle (Chapter 3). Suppose $f(k) = \theta^k \cdot (1-\theta), k \geq 0$, is the (shifted) geometric offspring distribution and let τ be an arbitrary tree rooted at $\langle 1 \rangle$ with total progeny n. Note that the probability of a singleton $\{\tau\}, \tau \in \Omega_n$ may be computed from (14.5) by passing to a limit since $\{\tau\} = \cap_{n=1}^{\infty} B_{\frac{1}{n}}(\tau)$, and for $\tau \in \Omega_n$ one has $\tau|n = \tau$ fall all n sufficiently large. Thus trees τ with total progeny n have, up to normalizing constant depending only on n, probability

$$\prod_{v \in \tau} f(\#v) = (1-\theta)^{n+1} \theta^{\sum_{v \in \tau} \#(v)} = (1-\theta)^{n+1} \theta^n,$$

where $\#(v)$ denotes the number of offspring of v. Thus the probability is the same for all trees with n progeny and, therefore, the conditional probability is the reciprocal of the number of such trees. For the converse proceed by induction. First note that in either case $n = 0$ or $n = 1$ there is only one path with probability one. For $n \geq 2$ define two trees with total progeny n by $\tau^{(1)} = \{\langle 1 \rangle, \langle 1j \rangle : j = 1, \ldots, n\}$ and $\tau^{(2)} = \{\langle 1 \rangle, \langle 1j \rangle : j = 1, 2, \ldots, n-1, \langle 111 \rangle\}$. By the induction hypothesis these trees are equally likely among trees with given total progeny n. The (conditional) probability of $\tau^{(1)}$, given n total progeny, is $C^{-1} f(n) f^n(0)$ and that of $\tau^{(2)}$ is $C^{-1} f(n-1) f^{n-1}(0) f(1)$, where C^{-1} is a positive normalizing constant. Equating these conditional probabilities one has that $\frac{f(n)}{f(n-1)} = \frac{f(1)}{f(0)}$. Now simply iterate this relation to complete the proof. ■

Corollary 22.3. For the Bienaymé–Galton–Watson distribution defined by a (shifted) geometric offspring distribution and single progenitor, the conditional distribution of the family tree τ given that $\|\tau\| = n$ coincides with the conditional distribution of a simple symmetric random walk starting at 1 and conditioned to reach 0 for the first time in $2n + 1$ steps.

Remark 22.1. One may check that the conditional distribution of the simple random walk $S_n := 1 + Y_1 + \cdots + Y_n, n \geq 1, S_0 = 1$, given $[S_0 = 1, S_{2n+1} = 0]$

does not depend on the probability $p \in (0,1)$ defining $P(Y_k = +1) = p = 1 - P(Y_k = -1)$. Thus the statement of the corollary extends accordingly to conditioned asymmetric simple random walks.

Let us now consider the simple symmetric random walk $S_0, S_1, S_2, \ldots$ started at $S_0 = 1$ and defined by i.i.d. equally likely Bernoulli ± 1 displacements $Y_1, Y_2, \ldots,$ i.e., $P(Y_k = 1) = P(Y_k = -1) = 1/2, k \geq 1$, and $S_n = 1 + Y_1 + Y_2 + \cdots Y_n, n \geq 1$. The associated *simple random walk reflected at* 0 may be defined by $R_n = |S_n|, n = 0, 1, 2 \ldots$. A graph of the sample paths reveals a sequence of contour path excursions which, in view of the following corollary and the strong Markov property, define a sequence of contours of i.i.d. Bienaymé–Galton–Watson distributed random trees.

Corollary 22.4. Let $N = ||\tau||$ denote the total progeny of a Bienaymé–Galton–Watson random tree τ, having the (shifted) geometric offspring distribution with $\theta = 1/2$. Then $2N + 1$ is distributed as $T_0 = \inf\{n : R_n \equiv |S_n| = 0\}$.

Proof. Observe that by conditioning on X_1 and using Proposition 14.3, the probability generating function $g(s) = \mathbb{E}s^N$ of $N = \sum_{n=1}^{\infty} X_n$ satisfies

$$g(s) = \mathbb{E}\{(sg(s))^{X_1}\} = \frac{1}{2}\frac{1}{1 - \frac{1}{2}sg(s)} = \frac{1}{2 - sg(s)}, \quad 0 \leq s \leq 1.$$

Thus, for $0 < s < 1$, the quadratic equation $sg^2(s) - 2g(s) + 1 = 0$ yields the solution

$$g(s) = \frac{1 - \sqrt{1-s}}{s} = \sum_{n=0}^{\infty} \binom{1/2}{n+1}(-1)^n s^n, \tag{22.4}$$

and therefore $P(N = n) = (-1)^n\binom{1/2}{n+1} = \frac{1}{n+1}\binom{2n}{n}2^{-(2n+1)}$, for $n = 0, 1, 2 \ldots,$ which agrees with the distribution of the hitting time at 0. ■

Remark 22.2 (Otter-Dwass Formulae). Corollary 22.4 may be viewed as a special case of the so-called Otter-Dwass formula for the distribution of total progeny in a critical branching process. That is, consider a discrete parameter Bienaymé–Galton–Watson branching process having single initial progenitor $Y_0 = 1$ and subcritical/critical offspring distribution $p_j, j \geq 0$. For this shift the contour paths to start at $S_0 = 0$ and end at $S_{2n+1} = -1$. The resulting contour walk will still be denoted by $\{S_n : n \geq 0\}$. According to the Otter-Dwass formulae in this representation, the total progeny $N = \sum_{j=0}^{\infty} Y_j$ of the subcritical or critical branching process coincides with the hitting time of -1 by $\{S_n : n \geq 0\}$, $S_0 = 0$, having increment distribution $p_{j+1}, j \geq -1$. The contour process clearly has skip-free sample paths to the left. Thus combining random walk properties, e.g., as in the case Corollary 22.3, with Kemperman's formula from Chapter 3, one obtains the following

$$P(N = n) = P(T_{-1} = n) = \frac{1}{n} P(S_n = -1). \tag{22.5}$$

This formula relating the total progeny distribution with the position of the left skip-free contour walk is often referred to as *Otter's formula* in the case of a single progenitor $Y_0 = 1$. The extension to $Y_0 = k \geq 2$ is also possible for the critical shifted geometric distribution (Exercise 22), making it a special case of *Dwass' formula*. In this case the tree labeling produces a forest of k genealogical labelings. What we have proven is only valid for the critical offspring distribution because one must show that the contour process is generally distributed as a random walk to apply Kemperman's formula. Proofs in the generality of subcritical/critical offspring distributions are both notationally cumbersome and index[authors]LeGall tricky.[1]

The following theorem provides a basic connection between "excursions" of reflected Brownian motion (defined below) and scaling of Bienaymé–Galton–Watson trees having the (shifted) geometric offspring distribution.

Remark 22.3. An extension[2] of this limit theorem to arbitrary critical offspring distributions with finite second moment is possible. This non-trivial theorem involves the construction of a secondary path process, referred to as the *Lukasiewicz path*, with i.i.d. increments having finite second moment and distributed as the random walk with increment distribution $p_{k+1}, k \geq -1$, up to the hitting time of -1, and such that the indicated weak convergence to reflected Brownian motions implies the same for the contour process.

Theorem 22.5. Let $\tau_1, \tau_2, \ldots$ be an i.i.d. sequence of Bienaymé–Galton–Watson random trees having the critical shifted geometric offspring distribution $f(k) = 2^{-k-1}, k = 0, 1, 2, \ldots$, each with a single progenitor. Denote the contour path of τ_j by $\{(k, s_k^{(j)}) : k = 0, 1, \ldots 2N_j + 1\}$ where $N_j = ||\tau_j||$. Define a continuation of successive contour paths $(n, s_n), n = 0, 1, 2, \ldots$ by $s_0 = 1$ and

$$s_n = s^{(j)}_{n-T^{(j-1)}-1}, \qquad T^{(j-1)} + 1 \leq n \leq T^{(j)},\ n, j \geq 1,$$

where $T^{(0)} = 0, T^{(j)} := T^{(j-1)} + 2N_j + 2, j \geq 1$. Next define a polygonal path continuous extension by linear interpolation $\{C_t^{(n)}, t \geq 0\}$ of the nodes $\{(k/n, s_k/\sqrt{n}) : k = 0, 1, \ldots\}$. Then $C_t^{(n)}$ converges weakly to the reflected Brownian motion $\{|B(t)| : t \geq 0\}$ where $\{B(t) : t \geq 0\}$ is standard Brownian motion starting at 0.

Proof. This follows immediately from the functional central limit theorem since $x \to |x|$ is a continuous function and, in view of Corollary 22.3 and Proposi-

[1]See Le Gall (2005) and Pitman (1997) for two different approaches.

[2]See Le Gall (2005), Aldous (1993).

tion 22.4, $\{S_k : k \geq 0\}$ has the same distribution as $\{|S_n| : n \geq 0\}$, where $\{S_n : n \geq 0\}$ is a simple symmetric random walk started at 1. ■

This result naturally motivates[3] an analysis of (positive) excursions of reflected Brownian motion from a perspective of branching processes. Recalling Proposition 7.25, $T := \inf\{t > 0 : |B(t)| = 0\} \equiv 0$ with probability one since $B(0) = 0$. In view of this a more suitable definition of Brownian excursions may be formulated as follows.

***Definition 22.2** (Brownian Excursion).* Let $B = \{B(t) : t \geq 0\}$ be standard Brownian motion starting at $B(0) = 0$ and defined on a probability space $(\Omega, \mathcal{F}, P)$. For $\omega \in \Omega$ such that $B(1, \omega) \neq 0$ define the *positive excursion path* of reflected Brownian motion $t \to |B(t, \omega)|$ about $t = 1$ by

$$B^{+*}(t, \omega) := \frac{|B(tR(\omega) + (1-t)L(\omega), \omega)|}{\sqrt{R(\omega) - L(\omega)}} \quad 0 \leq t \leq 1, \tag{22.6}$$

where $L := \sup\{s \leq 1 : |B(s)| = 0\}$ and $R := \inf\{s \geq 1 : |B(s)| = 0\}$. Define the excursion path $B^{+*}(t, \omega) := 0$ for each $0 \leq t \leq 1$ in the case $B(1, \omega) = 0$.

That is, the "excursion"is over the maximal zero-free interval about $t = 1$. The interval (L, R) is referred to as the *excursion interval about* $t = 1$. Recall from Chapter 18, Theorem 18.3, and Chapter 19, Exercise 6, that L has the arc-sine distribution

$$P(L \leq t) = \frac{2}{\pi} \sin^{-1}(\sqrt{t}).$$

A related process is the so-called *Brownian meander*, referring to the portion of the excursion ending at $t = 1$; see Exercise 22.

In preparation for the proof of the following proposition, define simple function approximations by

$$L_n = \sum_{j \leq 2^n} \frac{j-1}{2^n} \mathbf{1}_{[(j-1)2^{-n} \leq L < j2^{-n}]},$$

$$R_n = \sum_{k \geq 2^n} \frac{k}{2^n} \mathbf{1}_{[k2^{-n} \leq R < (k+1)2^{-n}]},$$

$$B_n^{+*}(t) = \frac{|B(tR_n + (1-t)L_n)|}{\sqrt{R_n - L_n}}, \qquad 0 \leq t \leq 1.$$

[3] Ideas related to the existence of continuum trees are developed in Aldous (1991) with corresponding functional limit theorems and invariance principles.

Proposition 22.6. $\{B^{+*}(t) : 0 \le t \le 1\}$ is a Markov process with continuous sample paths and (non-homogeneous) transition probabilities[4] $p^*(s,t;y,dz)$ given by

$$p^{+*}(0,t;0,dz) = \frac{2z^2}{\sqrt{2\pi t^3(1-t)^3}} e^{-\frac{z^2}{2t(1-t)}} dz, \quad 0 < t < 1,\ z > 0,$$

and for $0 < s < t < 1,\ y, z > 0$

$$p^{+*}(s,t;y,dz) = \{\frac{e^{-\frac{(z-y)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}} - \frac{e^{-\frac{(z+y)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}\}(\frac{1-s}{1-t})^{\frac{3}{2}} \frac{ze^{-\frac{z^2}{2(1-t)}}}{ye^{-\frac{y^2}{2(1-s)}}} dz.$$

Moreover, the excursion process is independent of the Brownian paths up to time L and the length of the excursion interval (L, R). The excursion is also independent of the Brownian path after time R.

Proof. In addition to the computation of transition probabilities we will show that the excursion is independent of the Brownian paths up to time L and the length of time between L and R. The independence of the excursion with the process after time R follows from the strong Markov property since $R = 1 + \tau_0(B^+(1))$ and $\tau_0(B^+(1)) := \inf\{t \ge 0 : B(1+t) = 0\}$ is a stopping time. In particular R is a stopping time (Exercise 22). Fix $0 < s_1 < \cdots < s_l$, $0 < t_1 < t_2 < \cdots < t_m < 1$, $m \ge 1$. Let g_1 be a bounded continuous function on $\mathbb{R}^m$, g_2 a bounded continuous function on $[0,1] \times [1,\infty)$ such that, for some $\epsilon > 0$, $g_2(s,t) = 0$ for $t - s < \epsilon$ and for $t - s > 1/\epsilon$, and let g_3 be a bounded continuous function on $[0,\infty)^m$. Consider the quantity

$$M_0 := \mathbb{E}_0\{g_1(B(s_1 \wedge L), \ldots, B(s_l \wedge L))g_2(L,R)g_3(B^{+*}(t_1), \ldots, B^{+*}(t_m))\}.$$

Then, in terms of the simple function approximations one has

$$\begin{aligned}
&g_1(B(s_1 \wedge L_n), \ldots, B(s_l \wedge L_n))g_2(L_n, R_n)g_3(B_n^{+*}(t_1), \ldots, B_n^{+*}(t_m)) \\
&= \sum_{j \le 2^n, k \ge 2^n} g_1(B(s_1 \wedge (j-1)2^{-n}), \ldots, B(s_l \wedge (j-1)2^{-n}))g_2(j2^{-n}, k2^{-n}) \\
&\times g_3(\frac{|B(t_1\Delta + j2^{-n})|}{\sqrt{\Delta}}, \ldots, \frac{|B(t_m\Delta + j2^{-n})|}{\sqrt{\Delta}})\mathbf{1}_{[(j-1)2^{-n} \le L < j2^{-n}, k2^{-n} \le R < (k+1)2^{-n}]},
\end{aligned}$$

where $\Delta \equiv \Delta(j,k,n) := (k-j)2^{-n}$. Thus,

[4]The derivation follows the masterful calculations of Itô and McKean (1974), with the trivial caveat that we consider excursions of the reflected Brownian motion while they consider signed excursions of Brownian motion. Obviously the sign does not change over the excursion interval in any case.

$$M_0 = \lim_n \sum_{j \le 2^n, k \ge 2^n} \mathbb{E}_0\{g_1(B(s_1 \wedge (j-1)2^{-n}), \ldots, B(s_l \wedge (j-1)2^{-n}))$$
$$\times g_2(j2^{-n}, k2^{-n}) g_3(\frac{|B(t_1\Delta + j2^{-n})|}{\sqrt{\Delta}}, \ldots, \frac{|B(t_m\Delta + j2^{-n})|}{\sqrt{\Delta}})$$
$$\times \mathbf{1}_{[(j-1)2^{-n} \le L < j2^{-n}, k2^{-n} \le R < (k+1)2^{-n}]}\}. \tag{22.7}$$

Note that $j2^{-n} < t_1\Delta + j2^{-n} < t_m\Delta + j2^{-n} \le k2^{-n}$ and $s_l \wedge (j-1)2^{-n} \le j2^{-n}$. The next step is to condition the term-wise expectations on the respective $\sigma-$fields $\mathcal{F}_{n,j} := \sigma(\mathbf{1}_{[(j-1)2^{-n} \le L < j2^{-n}]}, B(s), s \le j2^{-n})$. Notice that the σ-field gives the (unsigned) positions of the Brownian motion $B(\frac{j}{2^n})$ and the event that $(j-1)2^{-n} \le L < j2^{-n}$. From here

$$M_0 = \lim_n \sum_{j \le 2^n, k \ge 2^n} \mathbb{E}_0\{g_1(B(s_1 \wedge (j-1)2^{-n}), \ldots, B(s_l \wedge (j-1)2^{-n}))g_2(j2^{-n}, k2^{-n})$$
$$\times \mathbf{1}_{[(j-1)2^{-n} \le L < j2^{-n}]}$$
$$\times \mathbb{E}_0[\mathbf{1}_{[k2^{-n} \le R < (k+1)2^{-n}]} g_3(\frac{|B(t_1\Delta + j2^{-n})|}{\sqrt{\Delta}}, \ldots, \frac{|B(t_m\Delta + j2^{-n})|}{\sqrt{\Delta}})|\mathcal{F}_{n,j}]\}.$$

To compute the indicated conditional expectation recall the transition density $p^{(0)}(t; x, y) = \frac{1}{\sqrt{2\pi t}}\{e^{-\frac{(x-y)^2}{2t}} - e^{-\frac{(x+y)^2}{2t}}\}$ of Brownian motion on $(0, \infty)$ viewed up to the time to reach 0. Be mindful for the calculation to follow that according to these transition probabilities the process may either be positive or negative prior to reaching 0. Then, noting that $k2^{-n} - (t_m\Delta + j2^{-n}) = (1 - t_m)\Delta$, one has

$$\mathbb{E}_0\{\mathbf{1}_{[k2^{-n} \le R < (k+1)2^{-n}]} g_3(\frac{|B(t_1\Delta + j2^{-n})|}{\sqrt{\Delta}}, \ldots, \frac{|B(t_m\Delta + j2^{-n})|}{\sqrt{\Delta}})|\mathcal{F}_{n,j}\}$$
$$= \int_{[0,\infty)^m} \int_{[0,\infty)} g_3(\frac{y_1}{\sqrt{\Delta}}, \ldots, \frac{y_m}{\sqrt{\Delta}}) \frac{1}{2} p^{(0)}(t_1\Delta; B(\frac{j}{2^n}), y_1)$$
$$\times p^{(0)}((t_2 - t_1)\Delta; y_1, y_2) \cdots p^{(0)}((t_m - t_{m-1})\Delta; y_{m-1}, y_m) p^{(0)}((1 - t_m)\Delta; y_m, z)$$
$$P_z(\tau_0 < 2^{-n}) dz dy_m \cdots dy_1. \tag{22.8}$$

The factor of $\frac{1}{2}$ occurs as the probability of a positive excursion from 0. Next make the change of variable $y_i = \sqrt{\Delta} z_i$, $1 \le i \le m$, to write

$$p^{(0)}(t\Delta; \sqrt{\Delta}x, \sqrt{\Delta}y) = p^{(0)}(t; x, y)(\sqrt{\Delta})^{-1}$$

and observe by comparing definitions that

$$p^{(0)}(t - s; y, z) = p^{+*}(s, t; y, z)(\frac{1-t}{1-s})^{\frac{3}{2}} \frac{y e^{-\frac{y^2}{2(1-s)}}}{z e^{-\frac{z^2}{2(1-t)}}}.$$

In particular, this brings the asserted transition probabilities explicitly into the calculation by substituting these for the respective factors of the product $p^{(0)}((t_2 - t_1)\Delta; y_1, y_2) \cdots p^{(0)}((t_m - t_{m-1})\Delta; y_{m-1}, y_m)$. Remarkably, the telescoping cancellations in the resulting product render it as

$$\prod_{j=2}^{m} p^{+*}(t_{j-1}, t_j; z_{j-1}, z_j)(\frac{1-t_j}{1-t_{j-1}})^{\frac{3}{2}} \frac{z_{j-1} e^{-\frac{z_{j-1}^2}{2(1-t_{j-1})}}}{z_j e^{-\frac{z_j^2}{2(1-t_j)}}}. \tag{22.9}$$

To express the expectation in terms of the transition probabilities $p^{+*}(s, t; x, y)$ it is convenient to define (for $y = B(\frac{j}{2^n})$)

$$\begin{aligned} & r(\Delta, y, z_1, z_m, z) \\ &= \frac{p^{(0)}(t_1\Delta; y, \sqrt{\Delta} z_1) p^{(0)}((1-t_m)\Delta; \sqrt{\Delta} z_m, z)}{2p^{+*}(0, t_1; 0, z_1) p^{(0)}(\Delta; y, z)} (\frac{1-t_m}{1-t_1})^{\frac{3}{2}} \frac{z_1 e^{-\frac{z_1^2}{2(1-t_1)}}}{z_m e^{-\frac{z_m^2}{2(1-t_m)}}} \sqrt{\Delta}. \end{aligned}$$

The denominator of r permits the introduction of the two natural factors $p^{+*}(0, t_1; 0, z_1) p^{(0)}(\Delta; B_{\frac{j}{2^n}}, z)$ as follows:

$$\begin{aligned} & \mathbb{E}_0[g_3(\frac{|B(t_1\Delta + j2^{-n})|}{\sqrt{\Delta}}, \ldots, \frac{|B(t_m\Delta + j2^{-n})|}{\sqrt{\Delta}}) | \mathcal{F}_{n,j}] \\ &= \int_{[0,\infty)^m} \int_{[0,\infty)} g_3(z_1, \ldots, z_m) \frac{1}{2} p^{(0)}(t_1\Delta; B(\frac{j}{2^n}), \sqrt{\Delta} z_1) \\ & \times p^{+*}(t_1, t_2; z_1, z_2) \cdots p^{+*}(t_{m-1}, t_m; z_{m-1}, z_m) p^{(0)}((1-t_m)\Delta; \sqrt{\Delta} z_m, z) \\ & \times (\frac{1-t_m}{1-t_1})^{\frac{3}{2}} \frac{z_1 e^{-\frac{z_1^2}{2(1-t_1)}}}{z_m e^{-\frac{z_m^2}{2(1-t_m)}}} P_z(\tau_0 < 2^{-n}) dz dy_m \cdots dy_1 \\ &= \int_{[0,\infty)^m} \int_{[0,\infty)} g_3(z_1, \ldots, z_m) p^{(0)}(\Delta; B(\frac{j}{2^n}), z) p^{+*}(0, t_1; 0, z_1) p^{+*}(t_1, t_2; z_1, z_2) \cdots \\ & \cdot p^{+*}(t_{m-1}, t_m; z_{m-1}, z_m) P_z(\tau_0 < 2^{-n}) r(\Delta, B\frac{j}{2^n}, z_1, z_m, z) dz dy_m \cdots dy_1. \end{aligned} \tag{22.10}$$

Now, with a little algebraic manipulation (expanding the quadratic terms in the exponentials defining $p^{(0)}$) one can write the factors appearing in the expression for r in terms of $\sinh(x) = \frac{e^x - e^{-x}}{2}$ as

$$p^{(0)}(\Delta; y, z) = \frac{2}{\sqrt{2\pi\Delta}} e^{-\frac{y^2+z^2}{2\Delta}} \sinh(\frac{yz}{\Delta}),$$

$$p^{(0)}(t_1\Delta; y, \sqrt{\Delta} z_1) = \frac{2e^{-\frac{1}{2\Delta t_1}(\Delta z_1^2 + y^2)}}{\sqrt{2\pi\Delta t_1}} \sinh(\frac{yz_1}{\sqrt{\Delta} t_1}),$$

$$p^{(0)}((1-t_m)\Delta;\sqrt{\Delta}z_m,z)=\frac{2e^{-\frac{1}{2\Delta(1-t_m)}(z^2+\Delta z_m^2)}}{\sqrt{2\pi\Delta(1-t_m)}}\sinh(\frac{zz_m}{\sqrt{\Delta}(1-t_m)}),\qquad(22.11)$$

and

$$p^{+*}(0,t_1;0,z_1)=\frac{2z_1^2}{\sqrt{2\pi t_1^3(1-t_1)^3}}e^{-\frac{z_1^2}{2t_1(1-t_1)}}.$$

Performing the indicated multiplications, next observe that the factor $r = r(\Delta, y, z_1, z_m, z)$ may be expressed as

$$r=\frac{\frac{\sqrt{\Delta}t_1}{yz_1}\sinh(\frac{yz_1}{\sqrt{\Delta}t_1})\frac{\sqrt{\Delta}(1-t_m)}{zz_m}\sinh(\frac{zz_m}{\sqrt{\Delta}(1-t_m)})}{2\frac{\Delta}{yz}\sinh(\frac{yz}{\Delta})}\exp\{-\frac{y^2(1-t_1)}{2\Delta t_1}-\frac{z^2t_m}{2\Delta(1-t_m)}\}.$$

In particular the support of g_2 makes r bounded within the domain of integration and, using l'Hôpital's rule $\lim_{h\to 0}\frac{\sinh(ch)}{ch}=1$, so that $r\to 1/2$ as $y,z\to 0$. Substituting the expression for the unconditional expectation into M_0 and passing to the limit as $n\to\infty$ one obtains (noting $B(L)=0$ and $P_z(\tau_0<2^{-n})dz\Rightarrow\delta_0(dz)$, as $n\to\infty$),

$$\begin{aligned}M_0=\lim_n&\sum_{j\le 2^n,k\ge 2^n}\mathbb{E}_0\{g_1(B(s_1\wedge(j-1)2^{-n}),\ldots,B(s_l\wedge(j-1)2^{-n}))\\
&\times\mathbf{1}_{[(j-1)2^{-n}\le L<j2^{-n},k2^{-n}\le R<(k+1)2^{-n}]}\}\\
&\times\int_{[0,\infty)}g_2(j2^{-n},k2^{-n})p^{(0)}(\Delta;B(\frac{j}{2^n}),z)P_z(\tau_0<2^{-n})r(\Delta,B(\frac{j}{2^n}),z_1,z_m,z)dz\}\\
&\times\int_{[0,\infty)^m}g_3(z_1,\ldots,z_m)p^{+*}(0,t_1;0,z_1)p^{+*}(t_1,t_2;z_1,z_2)\\
&\cdots p^{+*}(t_{m-1},t_m;z_{m-1},z_m)dz_m\cdots dz_1\\
&=\mathbb{E}_0\{g_1(B(s_1\wedge L),\ldots,B(s_m\wedge L))g_2(L,R)\}\\
&\quad\times\int_{[0,\infty)^m}g_3(z_1,\ldots,z_m)p^{+*}(0,t_1;0,z_1)p^{+*}(t_1,t_2;z_1,z_2)\\
&\quad\cdots p^{+*}(t_{m-1},t_m;z_{m-1},z_m)dz_m\cdots dz_1.\end{aligned}$$

Since, letting $\epsilon\downarrow 0$ in the support of g_2, the functions g_1, g_2, g_3 are arbitrary bounded continuous functions this completes the proof. In particular taking g_1, g_2 identically one proves the Markov property and identifies the transition probability densities for the excursion. ■

Remark 22.4. A somewhat unsatisfying aspect of the above proof is that it relies on a-priori formula for p^{+*} that is then verified. On the other hand, it is a testimony to the power of computation by brute force simple function approximations and conditional expectation.

The literature on the distributions of various functionals of the Brownian excursion is quite extensive and results are obtainable by diverse methods,[5] including the distribution of the length of the excursion interval, and the distribution of the maximum of the Brownian excursion. Also the material in this chapter only scratches the surface of the deep index[authors]LeGall connections[6] between random trees and excursions. Some additional insights may be obtained from the computations outlined in the exercises.

Exercises

1. (*Dwass' Formula*) Let $k \geq 2$ and extend the formula (22.5) for the total progeny distribution in the case of a critical shifted geometric offspring distribution and the number of initial number of progenitors is $Y_0 = k$. [*Hint*: Refine the genealogical labeling to encode the dependence on each progenitor before applying the Kemperman formula.]
2. (*Borel–Tanner Distribution*) Let N be the total progeny in a branching process with a single progenitor and Poisson distribution with mean $\lambda \leq 1$. Show that $P(N = n) = \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}, n \geq 1$. [*Hint*: Use the Otter-Dwass' formula.]
3. Show that R in Definition 22.2 is a stopping time and the Brownian motion after time R is independent of the excursion process.
4. (*Brownian Meander*) The simple random walk starting at 0 and conditioned to remain positive until a first return at 0 at time $2n$ will be referred to as *simple random walk excursion over* 0 *to* $2n$ and denoted $\{S_k^{+*} : k = 0, 1, 2, \ldots, 2n\}$. The simple random walk starting at 0 and conditioned to remain positive over time $2n - 1$ is a process denoted $\{S_k^{+} : k = 0, 1, \ldots, 2n\}$ and referred to as *simple random walk meander over* 0 *to* $2n$. This exercise concerns limiting finite-dimensional distributions of the polygonal processes $\{\tilde{X}_t^{(n)+} : 0 \leq t \leq 1\}$ and $\{\tilde{X}_t^{(n)+*} : 0 \leq t \leq 1\}$ obtained by (continuous) linear interpolations of the points $(\frac{k}{2n}, \frac{S_k^+}{\sqrt{2n}}), k = 0, 1, \ldots, 2n$, and $(\frac{k}{2n}, \frac{S_k^{+*}}{\sqrt{2n}}), k = 0, 1, \ldots, 2n$, conditionally given $[X_1 > 0, T_0 > 2n]$ and $[X_1 > 0, T_0 = 2n]$, respectively, cf. (17.4). The goal is to compute the following limits.

---

[5]Much of this originated with Lévy (1945, 1965). Itô and McKean (1963) exploits this in their construction of one-dimensional diffusions as Markov processes on $[0, \infty)$ subject to a general analytic classification of possible boundary conditions due to Feller. Also see Chung (1976) and numerous references therein, for historic remarks and alternative approaches to excursions.

[6]While this chapter is limited to defining the processes and developing a few of their basic properties, elaborate theoretical developments with applications can be found in Aldous (1993); Neveu and Pitman (1989); Pitman (2002); Janson (2007); Le Gall (2005); Lyons and Peres (2016); Kovchegov and Zaliapin (2020).

(a)

$$\lim_{n\to\infty} P(\tilde{X}_{t_1}^{(n)+} \le x_1, \ldots, \tilde{X}_{t_k}^{(n)+} \le x_k)$$
$$= \int_0^{x_1} \cdots \int_0^{x_k} p^+(0,0;t_1,y_1)\ldots p^+(t_{k-1},y_{k-1};t_k,y_k)dy_k \ldots dy_1,$$

where, letting $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$, and $\Phi(z) = \int_{-\infty}^{z}\varphi(u)du$, for $x, y > 0$, $0 < s < t \le 1$,

(b)

$$p^+(s,x;t,y) = \frac{1}{\sqrt{2\pi(t-s)}}\{e^{-\frac{(y-x)^2}{2(t-s)}} - e^{-\frac{(x+y)^2}{2(t-s)}}\}\{\frac{\Phi(\frac{y}{\sqrt{1-t}}) - \Phi(-\frac{y}{\sqrt{1-t}})}{\Phi(\frac{x}{\sqrt{1-s}}) - \Phi(-\frac{x}{\sqrt{1-s}})}\}$$

$$p^+(0,0;t,y) = t^{-\frac{3}{2}}ye^{-\frac{y^2}{2t}}\{\Phi\left(\frac{y}{\sqrt{1-t}}) - \Phi(-\frac{y}{\sqrt{1-t}}\right)\},\ y > 0,\ 0 < t \le 1$$

$$P(\tilde{X}_{t_1}^{(n)+*} \le x_1, \ldots, \tilde{X}_{t_k}^{(n)+*} \le x_k)$$
$$= \int_0^{x_1} \cdots \int_0^{x_k} p^{+*}(0,0;t_1,y_1)p^{+*}(t_1,y_1;t_2,y_2)\ldots p^{+*}(t_{k-1},y_{k-1};t_k,y_k)dy_k \ldots dy_1,$$

where for $0 < s < t < 1$, $y, x > 0$,

$$p^{+*}(s,x;t,y) = \frac{(1-s)^{3/2}}{(1-t)^{3/2}}\frac{y}{x}\frac{\varphi(\frac{y}{\sqrt{1-t}})}{\varphi(\frac{x}{\sqrt{1-s}})}(t-s)^{-\frac{1}{2}}\{\varphi(\frac{y-x}{\sqrt{t-s}}) - \varphi(\frac{y+x}{\sqrt{t-s}})\},$$

$$p^{+*}(0,0;t,y) = 2(2\pi t^3(1-t)^3)^{-\frac{1}{2}}y^2\exp[-y^2/2t(1-t)]$$

are the transition probabilities for the positive excursion process of reflected Brownian motion. The calculations of the finite-dimensional limit distributions will use the m-dimensional version ($m \ge 1$) of the local limit theorem (Proposition 16.1) of Chapter 16, the reflection principle for simple symmetric random walk (Proposition 3.1) of Chapter 3, and Stirling's formula Chapter 3, as well. In particular since, as in (17.4)–(17.6), on the respective sets $[X_1 > 0, T_0 > 2n]$ and $[X_1 > 0, T_0 = 2n]$ the processes differ from the process $\{\frac{S_{[2nt]}}{\sqrt{2n}} : 0 \le t \le 1\}$ at time points $t_1 < \cdots < t_m$ by no more than $\frac{1}{\sqrt{2n}} = o(1)$, it is sufficient to consider $\{\frac{S_{[2nt]}}{\sqrt{2n}} : 0 \le t \le 1\}$ conditionally. Let $Y_1, Y_2, \ldots$ be i.i.d. ± 1-valued (Bernoulli) random variables with $P(Y_i = +1) = P(Y_i = -1) = \frac{1}{2}$. Let $S_0 = 0$, $S_n = Y_1 + \cdots + Y_n$, $n \ge 1$. Let $T_0 = \inf\{n \ge 1 : S_n = 0\}$.

(i) Show, for n even, $P(T_0 > 2n) = P(S_{2n-1} = -1) = P(S_{2n} = 0) \sim \sqrt{\frac{2}{\pi}}(2n)^{-\frac{1}{2}}$ [*Hint*: Use the reflection principle, symmetry and (2.2).]

(ii) $P(T_0 = 2n) = \frac{1}{2n-1}P(S_{2n} = 0) \sim \sqrt{\frac{2}{\pi}}(2n)^{-\frac{3}{2}}$[*Hint*: The assertion (ii) follows similarly or by direct application of (i).] in the sense that the ratios converge to 1 as $n \to \infty$.

(iii) Let $a > 0$ $b \ge 0$ and let k denote an arbitrary positive integer with $|b-a| \le k$ and such that $b-a$ and k have the same parity. Show $P(S_k = b|S_0 = a) = \binom{k}{\frac{k+b-a}{2}}\left(\frac{1}{2}\right)^k$.

(iv) Show for $a > 0$ that

$$P(S_k = b, S_i > 0, i \le k|S_0 = a) = P(S_k = b|S_0 = a) - P(S_k = b|S_0 = -a).$$

In the case $a = 0$ show $P(S_k = b, S_i > 0, i \le k) = \frac{b}{k}P(S_k = b)$. [*Hint*: Use the reflection principle in both cases]

(v) Show for $a > 0$, $P(S_i > 0, i \le k|S_0 = a) = P(T_{-a} > k)$, where T_y denotes the time to reach $y \ne 0$ by the simple symmetric random walk starting at 0.

(vi) Show for $y > 0$, that $P(\frac{1}{2n}T_{z\sqrt{2n}} > 1 - \frac{[2nt]}{2n}) \to 1 - P(\tau_{-y} \le 1-t)$. [*Hint*: $\frac{1}{2n}T_{-y\sqrt{2n}}$ converges in distribution to the corresponding hitting time τ_{-y} for Brownian motion by the functional central limit theorem.]

(vii) Show that the preceding probability may be expressed as $1 - 2P(B(1-t) > -y) \equiv 1 - 2P(B(1-t) \le y)$. [*Hint*: Use the reflection principle.]

(viii) Show for $y > 0, 0 < t < 1$,

$$\begin{aligned}
&\frac{\sqrt{2n}}{2}P(\frac{S_{[2nt]}}{\sqrt{2n}} = \frac{[y\sqrt{2n}]}{\sqrt{2n}}|X_1 > 0, T_0 > 2n)\\
&= \frac{\sqrt{2n}}{2}\frac{2}{P(T_0 > 2n)}P(\frac{S_{[2nt]}}{\sqrt{2n}} = \frac{[y\sqrt{2n}]}{\sqrt{2n}}, S_i > 0, 1 \le i \le [2nt])\\
&\quad \times P(S_i > 0, [2nt]+1 \le i \le 2n|S_{[2nt]} = [y\sqrt{2n}])\\
&= \frac{2}{P(T_0 > 2n)}\frac{[y\sqrt{2n}]}{[2nt]}\frac{\sqrt{2n}}{2}P\left(\frac{S_{[2nt]}}{\sqrt{2n}} = \frac{[y\sqrt{2n}]}{\sqrt{2n}}\right)P(\frac{1}{2n}T_{-[y\sqrt{2n}]} > 1 - \frac{[2nt]}{2n})\\
&\sim \frac{y}{t^{\frac{3}{2}}}e^{-\frac{y^2}{2t}}\left\{2\Phi\left(\frac{y}{\sqrt{1-t}}\right) - 1\right\} = \frac{y}{t^{\frac{3}{2}}}e^{-\frac{y^2}{2t}}\{\Phi(\frac{y}{\sqrt{1-t}}) - \Phi\left(-\frac{y}{\sqrt{1-t}}\right)\}.
\end{aligned}$$

[*Hint*: Use the preceding calculations together with the local central limit theorem and sample path combinatorial symmetries.]

(ix) Show for $0 \le s < t$,

$$\begin{aligned}
&(\frac{\sqrt{2n}}{2})^2 P(\frac{S_{[2ns]}}{\sqrt{2n}} = \frac{[x\sqrt{2n}]}{\sqrt{2n}}, \frac{S_{[2nt]}}{\sqrt{2n}} = \frac{[y\sqrt{2n}]}{\sqrt{2n}}|X_1 > 0, T_0 > 2n)\\
&= \frac{2}{P(T_0 > 2n)}\frac{\sqrt{2n}}{2}P(\frac{S_{[2ns]}}{\sqrt{2n}} = \frac{[x\sqrt{2n}]}{\sqrt{2n}}, S_i > 0, 1 \le i \le [2ns])\\
&\quad \times \frac{\sqrt{2n}}{2}P(\frac{S_{[2nt]}}{\sqrt{2n}} = \frac{[y\sqrt{2n}]}{\sqrt{2n}}, S_i > 0, [2ns]+1 \le i \le [2nt]|\frac{S_{[2ns]}}{\sqrt{2n}} = \frac{[x\sqrt{2n}]}{\sqrt{2n}})\\
&\quad \times P(S_i > 0, i = [2nt]+1, \dots, 2n|S_{[2nt]} = \left[y\sqrt{2n}\right])
\end{aligned}$$

$$
\begin{aligned}
&\sim \sqrt{2\pi}\sqrt{2n}\frac{[x\sqrt{2n}]}{[2ns]}\frac{1}{\sqrt{2\pi s}}e^{-\frac{1}{2s}x^2}\frac{1}{\sqrt{2\pi(t-s)}}\{e^{-\frac{1}{2(t-s)}(y-x)^2}\\
&\quad -e^{-\frac{1}{2(t-s)}(y+x)^2}\}\\
&\quad \times\{\Phi(\frac{y}{\sqrt{1-t}})-\Phi(-\frac{y}{\sqrt{1-t}})\} \quad (n\to\infty)\\
&\to \sqrt{\frac{\pi}{2}}\frac{x}{s}\frac{1}{\sqrt{s}}e^{-\frac{1}{2s}x^2}\frac{1}{\sqrt{2\pi(t-s)}}\{e^{-\frac{1}{2(t-s)}(y-x)^2}-e^{-\frac{1}{2(t-s)}(y+x)^2}\}\\
&\quad \times\{\Phi(\frac{y}{\sqrt{1-t}})-\Phi(-\frac{y}{\sqrt{1-t}})\}\\
&= p^+(0,0;s,x)p^+(s,x;t,y).
\end{aligned}
$$

(x) Calculate the higher finite-dimensional distributions in (a). [*Hint*: Condition as in the previous case k=2 and use induction.].

(xi) Provide similarly indicated calculations as above for the limit asserted in (b).

Chapter 23
Special Topic: The Geometric Random Walk and the Binomial Tree Model of Mathematical Finance

The pricing of options has a long mathematical history dating back to Luis Bachelier's remarkable pre-Einstein and Smoluchowski conceptions of Brownian motion. Option pricing is widely recognized among the most natural applications of martingale theory outside of mathematics. In this chapter, the basic discrete space-time model and underlying concepts are introduced in terms of a multiplicative (geometric) random walk. The key mathematical innovations[1] that result are (i) the natural occurrence of the notion of martingale change of measure in terms of arbitrage-freeness and (ii) the issue of martingale uniqueness as it pertains to market completeness.

Suppose that today's price ($t = 0$) of a share of some risky asset, e.g., a share of stock, is S_0. Also suppose that risk-free assets are available, e.g., US Treasury bonds, at today's price of B_0. The standard model for the (discrete time) evolution of bond prices is by deterministic growth at the risk-free interest rate $r \geq 0$. Accordingly,

$$B_{t+1} - B_t = rB_t, \quad t = 0, 1, \ldots \tag{23.1}$$

Solving (23.1) with the initial value B_0, one has

$$B_t = R^t B_0, \quad t = 0, 1, \ldots, \quad R := r + 1 \geq 1. \tag{23.2}$$

[1] A more comprehensive and contemporary treatment of these ideas is given by Foelmer and Schied (2002)

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_23

A standard model for stock prices assumes that in each unit of time, the price may move up (increase) by a factor $u > 1$ with probability p_u or may move down by a factor $d < 1$ with probability $p_d = 1 - p_u$. The probabilities (p_u, p_d) are referred to as the *historical* probabilities. Thus the temporal evolution of stock prices $S_0, S_1, \dots$ is governed by the stochastic difference equation

$$S_{t+1} - S_t = Y_{t+1} S_t, \quad t = 0, 1, 2, \dots, \tag{23.3}$$

where $Y_1, Y_2, \dots$ is an i.i.d. sequence of Bernoulli random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with $P(Y_1 = u - 1) = p_u$ and $P(Y_1 = d - 1) = p_d$. Equivalently, stock prices are said to be distributed as the *geometric random walk* process given by

$$S_t = \prod_{j=1}^{t} Z_j S_0, \ t = 0, 1, \dots, \quad Z_j := 1 + Y_j, \ j = 1, 2, \dots \tag{23.4}$$

Notice that the sequence $\ln S_0, \ln S_1, \ln S_2, \dots$ evolves by additive independent increments. Thus the stochastic process $\{\ln S_t : t = 0, 1, 2, \dots\}$ is an additive random walk. The geometric random walk model for stock prices is commonly referred to as the *binomial tree model*[2] in mathematical finance. This name is derived from the recombining tree graph illustrating possible stock movements $S_0 \to \{uS_0, dS_0\} \to \{uuS_0, udS_0 = duS_0, ddS_0\} \to \cdots$. The probability P is also referred to as the *historic* or *market* probability measure, to distinguish it from other related change of probability measures that arise naturally in this context. It is reasonable to restrict attention to the case of parameters $0 < d < R < u$; otherwise, for example, if $d, u < R$, the availability of risk-free assets at rate R would make the risky asset unattractive to investors.

Definition 23.1. An *option* or *contingent claim* on $S_0, S_1, \dots, S_T$ over a finite time horizon T is identified with a non-negative $\sigma(S_0, \dots, S_T)$-measurable payoff random variable $X = X(S_0, S_1, \dots, S_T)$ representing the value of a contract on evolution of the underlying asset prices.

Example 1. A *(European) call option* is a contract to allow the holder the right to purchase the underlying asset at the *expiration time* T for a previously agreed upon (contracted) *strike price* K. The qualifier "European" refers to contracts covering a time horizon $[0, T]$ that can only be exercised upon expiration time T. The value of the contingent claim at time T to the holder of the contract is its *payoff* $X = (S_T - K)^+$ since the holder will buy S_T at the price K upon the event $[S_T > K]$, but otherwise will not make the purchase, i.e., not exercise the option to buy. The question is to determine a "fair" purchase value of the contract at today's time $t = 0$.

[2]This model is a discretization of the continuous time model originally used for the derivation of the Black–Scholes formula in Black and Scholes (1973). The discretization was introduced by Cox et al. (1976).

Mathematically this is a *final value problem* in which one is given the final value of an evolving process at a specified time and then seeks the initial value.

The problem of pricing options may be viewed from two different perspectives, namely that of the holder and that of the writer of the contract. Let us first consider the pricing question from the perspective of the writer. A *hedging principle* assigns a price π_0, which is sufficient to render the writer's exposure to lose the amount X of the contingent claim, a risk-free exposure by appropriate investment of π_0 into an offsetting self-financing portfolio. For example, consider a $T = 1$ period contract for which the writer is paid π_0. The writer could purchase some φ_0 units of stock and ψ_o units of risk-free bond where

$$\pi_0 = \varphi_0 S_0 + \psi_0 B_0, \tag{23.5}$$

and φ_0 and ψ_0 can at most depend on S_0, but not tomorrow's value of S_1; i.e., $\sigma(S_0)$-measurable. At the end of one period, this portfolio will have the new value given by

$$V_1 = \varphi_0 S_1 + \psi_0 B_1 = \begin{cases} \varphi_0 u S_0 + \psi_0 R B_0 & \text{if } S_1 = uS_0 \\ \varphi_0 d S_0 + \psi_0 R B_0 & \text{if } S_1 = dS_0. \end{cases} \tag{23.6}$$

On the other hand, at the end of the $T = 1$ period, the writer is exposed to a loss in the contingent claim amount

$$X = X(S_0, S_1) = \begin{cases} X(S_0, uS_0) & \text{if } S_1 = uS_0 \\ X(S_0, dS_0) & \text{if } S_1 = dS_0. \end{cases} \tag{23.7}$$

Thus the writer of the contract may seek an amount π_0 sufficiently large to solve

$$V_1 \geq X = X(S_0, S_1) \tag{23.8}$$

for the respective portfolio amounts φ_0 and ψ_0 to hedge the risk of loss. Specifically, solving (23.8) for the least such amount under (23.6) and (23.7), one obtains

$$\varphi_0 = \frac{X(S_0, uS_0) - X(S_0, dS_0)}{uS_0 - dS_0} \qquad \psi_0 = \frac{dX(S_0, uS_0) - uX(S_0, dS_0)}{R(d-u)B_0}. \tag{23.9}$$

Substituting these values into (23.5) and collecting coefficients of $V(S_0, uS_0)$ and $V(S_0, dS_0)$, respectively, it follows that

$$\pi_0 = R^{-1}[X(S_0, uS_0)q_u + X(S_0, dS_0)q_d] = R^{-1}\mathbb{E}_Q\{X(S_0, S_1)|S_0\}, \tag{23.10}$$

where

$$q_u = \frac{R-d}{u-d}, \quad q_d = \frac{u-R}{u-d}, \tag{23.11}$$

and Q is the probability on $(\Omega, \mathcal{F})$ under which the stock price model $S_0, S_1, \ldots$ is distributed as the geometric random walk process

$$S_t = \prod_{j=1}^{t} Z_j S_0, \; t = 0, 1, \ldots, \quad j = 1, 2, \ldots, \tag{23.12}$$

with $Q(Z_1 = u) = q_u$ and $Q(Z_1 = d) = q_d$. The requirement $d < R < u$ imposed at the outset makes Q a probability. The formula (23.10) is the discrete time one-period *Black–Scholes formula* for option valuation; see Figure 23.1 as a guide to the derivation. Observe that it represents the price of the option as a discounted expected value under a *pricing probability* Q, which is generally different from the historic probability P. An agent is willing to accept π_0 for the contract because the risk can be completely removed by investment in a hedging portfolio with value $V_0 = \pi_0$ today and value $V_1 = X$ tomorrow at date of expiry ($T = 1$). Such an agent will sell a large number N of contracts for $\pi_0 + \epsilon$ to make a profit of $N\epsilon$.

The extension of the replicating hedging principle to a T-period contract is as follows: The writer seeks an amount π_0 sufficient to construct an offsetting portfolio defined by random variables $\varphi_t, \psi_t, t = 0, 1, 2, \ldots, T-1$ such that (a) φ_t and ψ_t are each $\mathcal{F}_{t-1} := \sigma(S_0, S_1, \ldots, S_{t-1})$-measurable; (b)$\pi_0 = V_0 := \phi_0 S_0 + \psi_0 B_0 \to V_1 := \varphi_0 S_1 + \psi_0 B_1 = \varphi_1 S_1 + \psi_1 B_1 \to \cdots \to V_T := \varphi_{T-1} S_T + \psi_{T-1} B_T$ such that $V_T \geq X$.

The arrows represent the transformation of the portfolio values due to movements in the underlying risky asset and bond prices, which are reapportioned in the amounts φ_t and ψ_t at each successive step t.

Definition 23.2. Random variables φ_t and ψ_t satisfying (a) are said to be *predictable*, and a *self-financing strategy* refers to a predictable strategy satisfying the condition (b).

Pricing over period T requires a sufficient amount π_0 that the portfolio can be reapportioned by a predictable strategy, i.e., without the aid of a "crystal ball," into shares of stocks and bonds in the course of their evolution such that the self-financing and attainability conditions hold, i.e.,

$$\varphi_t S_t + \psi_t B_t = \varphi_{t-1} S_t + \psi_{t-1} B_t \geq 0, t = 1, 2, \ldots, T-1, \quad \varphi_{T-1} S_T + \psi_{T-1} B_T = V_T. \tag{23.13}$$

The following proposition provides such a price π_0 and self-financing strategy for a given contract amount X.

Proposition 23.1. There is a self-financing hedging strategy (φ_t, ψ_t), $t = 0, 1, \ldots, T-1$ for any given contingency claim X provided $\pi_0 \geq R^{-T}\mathbb{E}_Q\{X|S_0\}$. Moreover, $V_t := \varphi_{t-1} S_t + \psi_{t-1} B_t \geq 0, t = 1, 2, \ldots, T$.

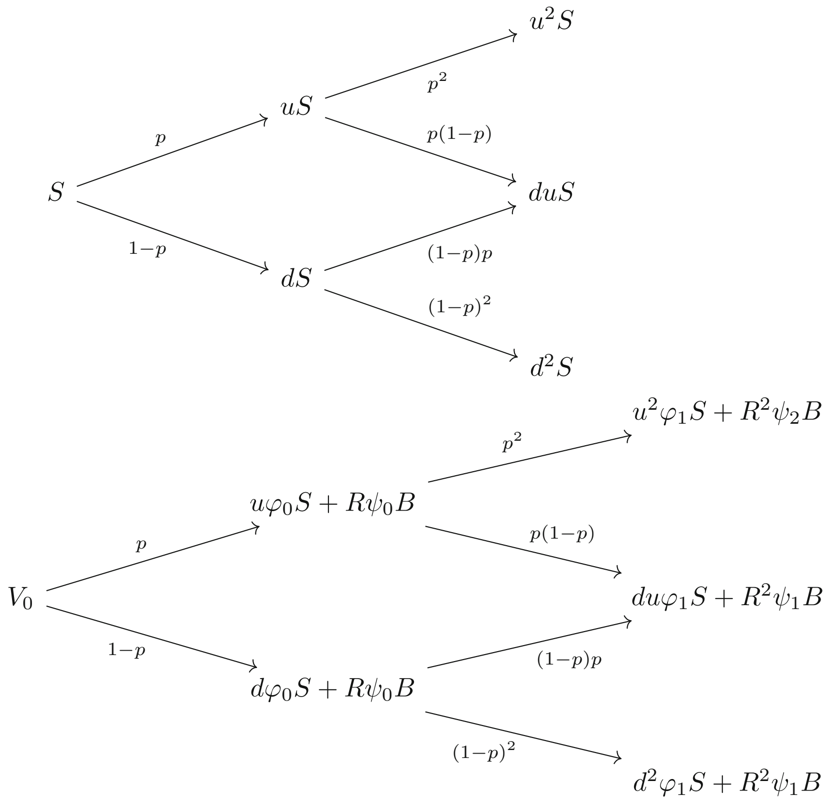


Fig. 23.1 Recombining Binomial Tree

Proof. The case $T = 1$ was solved in the derivation of (23.10). From here one proceeds by induction as follows. Consider the $T - 1$ period contingent claim

$$\tilde{X} = R^{-1}\mathbb{E}_Q(X|\mathcal{F}_{T-1})$$

over the *one period* from $T - 1$ to T. By the induction hypothesis, there is a self-financing trading strategy $\varphi_t, \psi_t, t = 0, 1, \dots, T - 2$, such that

$$\varphi_{T-2}S_{T-1} + \psi_{T-2}B_{T-1} = \tilde{X}$$

for the price

$$\tilde{\pi}_0 = R^{-(T-1)}\mathbb{E}_Q\{\tilde{X}|S_0\} = R^{-(T-1)}\mathbb{E}_Q\{\mathbb{E}_Q(R^{-1}X|\mathcal{F}_{T-1})|S_0\} = R^{-T}\mathbb{E}_Q\{X|S_0\}. \tag{23.14}$$

Now reapportion the amount $\tilde{X} = \varphi_{T-2}S_{T-1} + \psi_{T-2}B_{T-1}$ to complete the hedge over the last period by solving exactly as in the one-period case. That is, the least amount V_T required is $V_T = X$, so that one solves (for $\varphi_{T-1}, \psi_{T-1}$) the equations

$$\begin{aligned}
\tilde{X} &= \varphi_{T-1} S_T + \psi_{T-1} B_T \\
&= \begin{cases} \varphi_{T-1} u S_{T-1} + \psi_{T-1} R B_{T-1} = X(S_0, \ldots, S_{T-1}, u S_{T-1}) & \text{if } S_T = u S_{T-1} \\ \varphi_{T-1} d S_{T-1} + \psi_{T-1} R B_{T-1} = X(S_0, \ldots, S_{T-1}, d S_{T-1}) & \text{if } S_T = d S_{T-1}, \end{cases}
\end{aligned}$$

to obtain

$$\begin{aligned}
\varphi_{T-1} &= \frac{X(S_0, \ldots, S_{T-1}, u S_{T-1}) - X(S_0, \ldots, S_{T-1}, d S_{T-1})}{(u-d) S_{T-1}}, \\
\psi_{T-1} &= \frac{u X(S_0, \ldots, S_{T-1}, d S_{T-1}) - d X(S_0, \ldots, S_{T-1}, u S_{T-1})}{R(u-d) B_{T-1}}.
\end{aligned} \tag{23.15}$$

Also check that

$$\varphi_{T-1} S_{T-1} + \psi_{T-1} B_{T-1} = R^{-1} \mathbb{E}_Q(X|\mathcal{F}_{T-1}) = \tilde{X},$$

so that, by (23.14), $\varphi_{T-1} S_{T-1} + \psi_{T-1} B_{T-1} = \varphi_{T-2} S_{T-1} + \psi_{T-2} B_{T-1}$, completing the induction argument. No additional money is required to complete the hedge from $\tilde{X}$ to X in the last step so that the asserted non-negativity is also preserved by the induction argument. ■

Some additional terminologies used in this framework are as follows.

Definition 23.3. The contingent claim X with expiry T is said to be *attainable at price* π_0 if there is a self-financing strategy $(\varphi_t, \psi_t) : t = 0, 1, \ldots, T-1$, with the associated *market value portfolio process* $V_t = V(\varphi_0, \ldots, \varphi_t) := \varphi_t S_t + \psi_t B_t$, $t = 0, 1, \ldots, T$ such that $V_0 = \pi_0$ and $V_T = X$. Self-financing strategies such that $V_t \geq 0$ for each $t = 0, 1, \ldots, T$ are said to be *admissible*.

So we have proven that every contingent claim is attainable for the binomial tree model with parameters $0 < d < R < u$. This is desirable from the perspective of the individual who sells the contract (writer). Let us now turn to the contract buyer (holder). In particular, in what sense is the price $\pi_0 = R^{-T} \mathbb{E}_Q\{X|S_0\}$ a "fair" price to pay for the contract?

Definition 23.4. A market model is said to be *arbitrage-free* if there does *not* exist a self-financing trading strategy $\varphi_t, \psi_t, t = 0, 1, \ldots, T-1$ such that $\varphi_t S_t + \psi_t B_t = \varphi_{t-1} S_t + \psi_{t-1} B_t \geq 0, t = 1, \ldots, T$, but $\varphi_0 S_0 + \psi_0 B_0 = 0$ and $\mathbb{E}_P(\varphi_{T-1} S_T + \psi_{T-1} B_T) > 0$.

Note that the historic probability P is used in the definition of arbitrage-free market. The meaning is that there is not a self-financing strategy that can start with zero investment and attain a positive return with positive probability. This is a sense in which the writer's price may be viewed as "fair" to the buyer.

Proposition 23.2. The binomial tree model for risky asset prices and risk-free bonds $(S_t, B_t) : t = 0, 1, \dots, T$ is arbitrage-free.

Proof. The proof is based on two basic properties of the pricing measure Q obtained above. First is the existence of a measure Q, which is equivalent to P in the sense that for any event $A \in \mathcal{F}_T = \sigma(S_0, S_1, \dots, S_T)$, one has $P(A) = 0$ if and only if $Q(A) = 0$ (Exercise 1). Secondly, the following property holds under the probability Q

$$\mathbb{E}_Q(R^{-(t+1)}S_{t+1}|\mathcal{F}_t) = R^{-t}S_t, \quad t = 0, 1, \dots, T-1. \tag{23.16}$$

To verify (23.16), simply observe that

$$\begin{aligned}
\mathbb{E}_Q(R^{-(t+1)}S_{t+1}|\mathcal{F}_t) &= R^{-(t+1)}uS_t q_u + R^{-(t+1)}dS_t q_d \\
&= R^{-(t+1)}uS_t \frac{R-d}{u-d} + R^{-(t+1)}dS_t \frac{u-R}{u-d} \\
&= R^{-t}S_t.
\end{aligned}$$

Now, suppose for contradiction, that the model is not arbitrage-free. Then there is a self-financing trading strategy $\varphi_t, \psi_t, t = 0, 1, \dots, T-1$ such that $\varphi_t S_t + \psi_t B_t = \varphi_{t-1}S_t + \psi_{t-1}B_t \geq 0, t = 1, \dots, T$, but $\varphi_0 S_0 + \psi_0 B_0 = 0$ and $\mathbb{E}_P(\varphi_{T-1}S_T + \psi_{T-1}B_T) > 0$. Thus, by equivalence, $\mathbb{E}_Q(\varphi_{T-1}S_T + \psi_{T-1}B_T) > 0$. Now, using the property (23.16), one can iterate backward as follows:

$$\begin{aligned}
0 &< \mathbb{E}_Q(\varphi_{T-1}S_T + \psi_{T-1}B_T) \\
&= \mathbb{E}_Q\{\mathbb{E}_Q(\varphi_{T-1}S_T + \psi_{T-1}B_T|\mathcal{F}_{T-1})\} \\
&= \mathbb{E}_Q\{\varphi_{T-1}\mathbb{E}_Q(S_T|\mathcal{F}_{T-1}) + \psi_{T-1}B_T\} \\
&= \mathbb{E}_Q\{\varphi_{T-1}RS_{T-1} + \psi_{T-1}RB_{T-1}\} \\
&= R\mathbb{E}_Q\{\varphi_{T-1}S_{T-1} + \psi_{T-1}B_{T-1}\} \\
&= R\mathbb{E}_Q\{\varphi_{T-2}S_{T-1} + \psi_{T-2}B_{T-1}\}.
\end{aligned}$$

Iterating the successive conditioning, one arrives at

$$0 < \mathbb{E}_Q(\varphi_{T-1}S_T + \psi_{T-1}B_T) = R^T\mathbb{E}_Q\{\varphi_0 S_0 + \psi_0 B_0\},$$

which is a contradiction. ■

The property (23.16) is the martingale property of the stochastic process $M_t = R^{-t}S_t, t = 0, 1, \dots, T$ on $(\Omega, \mathcal{F}, Q)$ with respect to the filtration $\mathcal{F}_t, t \geq 0$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \sigma(S_0, S_1, \dots, S_t), 1 \leq t \leq T$. For this reason, Q is often referred to as the *martingale measure*.

Definition 23.5. An arbitrage-free market model is said to be *complete* for a time horizon T if every contingent claim X with expiry T is attainable by an admissible strategy.

To summarize what has been obtained so far for the binomial tree model, for a given finite time horizon T, there is a so-called *pricing probability* Q, equivalent to the *historic probability* P on $\sigma(S_0, S_1, \ldots, S_T)$ in the sense of mutual absolute continuity, such that

$$\mathbb{E}_Q(R^{-(t+1)}S_{t+1}|\mathcal{F}_t) = R^{-t}S_t, t = 0, 1, \ldots, T-1. \tag{23.17}$$

That is, the discounted stock prices $M_t = R^{-t}S_t, t = 0, 1 \ldots, T$, comprise a martingale under Q, making the model *arbitrage-free*. Moreover, the market is *complete* in the sense that *any* contingent claim $X = X(S_0, \ldots, S_T)$ offered at the arbitrage-free price $\pi_0 := R^{-T}\mathbb{E}_Q X$ could be perfectly hedged by the market price portfolio associated with a suitable self-financing strategy, i.e., is attainable.

In the remainder of this chapter, let us show a little more generally that (i) the arbitrage-freeness is a consequence of the *existence* of an equivalent probability, which makes the discounted risky asset prices a martingale, and (ii) *completeness* is equivalent to *uniqueness* of such an equivalent probability. Let us assume the same model for bonds $B_t = R^t B_0, t = 1, \ldots, T$, but replace the single risky security of the market price model $S_t = \prod_{j=1}^t Z_j S_0$, $j = 0, 1, \ldots, T$, by a vector $\mathbf{S} = \{\mathbf{S}_t : t = 0, \ldots, T\} := \{(S_t^{(1)}, \ldots, S_t^{(K)}) : t = 0, 1 \ldots, T\}$ consisting of d statistically independent risky securities $S_t^{(j)} := \prod_{i=1}^t Z_i^{(j)} S_0^{(j)}$, $j = 1, \ldots, d$, where $Z_1^{(j)}, Z_2^{(j)}, \ldots, Z_T^{(j)}$ are i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, P)$ taking values in $\{d_j, u_j\}$, $0 < d_j < u_j$ with probabilities $P(Z_t^{(j)} = u_j) = p_j = 1 - P(Z_t^{(j)} = d_j)$. The definition of a contingent claim $X = X(\mathbf{S}_0, \ldots, \mathbf{S}_T)$, namely a non-negative $\mathcal{F}_T$-measurable random variable, remains the same. The other terminology introduced above also directly extends to this setting with the single risky security replaced by the vector of risky securities.

Let $\mathcal{P}_T$ denote the collection of probabilities Q, which are equivalent to P in the sense of mutual absolute continuity and such that the discounted price vector $\{R^{-t}\mathbf{S}_t : t = 0, 1, \ldots, T\}$ is a (vector) martingale under Q with respect to the filtration $\mathcal{F}_t = \sigma(\mathbf{S}_0, \ldots, \mathbf{S}_t)$, $t = 1, \ldots, T$, $\mathcal{F}_0 = \sigma(\mathbf{S}_0) = \{\emptyset, \Omega\}$. That is, each component process $\{S_t^{(j)} : t = 0, 1, \ldots, T\}$ is a martingale under Q. We refer to $Q \in \mathcal{P}_T$ as an *equivalent martingale measure* (EMM).

Proposition 23.3 ***(Existence of EMM).*** The market price model is arbitrage-free over the period $0 \le t \le T$ if and only if $\mathcal{P}_T \neq \emptyset$.

Proof. Assume that there is an equivalent martingale measure Q. The proof is essentially as in the proof given above for the single security binomial tree model since the martingale property of $\{R^{-t}\mathbf{S}_t : t = 0, 1, \ldots, T\}$ implies that any market price portfolio associated with an admissible self-financing strategy $\boldsymbol{\varphi}_t = (\varphi_t^{(1)}, \ldots, \varphi_t^{(d)})$, ψ_t, $t = 0, \ldots, T-1$, namely

$$V_t := \boldsymbol{\varphi}_t \cdot \mathbf{S}_t + \psi_t B_t = \sum_{i=1}^{d} \varphi_t^{(i)} S_t^{(i)} + \psi_t B_t, \quad t = 1, \ldots, T,$$

is, upon being discounted as $R^{-t} V_t$, also a martingale. Specifically, the self-financing and predictability properties of φ, ψ together with the martingale property imply

$$\begin{aligned}
\mathbb{E}_Q(R^{-t} V_t | \mathcal{F}_{t-1}) &= \sum_{i=1}^{d} \varphi_{t-1}^{(i)} \mathbb{E}_Q(R^{-t} S_t^{(i)} | \mathcal{F}_{t-1}) + \psi_{t-1} R^{-t} B_t \\
&= \sum_{i=1}^{d} \varphi_{t-1}^{(i)} R^{-(t-1)} S_{t-1}^{(i)} + \psi_{t-1} B_{t-1} = R^{-(t-1)} V_{t-1},\ t = 1, \ldots, T.
\end{aligned}$$

Now one may see that there can be no arbitrage simply because martingales have constant expected values.

For the converse, suppose that the market model is arbitrage-free. If $0 < d_j < R < u_j$ for each $j = 1, 2, \ldots, d$, then using the result above, one may explicitly solve for a martingale measure for each single security model. The corresponding product measure belongs to $\mathcal{P}_T$. If however R is not contained in (d_{j_0}, u_{j_0}) for some j_0, then one can construct an arbitragable contingent claim based on the value $S_T^{(j_0)}$ of this single security at expiry as follows. In the case $R < d_{j_0} < u_{j_0}$, consider the option to sell the security at its market price $R^T S_0^{(j_0)}$. But observe that since $d_{j_0}^T S_0^{(j_0)} > R^T S_0^{(j_0)}$, $(S_T^{(j_0)} - R^T S_0^{(j_0)})^+ = S_T^{(j_0)} - R^T S_0^{(j_0)}$ regardless of the outcome of the model. This is arbitragable by borrowing $S_0^{(j_0)}/B_0$ units of risk-free bond, i.e., an amount $S_0^{(j_0)}$ of currency, and purchasing one unit of $S_0^{(j_0)}$ at the equivalent amount of currency and, hence, zero initial investment. At time T, one achieves a sure profit of at least $d_{j_0}^T S_0^{(j_0)} - R^T S_0^{(j_0)}$ after paying back the loan. That is, simply by holding these amounts, i.e., $\varphi_t = 1$, $\psi_t = -S_0^{(j_0)}/B_0$ for $t = 0, 1, \ldots, T-1$, one has

$$0 = S_0^{(j_0)} + \left(-\frac{S_0^{(j_0)}}{B_0}\right) B_0 \to V_T \ge d_{j_0}^T S_0^{(j_0)} - R^T S_0^{(j_0)} > 0.$$

Similarly one can construct an arbitrage opportunity in the case that $0 < d_{j_0} < u_{j_0} < R$ (Exercise 4). Thus if the market model is arbitrage-free, then $\mathcal{P}$ is non-empty. ■

Recall that an equivalent martingale measure serves as the *pricing measure*.

Corollary 23.4. If X is an attainable contingent claim, then

$$\pi_0 = R^{-T} \mathbb{E}_Q X \quad \text{for any} \quad Q \in \mathcal{P}_T.$$

Let us now turn to uniqueness of the EMM. Uniqueness of the equivalent martingale measure Q has the following economic implications. Recall that completeness of an arbitrage-free market refers to the attainability of all possible contingent claims.

Proposition 23.5 ***(Uniqueness of EMM).*** An arbitrage-fee market model is complete if and only if $\mathcal{P}_T = \{Q\}$ is a singleton.

Proof. For arbitrage-freeness, we have at the outset that $0 < d_j < R < u_j$ for each j. Suppose that the market is complete. Then for any contingent claim of the form $X = R^T \mathbf{1}_A$, where $A \in \mathcal{F}_T$, it follows from the above corollary that

$$\mathbb{E}_{Q_1}\mathbf{1}_A = \mathbb{E}_{Q_2}\mathbf{1}_A \quad \text{for any} \quad Q_1, Q_2 \in \mathcal{P}_T.$$

Thus $Q_1 = Q_2$. Conversely, suppose $\mathcal{P}_T = \{Q\}$. If X is any contingent claim, then for this market model, there is always an equivalent martingale measure under which X is attainable (Exercise 4). Thus this probability must coincide with Q and X is attainable, i.e., the market model is complete. ■

Exercises

1. (i) (*American Option*) An American option refers to a derivative contract, that permits exercise of the option at any time τ prior to expiry where τ has the (stopping time) property $[\tau \le m] \in \sigma\{S_j : j \le m\}$. (i) Show that for an American call option, there is no advantage to exercise prior to expiry, i.e., an American call is equivalent to a European call, in the sense that $\sup_\tau \mathbb{E}R^{-\tau}V(S_\tau) = \mathbb{E}R^{-T}V(S_T)$, where the supremum is over all stopping times $\tau \le T$. [*Hint*: Use Jensen's inequality.] (ii) Extend this to options with an arbitrary convex payoff function V such that $V(0) = 0$. [*Hint*: Check that $V(\lambda s) \le \lambda V(s)$ for $\lambda \ge 1$.]
   (ii) Show that the pricing measure Q is equivalent to the historic measure P in the sense that for any event $A \in \mathcal{F}_T = \sigma(S_0, S_1, \dots, S_T)$, one has $P(A) = 0$ if and only if $Q(A) = 0$, i.e., the restrictions of P and Q to the σ-field $\mathcal{F}_T = \sigma(S_0, S_1, \dots, S_T)$ are mutually absolutely continuous.
   (iii) Show that the probabilities cannot be mutually absolutely continuous on the (infinite horizon) σ-field $\sigma(S_0, S_1, \dots)$. [*Hint*: Use the law of large numbers to identify events that have positive probability under P and zero probability under Q, and vice verse.]
2. (*Regime Switching Model*) Let $Y_0, Y_1, Y_2, \dots$ be a two-state 0, 1 Markov chain with transition probabilities $P(Y_{n+1} = j | Y_n = i) = p_{ij} \in (0, 1)$, $i, j = 0, 1$, and initial distribution $P(Y_0 = i) = r_i \in (0, 1)$. Imagine Y_n as a state of the market economy, healthy (1) or unhealthy (0), on the n-th period. Assume a risk-free rate $R = 1 + r$ exists for bonds. Suppose that security prices $S_0, S_1, \dots$

evolve as $S_{n+1} = u(Y_n, Y_{n+1})S_n$, where $0 < u(0,0) < R < u(0,1)$ and $0 < u(1,0) < R < u(1,1)$.

(a) Show that such securities together with risk-free bonds provide an arbitrage-free market model and determine the price of a contingent claim X with expiry T. [*Hint*: Consider the role of the historic probability in the proof of Proposition 23.2.]
(b) Show that the market is not complete if $p_{00} \neq p_{10}$.

3. (*Put-Call Parity*) A *(European) put option* is a contract to allow the holder the right to sell the underlying asset at the *expiration time* T for a previously agreed upon (contracted) *strike price* K. The value of the contingent claim at time T to the holder of the contract is its *payoff* $X = (K - S_T)^+$ since the holder will sell S_T at the contracted price K upon the event $[S_T < K]$, but otherwise will not exercise the option. Let π, κ denote the present no-arbitrage prices of European put and call contracts, respectively, for a strike price K at expiry T on an asset with present price S_0.

   (a) Show that for a risk-free interest rate r, one has the put-call parity relation: $\pi = \kappa - S_0 + (1+r)^{-T}K$. Compute $\kappa - \pi = (1+r)^{-T}\mathbb{E}_Q\{(S_T - K)^+ - (K - S_T)^+\}$.
   (b) Give a no-arbitrage cash-flow argument based on the value at expiry of a position of a trader (contract writer and holder) who presently buys the asset at S_0, buys a put option at π, and sells a call option for a price κ. [*Hint*: Argue that the value of the trader's portfolio at expiry is K regardless of which of the two events $[S_T > K]$ and $[S_T \leq K]$ occur. Then discount this value of the portfolio to present value.]

4. Show that the finite market model given in this chapter with $0 < d_j < u_j$, $1 \leq j \leq d$, is arbitrage-free if and only if $d_j < R < u_j$ for each j. Show that if $d = 1$, then it is also complete. Give an example of an arbitrage-free but incomplete finite market model.
5. (*Discrete Time Lognormal Model*) Suppose that the $\{Z_1, Z_2, \ldots\}$ are i.i.d. lognormally distributed. Determine conditions on the parameters such that the discounted price sequence $\{R^{-t}S_t = R^{-t}\prod_{j=1}^{t} Z_j S_0 : t - 0, 1, \ldots T\}$ is a martingale.

Chapter 24
Special Topic: Optimal Stopping Rules

Optimal stopping rules are developed to maximize a reward or minimize a loss in a martingale framework by stopping the process at the right time. Applications include the pricing of American options and the "search for the best" (secretary problem) algorithm.

On a probability space $(\Omega, \mathcal{F}, P)$ an increasing sequence $\{\mathcal{F}_n : n \geq 0\}$ of sub-$\sigma-$fields of $\mathcal{F}$, i.e., a filtration, is given. For example, one may have $\mathcal{F}_n = \sigma\{X_0, X_1, \ldots, X_n\}$ where $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables. Also given are real-valued integrable random variables $\{Y_n : n \geq 0\}$, Y_n being $\mathcal{F}_n$-measurable. The objective is to find $\{\mathcal{F}_n\}_{n=1}^{\infty}$-*stopping times* τ^* that *minimize* $\mathbb{E}Y_\tau$ in the class $\mathcal{T}$ of all a.s. finite $\{\mathcal{F}_n\}_{n=1}^{\infty}$-stopping times τ. One may think of Y_n as the loss incurred by stopping at time n. Similar methods apply to the maximization problem when Y_n represents the gain in stopping at time n.

We begin with the simpler "finite horizon" problem of finding an optimal stopping time τ_m^* that minimizes $\mathbb{E}Y_\tau$ in the class $\mathcal{T}_m$ of all stopping times τ bounded by m. This problem will be solved by a method of *backward recursion*.

We first give a somewhat heuristic derivation of τ_m^*. If $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ $(n \geq 0)$ and $X_0, X_1, \ldots, X_{m-1}$ have been observed, then by stopping at time $m-1$ the loss incurred would be Y_{m-1}. On the other hand, if one decides to continue sampling, then the loss would be Y_m. But Y_m is not known yet, since X_m has not been observed at the time the decision is made to stop or not to stop sampling. The (conditional) expected value of Y_m, given $X_0, \ldots, X_{m-1}$, must then be compared to Y_{m-1}. In other words, τ_m^* is given, on the set $[\tau_m^* \geq m-1]$, by

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_24

$$\tau_m^* = \begin{cases} m-1 \text{ if } Y_{m-1} \le \mathbb{E}(Y_m \mid \mathcal{F}_{m-1}), \\ m \quad\quad \text{if } Y_{m-1} > \mathbb{E}(Y_m \mid \mathcal{F}_{m-1}). \end{cases} \tag{24.1}$$

As a consequence of such a stopping rule one's expected loss, conditionally given $\sigma\{X_0, \ldots, X_{m-1}\}$, is

$$V_{m-1} := \min\{Y_{m-1}, \mathbb{E}(Y_m \mid \mathcal{F}_{m-1})\} \quad \text{on } [\tau_m^* \ge m-1]. \tag{24.2}$$

Similarly, suppose one has already observed $X_0, X_1, \ldots, X_{m-2}$ (so that $\tau_m^* \ge m-2$). Then one should continue sampling only if Y_{m-2} is greater than the conditional expectation (given $\{X_0, \ldots, X_{m-2}\}$) of the loss that would result from continued sampling. That is,

$$\tau_m^* = \begin{cases} = m-2 \text{ if } Y_{m-2} \le \mathbb{E}(V_{m-1} \mid \mathcal{F}_{m-2}), \\ \ge m-1 \text{ if } Y_{m-2} > \mathbb{E}(V_{m-1} \mid \mathcal{F}_{m-2}), \quad \text{on } [\tau_m^* \ge m-2]. \end{cases} \tag{24.3}$$

The conditional expected loss, given $\{X_0, \ldots, X_{m-2}\}$, is then

$$V_{m-2} := \min\{Y_{m-2}, \mathbb{E}(V_{m-1} \mid \mathcal{F}_{m-2})\} \quad \text{on } [\tau_m^* \ge m-2]. \tag{24.4}$$

Proceeding backwards in this manner one finally arrives at

$$\tau_m^* = \begin{cases} = 0 \text{ if } Y_0 \le \mathbb{E}(V_1 \mid \mathcal{F}_0) \\ \ge 1 \text{ if } Y_0 > \mathbb{E}(V_1 \mid \mathcal{F}_0), \quad \text{on } [\tau_m^* \ge 0] \equiv \Omega. \end{cases} \tag{24.5}$$

The conditional expectation of the loss, given $\mathcal{F}_0$, is then

$$V_0 := \min\{Y_0, \mathbb{E}(V_1 \mid \mathcal{F}_0)\}. \tag{24.6}$$

More precisely, V_j are defined by *backward recursion,*

$$V_m := Y_m, \quad V_j := \min\{Y_j, \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)\} \quad (j = m-1, m-2 \ldots, 0), \tag{24.7}$$

and the stopping time τ_m^* is defined by

$$\tau_m^* := \min\{j : 0 \le j \le m,\ Y_j = V_j\}. \tag{24.8}$$

Although the optimality of τ_m^* is intuitively clear, a formal proof of its optimality is worthwhile.

Theorem 24.1. (a) The sequence $\{V_j : 0 \le j \le m\}$ is a $\{\mathcal{F}_j\}$-submartingale. (b) The sequence $\{V_{\tau_m^* \wedge j} : 0 \le j \le m\}$ is a $\{\mathcal{F}_j\}$-martingale. (c) One has

$$\mathbb{E}(Y_\tau \mid \mathcal{F}_0) \ge V_0 \quad \text{a.s.} \ \forall\, \tau \in \mathcal{T}_m,$$

$$\mathbb{E}\left(Y_{\tau_m^*} \mid \mathcal{F}_0\right) = V_0 \quad \text{a.s.} \tag{24.9}$$

In particular, τ_m^* is optimal in the class $\mathcal{T}_m$.

Proof. (a) By (24.7), $V_j \le \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)$.

(b) We need to prove $\mathbb{E}\left(V_{\tau_m^* \wedge (j+1)} \mid \mathcal{F}_j\right) = V_{\tau_m^* \wedge j}$. This requires showing for an arbitrary $\mathcal{F}_j$-measurable bounded real-valued random variable Z that

$$\mathbb{E}\left(Z V_{\tau_m^* \wedge j}\right) = \mathbb{E}\left(Z V_{\tau_m^* \wedge (j+1)}\right) \qquad (0 \le j \le m-1). \tag{24.10}$$

For this write

$$\begin{aligned}\mathbb{E}\left(Z V_{\tau_m^* \wedge j}\right) &= \mathbb{E}\left(Z V_{\tau_m^* \wedge j} \mathbf{1}_{[\tau_m^* \le j]}\right) + \mathbb{E}\left(Z V_{\tau_m^* \wedge j} \mathbf{1}_{[\tau_m^* > j]}\right) \\ &= \mathbb{E}\left(Z V_{\tau_m^* \wedge (j+1)} \mathbf{1}_{[\tau_m^* \le j]}\right) + \mathbb{E}\left(Z V_j \mathbf{1}_{[\tau_m^* > j]}\right).\end{aligned} \tag{24.11}$$

But, on $[\tau_m^* > j]$, $V_j = \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)$. Also, $[\tau_m^* > j] \in \mathcal{F}_j$. Therefore,

$$\begin{aligned}\mathbb{E}\left(Z V_j \mathbf{1}_{[\tau_m^* > j]}\right) &= \mathbb{E}\left[Z \mathbf{1}_{[\tau_m^* > j]} \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)\right] \\ &= \mathbb{E}\left(Z V_{j+1} \mathbf{1}_{[\tau_m^* > j]}\right) = \mathbb{E}\left(Z V_{\tau_m^* \wedge (j+1)} \mathbf{1}_{[\tau_m^* > j]}\right).\end{aligned} \tag{24.12}$$

Using (24.11) in (24.12) one gets (24.10).

(c) Let $\tau \in \mathcal{F}_m$. Since $Y_j \ge V_j$ for all j (see (24.7)), one has $Y_\tau \ge V_\tau$. By Theorem 11.1 and the submartingale property of $\{V_j\}_{j=0}^m$ it now follows that

$$\mathbb{E}(Y_\tau \mid \mathcal{F}_0) \ge \mathbb{E}(V_\tau \mid \mathcal{F}_0) \ge V_0. \tag{24.13}$$

This gives the first relation in (24.9). The second relation in (24.9) follows by the martingale property of $\{Y_{\tau_m^* \wedge j}\}_{j=0}^m$ (and Theorem 11.1). By taking expectations in (24.9) the optimality of τ_m^* is established. ■

Remark 24.1. In the minimization of $\mathbb{E}Y_\tau$ over $\mathcal{T}_m$, Y_n need not be $\mathcal{F}_n$-measurable. In such cases one may replace Y_n by $\mathbb{E}(Y_n \mid \mathcal{F}_n) = U_n$, say, and note that, for every $\{\mathcal{F}_n\}_{n=1}^\infty$-stopping time τ,

$$\mathbb{E}(Y_\tau) = \sum_{j=0}^m \mathbb{E}(Y_j \mathbf{1}_{[\tau=j]}) = \sum_{j=0}^m \mathbb{E}\left[\mathbb{E}(Y_j \mathbf{1}_{[\tau=j]} \mid \mathcal{F}_j)\right] = \sum_{j=0}^m \mathbb{E}[\mathbf{1}_{[\tau=j]} U_j) = \mathbb{E}U_\tau.$$

Hence, the minimization of $\mathbb{E}Y_\tau$ reduces to the minimization of $\mathbb{E}U_\tau$ over $\mathcal{T}_m$.

Also, instead of minimization, one could as easily maximize $\mathbb{E}Y_\tau$ over $\mathcal{T}_m$. Simply replace min by max in (24.7), and replace $\ge$ by $\le$ in (24.9). We state this as a corollary below.

Corollary 24.2. Let $V_j := \max\{Y_j, \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)\}$ $(j = m-1, m-2, \ldots, 1, 0)$, $V_m = Y_m$, $\tau_m^* = \min\{j : 0 \le j \le m, Y_j = V_j\}$. Then:

a. $\{V_j : j = 0, 1, \ldots, m\}$ is a $\{\mathcal{F}_j\}$-supermartingale.
b. $\{V_{\tau_m^* \wedge j} : j = 0, 1, \ldots, m\}$ is a $\{\mathcal{F}_j\}$-martingale.
c. $\mathbb{E}(Y_\tau \mid \mathcal{F}_0) \leq V_0$ a.s. $\forall\, \tau \in \mathcal{T}_m$.
d. $\mathbb{E}(Y_{\tau_m^*} \mid \mathcal{F}_0) = V_0$ a.s. In particular, τ_m^* is optimal in $\mathcal{T}_m$ for the problem of maximizing $\mathbb{E}(Y_\tau)$, and the maximum value is $\mathbb{E}(V_0)$.

Remark 24.2. Let $\mathcal{T}_{j,m}$ denote the set of all stopping times τ such that $j \leq \tau \leq m$ a.s. It follows from the backward recursion, in the construction of the optimal stopping rule in Theorem 24.1, that the optimal stopping rule in the class $\mathcal{T}_{j,m}$ is $\tau_{j,m}^* := \min\{j \leq k \leq m : Y_k = V_k\}$, and the corresponding optimal value is $\mathbb{E}V_j$, and that $V_j = \mathbb{E}[Y_{\tau_{j,m}^*} \mid \mathcal{F}_j]$. Note that V_j maximizes the conditional expectation $\mathbb{E}(Y_\tau \mid \mathcal{F}_j)$ in the class of all τ in $\mathcal{T}_{j,m}$.

Example 1 *(Search for the Best Algorithm (Secretary Problem)).* Suppose there are $m \geq 2$ candidates for a position. Assume their qualifications can be measured on a numerical scale from the worst to the best, with no ties. To be precise, suppose that they can be ranked from worst to best as $1, 2, \ldots, m$. Assume also that the candidates are interviewed one after another and rejected until the j-th candidate is chosen, say $j = 1, \ldots, m$. Let W_j be the probability that j-th person chosen is the best. The problem[1] is to find a stopping rule τ that maximizes the probability W_τ, among all stopping rules τ. Let $\mathcal{F}_j$ denote the σ-field generated by the first j observations $X_1, \ldots, X_j$.

Define the $\mathcal{F}_j$-measurable random variables

$$Y_j := \mathbb{E}(W_j \mid \mathcal{F}_j) = P(X_j = M_j \mid \mathcal{F}_j), \tag{24.14}$$

where $M_j := \max\{X_1, \ldots, X_j\}$, $j = 1, \ldots, m$. If $X_j = M_j$, i.e., X_j, is the top ranked among the first j observations, then the condition probability (given $X_1, \ldots, X_j$) that $X_j = m$ is the top ranked (maximum) overall is j/m; if $X_j < M_j$ then of course this conditional probability is zero. Therefore,

$$Y_j = \frac{j}{m}\mathbf{1}_{\{X_j = M_j\}}. \tag{24.15}$$

Also, for every $\tau \in \mathcal{T}_m$, as explained above,

$$\mathbb{E}Y_\tau = \sum_{j=1}^{m} \mathbb{E}\left(Y_j \mathbf{1}_{\{\tau=j\}}\right) = \sum_{j=1}^{m} \mathbb{E}\left[\mathbb{E}(W_j \mid \mathcal{F}_j)\mathbf{1}_{\{\tau=j\}}\right]$$

[1]The secretary problem has an interesting history recorded by Ferguson (1989), with numerous references. There is also a rather large literature on generalizations to lower order preferences than the best, e.g., search for second best, to partially ordered preferences, and so-on of interest in theories and models for ecological foraging, on-line marketing, and others.

$$= \sum_{j=1}^{m} \mathbb{E}\left[\mathbb{E}(\mathbf{1}_{\{\tau=j\}} W_j \mid \mathcal{F}_j)\right]$$

$$= \sum_{j=1}^{m} \mathbb{E}\left(\mathbf{1}_{\{\tau=j\}} W_j\right) = \mathbb{E}W_\tau.$$

Hence, the maximum of $\mathbb{E}W_\tau$ over $\mathcal{T}_m$ is also the maximum of $\mathbb{E}Y_\tau$ over $\mathcal{T}_m$. In order to use Theorem 24.1 with *min* replaced by *max* in (24.7) and "$\geq$" replaced by "$\leq$" in (24.9), we need to calculate V_j $(1 \leq j \leq m)$. Now $V_m = Y_m$ and (see (24.15))

$$\mathbb{E}(Y_m \mid \mathcal{F}_{m-1}) = \frac{m}{m} P(X_m = M_m \mid \mathcal{F}_{m-1}) = \frac{1}{m} \qquad (M_m = M).$$

Since $(m-1)/m \geq 1/m$, it then follows that

$$V_{m-1} := \max\{Y_{m-1}, \mathbb{E}(Y_m \mid \mathcal{F}_{m-1})\}$$
$$= \frac{m-1}{m} \mathbf{1}_{\{X_{m-1}=M_{m-1}\}} + \frac{1}{m} \mathbf{1}_{\{X_{m-1}<M_{m-1}\}}.$$

Then,

$$\mathbb{E}(V_{m-1} \mid \mathcal{F}_{m-2})$$
$$= \frac{m-1}{m} P\left(X_{m-1} = M_{m-1} \mid \mathcal{F}_{m-2}\right) + \frac{1}{m} P\left(X_{m-1} < M_{m-1} \mid \mathcal{F}_{m-2}\right)$$
$$= \frac{m-1}{m} \frac{1}{m-1} + \frac{1}{m} \frac{m-2}{m-1} = \frac{m-2}{m}\left(\frac{1}{m-2} + \frac{1}{m-1}\right). \tag{24.16}$$

To evaluate

$$V_{m-2} := \max\{Y_{m-2}, \mathbb{E}(V_{m-1} \mid \mathcal{F}_{m-2})\}$$

note that if $(m-2)^{-1} + (m-1)^{-1} \leq 1$ then (see (24.14) and (24.16))

$$V_{m-2} = \frac{m-2}{m} \mathbf{1}_{\{X_{m-2}=M_{m-2}\}} + \frac{m-2}{m}\left(\frac{1}{m-2} + \frac{1}{m-1}\right) \mathbf{1}_{\{X_{m-2}<M_{m-2}\}} \tag{24.17}$$

so that a calculation akin to (24.16) yields

$$\mathbb{E}(V_{m-2} \mid \mathcal{F}_{m-3}) = \frac{m-3}{m}\left(\frac{1}{m-3} + \frac{1}{m-2} + \frac{1}{m-1}\right).$$

Assume, as a backward induction hypothesis, that

$$\mathbb{E}(V_{j+1} \mid \mathcal{F}_j) = \frac{j}{m}\left(\frac{1}{j} + \frac{1}{j+1} + \cdots + \frac{1}{m-1}\right), \tag{24.18}$$

for some j such that

$$a_j := \frac{1}{j} + \frac{1}{j+1} + \cdots + \frac{1}{m-1} \leq 1. \tag{24.19}$$

Then,

$$\begin{aligned} V_j &:= \max\{Y_j, \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)\} = \max\left\{\frac{j}{m}\mathbf{1}_{\{X_j=M_j\}}, \frac{j}{m}a_j\right\} \\ &= \frac{j}{m}\mathbf{1}_{\{X_j=M_j\}} + \frac{j}{m}a_j\mathbf{1}_{\{X_j<M_j\}}, \end{aligned}$$

and, since $P(X_j = M_j \mid \mathcal{F}_{j-1}) = 1/j$, it follows that

$$\mathbb{E}(V_j \mid \mathcal{F}_{j-1}) = \frac{j-1}{m}\left(\frac{1}{j-1} + \frac{1}{j} + \cdots + \frac{1}{m-1}\right) = \frac{j-1}{m}a_{j-1}.$$

The induction is complete, i.e., (24.18) holds for all $j \geq j^*$, where

$$j^* := \max\{j : 1 \leq j \leq m, a_j > 1\}. \tag{24.20}$$

In other words, one gets

$$\mathbb{E}(V_{j+1} \mid \mathcal{F}_j) = \frac{j}{m}a_j \qquad \text{for } \ j^* \leq j \leq m. \tag{24.21}$$

Also,

$$V_{j^*} \equiv \max\left\{Y_{j^*}, \mathbb{E}(V_{j^*+1} \mid \mathcal{F}_{j^*})\right\} = \frac{j^*}{m}a_{j^*}, \tag{24.22}$$

since $a_{j^*} > 1$. In particular, V_{j^*} is *nonrandom*, which implies

$$\mathbb{E}(V_{j^*} \mid \mathcal{F}_{j^*-1}) = \frac{j^*}{m}a_{j^*},$$

which in turn leads to

$$V_{j^*-1} \equiv \max\left\{Y_{j^*-1}, \mathbb{E}(V_{j^*} \mid \mathcal{F}_{j^*-1})\right\} = \frac{j^*}{m}a_{j^*}.$$

Continuing in this manner one get

$$V_j = \frac{j^*}{m} a_{j^*} \qquad \text{for } \ 1 \le j \le j^*. \tag{24.23}$$

The optimal stopping rule is then given by (see (24.8))

$$\tau_m^* = \begin{cases} \min\{j : j \ge j^* + 1, X_j = M_j\}, & \text{if } X_j = M_j \text{ for some } j \ge j^* + 1, \\ m, & \text{if } X_j \ne M_j \text{ for all } j \ge j^* + 1. \end{cases} \tag{24.24}$$

For, if $j \le j^*$, then $Y_j < \mathbb{E}(V_{j+1} \mid \mathcal{F}_j) = (j^*/m)a_{j^*}$. On the other hand, if $j > j^*$, then $a_j \le 1$ so that: (i) $Y_j \ge \mathbb{E}(V_{j+1} \mid \mathcal{F}_j) = (j/m)a_j$ on $\{X_j = M_j\}$ and (ii) $0 = Y_j < \mathbb{E}(V_{j+1} \mid \mathcal{F}_j)$ on $\{X_j < M_j\}$. Simply stated, the optimal stopping rule is to *draw j^* observations and then continue sampling until an observation larger than all the preceding shows up* (and if this does not happen, stop after the last observation has been drawn). The maximal probability of stopping at the maximum value is then

$$\mathbb{E}(V_1) = V_1 = \frac{j^*}{m} a_{j^*} = \frac{j^*}{m} \left(\frac{1}{j^*} + \frac{1}{j^* + 1} + \cdots + \frac{1}{m-1} \right). \tag{24.25}$$

Finally, note that, as $m \to \infty$,

$$a_{j^*} \equiv \frac{1}{j^*} + \frac{1}{j^* + 1} + \cdots + \frac{1}{m-1} \to 1, \tag{24.26}$$

where the difference between the two sides of the relation "$\approx$" goes to zero. This follows since j^* must go to infinity (as the series $\sum_1^\infty (1/j)$ diverges and j^* is defined by (24.20)) and $a_{j^*} > 1$, $a_{j^*+1} \le 1$. Now,

$$a_{j^*} = \frac{1}{m} \sum_{j=j^*}^{m-1} \frac{1}{j/m} \approx \int_{j^*/m}^{1} \frac{1}{x}\, dx = -\log(j^*/m). \tag{24.27}$$

Combining (24.26) and (24.27) one gets

$$-\log \frac{j^*}{m} \approx 1, \qquad \frac{j^*}{m} \sim e^{-1}, \tag{24.28}$$

where the ratio of the two sides of "$\sim$" goes to one, as $m \to \infty$. Thus,

$$\lim_{m \to \infty} \frac{j^*}{m} = e^{-1}, \qquad \lim_{m \to \infty} \mathbb{E}(V_1) = e^{-1}. \tag{24.29}$$

Example 2 ***(****American Options****)****.* An *American option* in mathematical finance refers to contracts that may be exercised at any stopping time prior to the contracted expiry date T. For example, an American call would permit the holder the option to buy the stock at the strike price K at any stopping time $\tau \le T$; see Exercise 6.

Similarly an American put option gives the holder the right to sell the stock at the strike price K at any stopping time prior to T. Recall the binomial tree model for a single security $S_t = \prod_{j=1}^{t} Z_j S_0$ and risk-free bonds $B_t = R^t B_0$, $t = 0, 1, \ldots, T$, where $R = 1 + r$ for the risk-free rate $r \geq 0$ treated in Chapter 23. For an American put option over a time horizon of length T in this market model it is natural to consider the quantity

$$\pi_0 = \max_{\tau \in \mathcal{T}_T} \mathbb{E}_Q(R^{-\tau}(K - S_\tau)^+ | \mathcal{F}_\tau),$$

where Q is an equivalent martingale measure (pricing measure); see Chapter 23. From the perspective of the writer this is the price that would prepare for the worst case exercise scenario by the holder of the put option. The maximum is achieved at the exercise date

$$\tau^* = \min\{t : (K - S_t)^+ = R^{-1}\mathbb{E}_Q(V_{t+1}|\mathcal{F}_t)\},$$

where V_t is recursively defined backwards in time by

$$V_t = \max\{(K - S_t)^+, R^{-1}\mathbb{E}_Q(V_{t+1}|\mathcal{F}_t)\} \quad t = T-1, T-2, \ldots 0$$

and $V_T = (K - S_T)^+$. In other words, one proceeds recursively backwards through the tree comparing the intrinsic value of the option with the discounted expected value (under Q) at each node and recording the larger of the two. The stopping rule is to stop the first time the intrinsic value is larger than the discounted expected value.

To complete this chapter we will consider the *infinite horizon* problem of finding an optimal stopping rule in the class $\mathcal{T}$ and write

$$V_j^{(m)} = V_j \tag{24.30}$$

to emphasize the dependence of V_j on m.

Since $V_m^{(m)} = Y_m$, one has $V_m^{(m+1)} \equiv \min\{Y_m, \mathbb{E}(Y_{m+1} \mid \mathcal{F}_m)\} \leq V_m^{(m)}$. Assuming, as (backward) induction hypothesis, that $V_{j+1}^{(m+1)} \leq V_{j+1}^{(m)}$ for some j $(0 \leq j < m)$, one has

$$V_j^{(m+1)} \equiv \min\left\{Y_j, \mathbb{E}\left(V_{j+1}^{(m+1)} \mid \mathcal{F}_j\right)\right\} \leq \min\left\{Y_j, \mathbb{E}\left(V_{j+1}^{(m)} \mid \mathcal{F}_j\right)\right\} \equiv V_j^{(m)}. \tag{24.31}$$

Hence, $V_j^{(m+1)} \leq V_j^{(m)}$ for $0 \leq j \leq m$, and for each j the sequence $\{V_j^{(m)} : m \geq j\}$ is *decreasing* (as $m \uparrow$). Let $V_j^{(\infty)}$ denote the limit of $V_j^{(m)}$ as $m \to \infty$. Since $V_j^{(m)}$ is $\mathcal{F}_j$-measurable (for every m) so is $V_j^{(\infty)}$. If one assumes

$$\mathbb{E}\left(\sup_{n\geq 0} Y_n^-\right) < \infty, \tag{24.32}$$

then, writing $Z = \sup_{n\geq 1} Y_n^-$, one has

$$-\mathbb{E}(Z \mid \mathcal{F}_j) \leq V_j^{(\infty)} \leq Y_j \quad (j \geq 0). \tag{24.33}$$

The right side of the inequality follows from $V_j^{(m)} \leq Y_j$. For the left side use: (i)$V_m^{(m)} \equiv Y_m \geq -Y_m^- \geq -Z$, so that $V_m^{(m)} \equiv \mathbb{E}(V_m^{(m)} \mid \mathcal{F}_m) \geq -\mathbb{E}(Z \mid \mathcal{F}_m)$ and (ii)backward induction, to get $V_j^{(m)} \geq -\mathbb{E}(Z \mid \mathcal{F}_j)$ (Exercise 1). Thus (24.33) holds for $V_j^{(m)}$ $(m \geq j)$ in place of $V_j^{(\infty)}$. Now take limit as $m \uparrow \infty$. In particular, $V_j^{(\infty)}$ is integrable. It is natural to ask if $\mathbb{E}V_0^{(\infty)}$ is the optimum value for the original problem, since $\mathbb{E}V_0^{(m)}$ is the optimum value for the truncated problem. There are simple examples to show that this is not true without additional assumptions (Exercises 2 and 3). To see what needs to be assumed, notice first that the optimum value is finite; indeed,

$$-\mathbb{E}(Z) \leq \inf_{\tau\in\mathcal{T}} \mathbb{E}Y_\tau \leq \mathbb{E}Y_1. \tag{24.34}$$

Therefore, given $\varepsilon > 0$, there exists $\tau_\varepsilon \in \mathcal{T}$ such that $\mathbb{E}Y_{\tau_\varepsilon} < \inf \mathbb{E}Y_\tau + \varepsilon$. Such a τ_ε is called a *ε-optimal stopping time*. Suppose that, for each $\varepsilon > 0$,

$$\lim_{m\to\infty} \mathbb{E}Y_{\tau_\varepsilon \wedge m} = \mathbb{E}Y_{\tau_\varepsilon}. \tag{24.35}$$

Then by taking m_ε sufficiently large one has

$$\mathbb{E}Y_{\tau_\varepsilon\wedge m_\varepsilon} < \inf_{\tau\in\mathcal{T}} \mathbb{E}Y_\tau + \varepsilon. \tag{24.36}$$

But $\mathbb{E}V_0^{(m_\varepsilon)} \leq \mathbb{E}Y_{\tau_\varepsilon\wedge m_\varepsilon}$. Therefore,

$$\mathbb{E}Y_{\tau^*_{m_\varepsilon}} - \mathbb{E}V_0^{(m_\varepsilon)} < \inf_{\tau\in\mathcal{T}} \mathbb{E}Y_\tau + \varepsilon, \tag{24.37}$$

implying $\mathbb{E}V_0^{(\infty)} < \inf_{\tau\in\mathcal{T}} \mathbb{E}Y_\tau + \varepsilon$, so that

$$\mathbb{E}V_0^{(\infty)} \leq \inf_{\tau\in\mathcal{T}} \mathbb{E}Y_\tau. \tag{24.38}$$

A simple sufficient condition for (24.35) (as well as (24.32)) is that the sequence $\{|Y_n|\}_{n=0}^\infty$ is bounded by an integrable random variable. For some applications this assumption is too stringent, although a direct verification of (24.35) is possible.

Theorem 24.3. Let $\{Y_n : n \geq 0\}$ be an integrable sequence of random variables such that (24.32) holds:

a. Then for every $\varepsilon > 0$ a bounded ε-optimal stopping time exists.
b. Assume (24.35) for a family of ε-optimal stopping times defined for all sufficiently small ε. Then $V_j^{(\infty)} := \lim_{m\to\infty} V_j^{(m)}$ $(j \geq 0)$ satisfy

$$-\mathbb{E}(Z \mid \mathcal{F}_j) \leq V_j^{(\infty)} \leq Y_j, \quad (Z := \sup_{n\geq 0} \mathbb{E}Y_n^-). \tag{24.39}$$

$$V_j^{(\infty)} = \min\left\{Y_j, \mathbb{E}(V_{j+1}^{(\infty)} \mid \mathcal{F}_j)\right\}, \tag{24.40}$$

$$\mathbb{E}V_0^{(\infty)} \leq \mathbb{E}Y_\tau \quad \forall\ \tau \in \mathcal{T}. \tag{24.41}$$

c. If Y_n are non-negative, (24.35) holds, and the stopping time

$$\tau^* := \inf\left\{j \geq 0 : V_j^{(\infty)} = Y_j\right\} \tag{24.42}$$

is a.s. finite, then τ^* is optimal, $\{V_{\tau^*\wedge j}^{(\infty)} : j \geq 0\}$ is a $\{\mathcal{F}_j\}$-martingale, and

$$\mathbb{E}(Y_{\tau^*} \mid \mathcal{F}_0) = V_0^{(\infty)} \text{ a.s.}, \quad \mathbb{E}Y_{\tau^*} = \mathbb{E}V_0^{(\infty)}. \tag{24.43}$$

Proof. We have already proved (a), (b) except for (24.40), which follows on letting $m \to \infty$ in (24.7).

(c) By relations (24.11), (24.12), with τ_m^* replaced by τ^*, it follows that $\{V_{\tau^*\wedge j}^{(\infty)} : j \geq 0\}$ is a $\{\mathcal{F}_j\}_{j=0}^{\infty}$-martingale. Therefore,

$$\mathbb{E}\left(V_{\tau^*\wedge j}^{(\infty)} \mid \mathcal{F}_0\right) = \mathbb{E}\left(V_{\tau^*\wedge 0}^{(\infty)} \mid \mathcal{F}_0\right) = \mathbb{E}\left(V_0^{(\infty)} \mid \mathcal{F}_0\right) = V_0^{(\infty)} \text{ a.s.} \tag{24.44}$$

Since $V_n^{(\infty)} \geq 0$ for all n (in view of the non-negativity of Y_n), one may apply Fatou's Lemma in (24.44) to obtain

$$\mathbb{E}\left(V_{\tau^*}^{(\infty)} \mid \mathcal{F}_0\right) = \mathbb{E}\left(\lim_{j\to\infty} V_{\tau^*\wedge j}^{(\infty)} \mid \mathcal{F}_0\right) \leq \lim_{j\to\infty} \mathbb{E}\left(V_{\tau^*\wedge j}^{(\infty)} \mid \mathcal{F}_0\right) = V_0^{(\infty)} \text{ a.s.} \tag{24.45}$$

Since $V_{\tau^*}^{(\infty)} = Y_{\tau^*}$, it follows that

$$\mathbb{E}(Y_{\tau^*} \mid \mathcal{F}_0) \leq V_0^{(\infty)} \text{ a.s.} \tag{24.46}$$

On the other hand, $\mathbb{E}Y_{\tau^*} \geq \mathbb{E}V_0^{(\infty)}$ by (24.41). Therefore, (24.46) must be an equality a.s. Hence, (24.43) is true. The second relation in (24.43) and the inequality (24.41) prove the optimality of τ^*. ■

Remark 24.3. For the problem of maximization of $\mathbb{E}Y_\tau$ (or, $\mathbb{E}(Y_\tau \mid \mathcal{F}_0)$), (24.32) is to be replaced by $Z := \mathbb{E}(\sup_n Y_n^+) < \infty$. The conclusions of Theorem 24.3 are modified accordingly, with (24.39) replaced by $Y_j \le V_j^{(\infty)} \le \mathbb{E}(Z \mid \mathcal{F}_j)$, in (24.40). "min" is to be replaced by max, and the inequality in (24.41) is reversed. The assumption of non-negativity in part (c) is to be replaced by the assumption that $\{Y_n\}_{n=0}^\infty$ is bounded above by a constant.

Exercises

1. Prove the first inequality in (24.33). [*Hint*: Assume $V_j^{(m)} \ge \mathbb{E}(-Z \mid \mathcal{F}_j)$ for some $j \le m$. Then use $Y_{j-1} \ge \mathbb{E}(-Z \mid \mathcal{F}_{j-1})$ and $\mathbb{E}(V_j^{(m)} \mid \mathcal{F}_{j-1}) \ge \mathbb{E}(-Z \mid \mathcal{F}_{j-1})$ to get $\mathbb{E}(V_{j-1}^{(m)} \mid \mathcal{F}_{j-1}) \ge \mathbb{E}(-Z \mid \mathcal{F}_{j-1})$.]
2. Let X_n $(n \ge 1)$ be i.i.d., $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Let $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$ $(n \ge 1)$. Define $Y_n := \frac{2n}{(n+1)} \prod_{i=1}^n (X_i + 1)$ $(n \ge 1)$.
   (a) Show that there does not exist an optimal $\{\mathcal{F}_n\}_{n=1}^\infty$-stopping time τ^* such that $\mathbb{E}Y_{\tau^*}$ maximizes $\mathbb{E}Y_\tau$ in the class of all $\tau \in \mathcal{T}$. [*Hint*: Use the strong Markov property of $\{X_n\}_{n=1}^\infty$ to prove that, for every $\tau \in \mathcal{T}$, $\mathbb{E}Y_\tau < \mathbb{E}Y_{\tau+1}$.]
   (b) Find an optimal τ_m^* in the class $\mathcal{T}_m$, and calculate $V_j^{(\infty)} := \lim_{m\to\infty} V_j^{(m)}$. [*Hint*: Use Corollary24.2.]
   (c) Show that τ^* defined by (24.42) is ∞ a.s. What other assumptions of the analog of Theorem 24.3(b) are violated? (See Remark 24.3).
   (d) Show that $\sup\{\mathbb{E}Y_\tau : \tau \in \mathcal{T}\} = 2$.
3. Let $\{X_n : n \ge 1\}$ be as in Exercise 2, $Y_n := \min\{1, X_1 + \cdots + X_n\} - \frac{n}{n+1}$, $\mathcal{F}_n := \sigma\{X_1, \dots, X_n\}$ $(n \ge 1)$.
   (a) Show that, for every $\tau \in \mathcal{T}_m$, $\mathbb{E}Y_\tau \le -\frac{1}{2}$. [*Hint*: $\mathbb{E}(Y_\tau) \le \mathbb{E}(X_1 + \cdots + X_\tau) - E\frac{\tau}{\tau+1} = -E\frac{\tau}{\tau+1}$.]
   (b) Let $\tau := \inf\{n \ge 1 : X_1 + \cdots + X_n = 1\}$. Show that $\tau \in \mathcal{T}$ and $\mathbb{E}Y_\tau > 0$.
   (c) Show that $\sup\{\mathbb{E}Y_\tau : \tau \in \mathcal{T}\}$ is not the limit of $\mathbb{E}Y_{\tau_m^+}$ as $m \to \infty$, where τ_m^* is a stopping time that maximizes $\mathbb{E}Y_\tau$ in the class of all τ in $\mathcal{T}_m$.
   (d) Provide a computation of $\mathbb{E}Y_\tau$ in (b).
4. (a) Let $\gamma \to Y(\gamma, \omega)$ be a concave function (on some interval I) for a.s. all $\omega \in \Omega$, and $\mathbb{E}|Y(\gamma_j)| < \infty$ for every $\gamma \in I$. Prove that $\gamma \to \mathbb{E}Y(\gamma_j)$ is concave on I.
   (b) Suppose that $\gamma \to \nu_\theta(\gamma)$, $\theta \in \Theta$, is a family of uniformly bounded concave functions on I. Show that $\gamma \to \nu_*(\gamma) = \inf\{\nu_\theta(\gamma) : \theta \in \Theta\}$ is concave on I.
5. (i) Apply the solution to the secretary problem for the case $m = 10$ to 25 random samples, each of size $m = 10$ independently chosen from a continuous distribution.

(a) Calculate the proportion of times among these 25 in which the highest score is obtained by the optimal stopping rule, and compare that with the theoretical result (probability). [*Hint*: Check that $j^* = 3$.]
(b) Calculate the proportion of times in which one of the two highest scores is obtained [Note that in (a), (b) it does not matter which continuous distribution is sampled because the $m!$ permutations of the orders are equally likely.]

(ii) For the secretary problem show that

$$P(\tau_m^* = j^* + j) = \frac{1}{j^* + j} \sum_{k=0}^{j-1} (-1)^k \binom{j-1}{k} \prod_{i=1}^{k} \frac{1}{j^* + j - i},$$
$$j = 1, \ldots, m - j^* - 1,$$
$$P(\tau_m^* = m) = 1 - \sum_{j=1}^{m-j^*-1} P(\tau_m^* = j^* + j).$$

[*Hint*: Let $A_j = [X_{j^*+j} = M_{j^*+j}]$, $j = 1, \ldots, m - j^*$ and use inclusion/exclusion to check that

$$P(A_j \cap A_{j-1}^c \cap \cdots \cap A_1^c)$$
$$= P(A_j) + \sum_{k=1}^{j-1} (-1)^k \sum_{1 \le i_1 < \cdots < i_k < j} P(A_j \cap A_{i_1} \cap \cdots \cap A_{i_k})$$

and use combinatorial principles for counting permutations.]

(iii) As an alternative, letting $(n)_k = n(n-1)\cdots(n-k+1), 1 \le k \le n$, $(n)_k = 0, k > n$, $(n)_0 = 1$, derive the following formula for the secretary problem:

$$P(\tau_m^* = j^* + j) = \sum_{k=j^*}^{m-1} \frac{(k-1)_{j^*-1}(k-j^*)_{j-1}(m-k)_1(m-j^*-j)!}{m!},$$
$$j = 1, \ldots, m - j^* - 1,$$
$$P(\tau_m^* = m) = P(\tau_m^* = m, M_{j^*} = m-1) + P(\tau_m^* = m, M_{j^*} = m)$$
$$= j^* \frac{(m-2)_{j^*-1}(m-1-j^*)_{m-j^*-1}(1)_1}{m!}$$
$$+ j^* \frac{(m-1)_{j^*-1}(m-j^*)!}{m!}.$$

[*Hint*: Express $P(\tau_m^* = j^* + j) = \sum_{k=j^*}^{m-1} P(\tau_m^* = j^* + j, M_{j^*} = k)$ and compute the terms using combinatorial principles for permutations.]

(iv) Tabulate values for $\frac{1}{m}\mathbb{E}\tau_m^* = \frac{1}{m}\sum_j (j^*+j)P(\tau_m^* = j^*+j)$ for the secretary problem for $m = 5, 6, 7, 8$.

6. (*American vs. European Call Options*) Show that if the payoff function V is convex and $V(0) = 0$, then there is no advantage to early exercise of the option. Show that an American call option is equivalent to a European call option.
7. (*American Put Option*) An *American put option* is a contract that provides the right to sell the stock at any time $\tau \leq T$ for the strike price K, where τ is any stopping time. The final value of such a contract is $(K - S_\tau)^+$. Compute the price and stopping time strategy for an American call option in the binomial tree model with parameters $T = 3, r = .05, K = 50, S_0 = 52, u = 1.07, d = .93, p_u = p_d = .5$.

Chapter 25
Special Topic: A Comprehensive Renewal Theory for General Random Walks

This chapter extends the analysis of renewal processes initiated in Chapter 8. The renewal theorem has a rich history that culminated in a unified approach due to David Blackwell.[1] Among the most important ideas introduced by Blackwell in this context is the notion of *ladder variables*. A comprehensive treatment of this theory is presented here. Example applications include the computation of certain self-similar fractal dimensions arising in iterated function systems in this chapter, and ruin problems in insurance in special topics Chapter 26.

Fix a non-degenerate probability distribution $Q \neq \delta_0$ on $[0, \infty)$ and a measurable function $g : [0, \infty) \to \mathbb{R}$, which is locally bounded, i.e., bounded on finite intervals. The renewal equation is the fixed point equation

$$Tu = u \tag{25.1}$$

for locally bounded measurable functions u, where

$$Tu(x) = g(x) + \int_{[0,x)} u(x-y)Q(dy) = g(x) + (u * Q)(x), \quad x \geq 0. \tag{25.2}$$

Adopting the convention that functions and measures are zero on the negative half-line, the equation may be viewed as an equation on all of $\mathbb{R}$. Formally, repeated iterations of (25.2) yield

[1] Blackwell (1953).

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_25

$$
\begin{aligned}
u &= Tu = g + u * Q = g + (g + (u * Q)) * Q \\
&= g + g * Q + u * Q^{*2} \\
&\vdots \\
&= g + \sum_{m=1}^{n} g * Q^{*m} + u * Q^{*(n+1)} \to g * \sum_{m=0}^{\infty} Q^{*m} = g * U,
\end{aligned} \tag{25.3}
$$

where the m-fold convolution Q^{*m} is defined recursively by $Q^{*0}(dy) = \delta_0(dy)$, $Q^{*(m+1)}(dy) = Q * Q^{*m}(dy)$, $m \geq 0$. So, formally, a solution to (25.1) is given by

$$
u(x) = \int_{[0,x)} g(x-y)U(dy) = (g * U)(x), \tag{25.4}
$$

where $U(dy)$ is the infinite convolution measure, referred to as the *renewal measure*.

$$
U(dy) = \sum_{n=0}^{\infty} Q^{*n}(dy). \tag{25.5}
$$

While this solution is not rigorous, the following proposition provides conditions for when it is valid.

Proposition 25.1. Let $U(dy)$ denote the renewal measure defined by (25.5). Then, $U[0, y) < \infty$ for all $y \geq 0$. Moreover, for locally bounded, measurable g, the function $u(x) = \int_0^x g(x-y)U(dy)$, $x \in \mathbb{R}$, solves (25.1) uniquely among functions bounded on finite intervals.

Proof. Let X_j, $j \geq 1$, be i.i.d. with distribution Q, and $\delta > 0$ such that $\theta = P(X_1 > \delta) > 0$. Let r be the smallest integer greater than or equal to x/δ. Write $N = N[0, x)$ for the number of visits to $[0, x)$ by the random walk $\{S_n = X_1 + \cdots + X_n : n \geq 1\}$. Then $P(N \geq n) \leq P(\#\{1 \leq j \leq n : X_j > \delta\} < r) = \sum_{m=0}^{r-1} \binom{n}{m} \theta^m (1-\theta)^{n-m} \leq rn^r \alpha^n$, $\alpha := \max\{\theta, 1-\theta\}$. Hence $U[0, x) \leq \sum_{n=0}^{\infty} P(N \geq n) < \infty$. To prove that (25.1) is a locally bounded solution, first note that it is clearly locally bounded. Also,

$$
\begin{aligned}
u(x) &= \int_{[0,x]} g(x-y) \sum_{n=0}^{\infty} Q^{*n}(dy) \\
&= g(x) + \int_{[0,x]} g(x-y) \sum_{n=1}^{\infty} Q^{*n}(dy) \\
&= g(x) + \int_{[0,x]} g(x-y) \sum_{n=0}^{\infty} Q^{*n} * Q(dy) \\
&= g(x) + u * Q(x)
\end{aligned}
$$

$$= g(x) + \int_{[0,x]} u(x-y)Q(dy). \tag{25.6}$$

So u satisfies the renewal equation (25.1) as well. For uniqueness, let v be the difference of two solutions. Then $v = Q * v$, and, in particular therefore,

$$v(x) = \int_{[0,x)} v(x-y)Q^{*j}(dy), \quad j = 1, 2, \ldots.$$

Since, $Q \neq \delta_0$, by the law of large numbers, $Q^{*j}[0,x] \to 0$ as $j \to \infty$, for v bounded on $[0,x]$, one has $v(x) \leq ||v||_\infty Q^{*j}[0,x] \to 0$ as $j \to \infty$, i.e., $v(x) = 0$ for all x. ■

Remark 25.1. Another proof of the finiteness of $U[0,x)$ for all x follows from Stein's lemma in Lemma 1 below.

Example 1. For an especially simple example that can be solved directly, suppose that $Q(dx) = \delta_h(dx)$, for a fixed $h > 0$, in the renewal equation (25.1) and (25.2). That is, $u(x) = g(x), 0 \leq x \leq h$, and

$$u(x) = g(x) + u(x-h), x \geq h. \tag{25.7}$$

Then simple iteration yields, inductively, that $u(nh) = \sum_{j=0}^{n} g(jh), n = 1, 2, \ldots.$ More generally, $u(x) = \sum_{j=0}^{[\frac{x}{h}]} g(x - jh), n \geq 0$. In particular, assuming $\sum_{k=0}^{\infty} g(k)$ converges, then

$$\lim_{n\to\infty} u(nh) = \sum_{k=0}^{\infty} g(kh) \equiv \frac{h}{\mu} \sum_{k=0}^{\infty} g(kh),$$

where $\mu = h$ is the mean of Q and h is the lattice spacing. While technically simple, this and related examples occur in the computation of certain fractal dimensions as illustrated in Example 3 at the end of this chapter.

As indicated by the above proof, a probabilistic interpretation of the *renewal measure* $U(dy)$ is obtained as follows. Let $S_n, n \geq 0$, denote the random walk starting at 0 on $\mathbb{R}$ with increments distributed as Q. Then, for $B \in \mathbb{B}$,

$$\begin{aligned} U(B) &= \sum_{n=0}^{\infty} Q^{*n}(B) = \sum_{n=0}^{\infty} P(S_n \in B) \\ &= \sum_{n=0}^{\infty} \mathbb{E}\mathbf{1}_B(S_n) = \mathbb{E}\sum_{n=0}^{\infty} \mathbf{1}_B(S_n), \end{aligned} \tag{25.8}$$

is the *expected number of visits to* B by the random walk.

Definition 25.1. A random variable X is said to have a *lattice distribution* if there are numbers b and $h > 0$ such that $P(X \in \{b + nh : n \in \mathbb{Z}\}) = 1$. The largest such h is the lattice span. If X does not have a lattice distribution, then one says that X has a *non-lattice distribution*. In the case $b = 0$, the lattice distribution is said to be *arithmetic*.

Remark 25.2. The issue of whether a lattice distribution is purely arithmetic or not is not consequential to the renewal theory presented below since $X - b$ is arithmetic without changing h when X is lattice distributed. While the results are stated for arithmetic distributions, they easily extend to the more general lattice distributions by a shift in the increments and, therefore, the mean.

Example 2. While renewal equations occur quite generally in a variety of situations, a simple probabilistic context is obtained as follows. Consider the simple asymmetric random walk $\{Y_n : n \geq 0\}$ on the integers, starting at 0, with ± 1-valued increments such that $P(Y_{n+1} - Y_n = 1) = p > 1/2$. Next let $Q(\{n\}) = P(\tau_0 = n)$, the (defective) distribution of the first return (renewal) time $\tau_0 = \inf\{n \geq 1 : Y_n = 0\}$. Let $g(n) = \delta_0(n), n \geq 0$. Let $S_n, n \geq 0$, denote the random walk on the positive integers starting at zero with increment distribution Q. Then the state S_n of the renewal process is the time of the n-th return to zero by the simple random walk Y, and Q is an arithmetic distribution with lattice span two. One has simply by countable additivity and conditioning that $u(n) = P(Y_n = 0) = \delta_0(n) + \sum_{j=0}^{n} u(n-j)Q(\{j\}), n \geq 0$, solves the renewal equation (25.1) for this choice of g and Q.

Remark 25.3. For a bit of terminology, more generally, a random walk $S_n, n \geq 0$, having non-negative increments with $S_0 = 0$ is referred to as an *ordinary renewal process*. If $S_0 \neq 0$ a.s. is a non-negative random variable, then one says that $S_n, n \geq 0$, is a *delayed renewal process*. In either case, the values S_n are referred to as *renewal times* and the increments $X_n = S_n - S_{n-1}, n \geq 1$, are referred to as the *time between renewals*. The stochastic process recording the number of renewals in time t, $N(t) = \sup\{n : S_n \leq t\}, t \geq 0$, is referred to as the *renewal counting process*. In Chapter 5 it was shown that if the times $X_1, X_2, \ldots$ are i.i.d. exponentially distributed random variables, then the counting process $N(t), t \geq 0$, is a Poisson process.

An important class of renewal processes arise when considering the successive returns (recurrences) of a Markov chain to some specific state. More generally, a key probability problem for renewal theory is the determination of $\lim_{n\to\infty} u(n)$ from the renewal structure embodied by the renewal equation, e.g., see Corollary 25.7.

The history of theoretical developments in this area is one of gradual accomplishments under various hypotheses on the distribution Q. However, the penultimate theorem[2] is David Blackwell's synthesis of these accomplishments that yields a generalization to distributions Q on $\mathbb{R}$ having positive mean.

***Theorem 25.2** (Blackwell's Renewal Theorem).* Let $X_1, X_2, \ldots$ be an i.i.d. sequence of real-valued random variables having common distribution Q, defined on a probability space $(\Omega, \mathcal{F}, P)$ with $\mu = \mathbb{E}X_1 \in (0, \infty]$. Let $S_0 = 0$, $S_n = X_1 + \cdots + X_n, n \geq 1$. For any $y > 0$ let $U[a, a+y) = \sum_{n=0}^{\infty} Q^{*n}[a, a+y)$, $a \in \mathbb{R}$.

1. If Q is non-lattice, then, with the convention that $\frac{y}{\infty} = 0$, $\lim_{a\to\infty} U[a, a+y) = \frac{y}{\mu}$, and $\lim_{a\to-\infty} U[a, a+y) = 0$.
2. If Q is an arithmetic distribution with lattice span h, then, with the convention that $\frac{h}{\infty} = 0$, $\lim_{m\to\infty} U[mh, (m+k)h) = \frac{kh}{\mu}$, $(m, k \in \mathbb{Z}_+)$, and $\lim_{a\to-\infty} U[a, a+y) = 0$.

Let us note that, as observed earlier, $U[a, a+y)$ denotes the expected number of $n \geq 0$ that $a \leq S_n < a+y$. By the law of large numbers $\frac{S_n}{n} \to \mu$ a.s. as $n \to \infty$. In particular, for $\mu > 0$, one has $S_n \to \infty$ almost surely. Therefore, with probability one, for fixed a and $y > 0$, $a \leq S_n < a+y$ for at most finitely many n.

The proof of Theorem 25.2 will be obtained as a result of Blackwell's overall synthesis of earlier special cases involving non-negative random variables that lead to his relaxation to real-valued random variables having a positive mean.

Non-negative lattice case: Noting that $U[mh, (m+k)h) = \mathbb{E}(N_{(m+k)h} - N_{mh})$, this case was treated in Chapter 8. Theorem 25.2(**2**) follows directly from Theorem 8.5.

Non-negative, non-lattice case: Next let us consider the case of non-lattice but non-negative random variables. First let

$$N_k[a, a+y) = \sum_{j=1}^{\infty} \mathbf{1}_{[a,a+y)}(X_{k+1} + \cdots + X_{k+j}), \quad y > 0, a \in \mathbb{R}. \tag{25.9}$$

Note that for constants a, y, the distribution of $N_k[a, a+y)$ does not depend on k. Let $N[a, a+y) = N_0[a, a+y)$, and $U[a, a+y) = \mathbb{E}N_k[a, a+y)$. So $U(T) \equiv U[0, T)$ is the expected number of sums $S_k = \sum_{j=1}^{k} X_j, k \geq 1$, such that $0 \leq S_k < T$. In the case of a random variable Z, the random variable $U[Z, Z+y)$ is defined by the composite function $U[\cdot, \cdot + y) \circ Z$ on Ω.

Lemma 1. $U[a, a+y) < \infty$ for all $a \in \mathbb{R}, y > 0$.

Proof. This follows immediately from Stein's lemma (Corollary 11.5). ■

Lemma 2. Suppose $Z = g(X_1, \ldots, X_k)$ for some measurable function g. Then, for any $A \in \sigma(X_1, \ldots, X_k)$

[2] In recent years a proof by coupling methods has emerged in some generality. The difficulty in establishing finiteness of the coupling time is significantly more non-trivial than in the lattice case treated in Chapter 8. Blackwell's approach using ladder variables nonetheless stands the test of time since ladder variable techniques continue to find new applications.

$$\mathbb{E}\mathbf{1}_A N_k[Z, Z+y) = \mathbb{E}\mathbf{1}_A U[Z, Z+y).$$

Proof. The proof essentially follows by conditioning and the substitution property[3] for conditional expectation.

$$\begin{aligned}\mathbb{E}\mathbf{1}_A N_k[Z, Z+y) &= \mathbb{E}\mathbf{1}_A \mathbb{E}[N_k[Z, Z+y)|\sigma(X_1, \dots, X_k)] \\ &= \mathbb{E}\mathbf{1}_A U[a, a+y)|_{a=Z} \\ &= \mathbb{E}\mathbf{1}_A U[Z, Z+y), \end{aligned} \tag{25.10}$$

by definition of $U[Z, Z+y)$. ■

Lemma 3. For any $a \in \mathbb{R}$, $y > 0$,

$$U[a, a+y) \leq U[0, y) + 1.$$

Proof. Let $A_k = [S_{k-1} < a \leq S_k]$. In view of the non-negativity assumption on the event A_k, k is the first possible sum S_k to be counted in $N[a, a+y)$, and the count $N[a, a+y)$ depends on how many additional sums $X_k, X_k + X_{k+1}, \dots$ may be added so that $S_k + X_{k+1} + \cdots + X_{k+j} < a + y$. Thus, from this and Lemma 2, one has

$$\begin{aligned}\mathbb{E}N[a, a+y) &= \sum_k \mathbb{E}N[a, a+y)\mathbf{1}_{A_k} \\ &\leq \sum_k \mathbb{E}\{N_k[0, y) + 1\}\mathbf{1}_{A_k} \\ &= \sum_k \mathbb{E}\{U[0, y) + 1\}\mathbf{1}_{A_k} \\ &\leq U[0, y) + 1.\end{aligned}$$

■

Lemma 4. Suppose $g_n, n = 1, 2, \dots$ is a sequence of non-negative measurable functions and D is a real number such that $\limsup_n g_n = D$ a.s., and $\lim_n \mathbb{E}g_n = D$. Then $g_n \to D$ a.s. as $n \to \infty$.

Proof. Suppose, for contradiction, $\lim_n g_n$ is not a.s. D. Then $\liminf_n g_n < D$ on a set of positive probability. This means there exist some $\delta > 0$ and a subsequence $g_{n_k} \leq D - \epsilon$ for all k on a set of probability δ, and $g_{n_k} \leq D$ outside of this set. Thus, $\liminf_k \mathbb{E}g_{n_k} < D$, which contradicts $\lim_{\mathbb{E}} g_n = D$. It follows that $\liminf_n g_n \geq D - \epsilon$ a.s. This being true for all $\epsilon > 0$, one must have $\liminf_n g_n = D$ a.s. so that $g_n \to D$ almost surely. ■

[3] See BCPT p. 38.

Theorem 25.3. For any fixed integer $k \geq 1$, $y > 0$, and sequences $\alpha_n, \beta_n, n \geq 1$, such that $\alpha_n \to \infty$, $\beta_n \to \infty$, and

$$D(y) = \limsup_{a\to\infty} U[a, a+y) = \lim_{n\to\infty} U[\alpha_n, \alpha_n + y),$$

$$d(y) = \liminf_{a\to\infty} U[a, a+y) = \lim_{n\to\infty} U[\beta_n, \beta_n + y),$$

one has the following almost sure[4] limits

$$\varphi_n = U[\alpha_n - S_k, \alpha_n - S_k + y) \to D(y),$$

and

$$\psi_n = U[\beta_n - S_k, \beta_n - S_k + y) \to d(y).$$

Proof. Observe that

$$N[a, a+y)\mathbf{1}_{[0,a]}(S_k) = N_k[a - S_k, a - S_k + y)\mathbf{1}_{[0,a]}(S_k),$$

and

$$N[a, a+y)\mathbf{1}_{[a,\infty}(S_k) \leq \{k + N_k[0, y)\}\mathbf{1}_{[a,\infty)}(S_k).$$

Thus,

$$\begin{aligned}
&\mathbb{E}U[a - S_k, a - S_k + y)\mathbf{1}_{[0,a]}(S_k) \\
&\quad \leq U[a, a+y) \\
&\quad \leq \mathbb{E}U[a - S_k, a - S_k + y)\mathbf{1}_{[0,a]}(S_k) + \{k + U[0, y)\}P(S_k \geq a). \quad (25.11)
\end{aligned}$$

Also, by Lemma 3

$$\begin{aligned}
&\mathbb{E}U[a - S_k, a - S_k + y)\mathbf{1}_{[0,a]}(S_k) \geq \mathbb{E}U[a - S_k, a - S_k + y) \\
&\quad - \{1 + U[0, y)\}P(S_k \geq a). \quad (25.12)
\end{aligned}$$

Combining these bounds, one has

$$\begin{aligned}
D(y) &= \lim_n U[\alpha_n, \alpha_n + y) \\
&\leq \liminf_n \big(\mathbb{E}U[\alpha_n - S_k, \alpha_n - S_k + y)\mathbf{1}_{[0,\alpha_n]}(S_k) + \{k + U[0, y)\}P(S_k \geq \alpha_n)\big)
\end{aligned}$$

[4] In Blackwell (1948) this limit is obtained in probability and involves a diagonal subsequence argument to extend to an almost sure limit in a Corollary.

$$\leq \liminf_n \left(\mathbb{E}U[\alpha_n - S_k, \alpha_n - S_k + y) + \{k + U[0, y)\}P(S_k \geq \alpha_n)\right) = \liminf_{n\to\infty} \mathbb{E}\varphi_n,$$

and, similarly,

$$d(y) \geq \limsup_n \mathbb{E}\psi_n.$$

Moreover,

$$\limsup_{n\to\infty} \varphi_n \leq \limsup_{a\to\infty} U[a, a+y) = D(y), \ \liminf_{n\to\infty} \psi_n \geq \liminf_{a\to\infty} U[a, a+y) \geq d(y).$$

In summary,

$$\limsup_n \varphi_n \leq D(y) \leq \liminf_n \mathbb{E}\varphi_n, \quad \limsup_n \mathbb{E}\psi_n \leq d(y) \leq \liminf_n \psi_n,$$

and both sequences are uniformly bounded. For fixed y, apply Fatou's lemma to $U[0, y) + 1 - \varphi_n \geq 0$, to obtain with a little algebra, that $\limsup_n \mathbb{E}\varphi_n \leq \mathbb{E}\limsup_n \varphi_n \leq D(y)$. Observe that if $P(\limsup_n \varphi_n < D(y)) > 0$, then the inequality $\mathbb{E}\limsup_n \varphi_n < D(y)$ is also strict. But then, as noted in summary, $\limsup_n \mathbb{E}\varphi_n \leq \mathbb{E}\limsup \varphi_n < D(y) \leq \liminf_n \mathbb{E}\varphi_n$, a contradiction. Thus, one has $\limsup_n \varphi_n = D(y)$ almost surely, and $\lim_n \mathbb{E}\varphi_n = D(y)$. The assertion follows from Lemma 4. The proof that $\psi_n \to d(y)$ is entirely analogous. ■

Lemma 5. Assume that the distribution of X_1 is non-lattice. Then for every $\epsilon > 0$, there is a number $T > 0$, such that for all $t > T$, there exist numbers c_1, c_2 for which $|t - c_i| < \epsilon, i = 1, 2$, and $\lim_{n\to\infty} U[\alpha_n - c_1, \alpha_n - c_1 + y) = D(y)$, $\lim_{n\to\infty} U[\beta_n - c_2, \beta_n - c_2 + y) = d(y)$.

Proof. Let $V = \{x : P(X_1 \in J) > 0$ for every open interval J containing $x\}$. In view of the hypothesis, V cannot consist of integral multiples of a fixed number. So there is a number T such that every open interval (a, b) of length ϵ with $a > T$ contains points of the form $s = \sum_{i=1}^{\infty} n_i x_i$, $x_i \in V, i \geq 1$, and n_i are non-negative integers[5] all but finitely many of which are zero. For any open interval J containing s, one has $P(S_k \in J) > 0$, where $k = \sum_i n_i$. Now, Lemma 5 implies the existence of $c_1, c_2 \in J$ such that $U[\alpha_n - c_1, \alpha_n - c_1 + y) \to D(y)$, and $U[\beta_n - c_2, \beta_n - c_2 + y) \to d(y)$ as $n \to \infty$. ■

Lemma 6. For all positive numbers a, y, ℓ, one has

$$\frac{P(N[a, a+y) > 0)}{P(X_1 > 0)} \leq U[a, a+y) \leq \frac{P(N[a, a+y) > 0)}{P(X_1 \geq y)} \tag{25.13}$$

[5] This is a number theory fact whose proof may be found in Feller (1971) p. 147, Lemma 2(b). It also follows from Kronecker's theorem; see Hardy and Wright (1938), Chap. III.

and

$$\frac{P(X_1 \geq \ell + y)}{P(X_1 > 0)} \leq P(N[a+y, a+y+\ell) = 0 | N[a, a+y) > 0) \leq \frac{P(X_1 \geq \ell)}{P(X_1 \geq y)}. \tag{25.14}$$

Proof. To simplify notation, write $N = N[a, a+y)$, and $N^* = N[a+y, a+y+\ell)$. Let

$$E_j = [S_{j-1} < a \leq S_j].$$

Then $[N > 0] = \cup_{j=1}^{\infty} E_j$ is a disjoint union. Also $U[a, a+y) = \sum_{k=1}^{\infty} kP(N = k) = \sum_{k=1}^{\infty} P(N \geq k)$, and

$$E_j \cap \cap_{i=1}^{k-1}[X_{j+i} = 0] \subset E_j \cap [N \geq k] \subset E_j \cap \cap_{i=1}^{k-1}[X_{j+i} < y].$$

Thus,

$$P(E_j)P^{k-1}[X_1 = 0] \leq P(E_j \cap [N \geq k]) \leq P(E_j)P^{k-1}(X_1 < y).$$

Sum over j to obtain

$$P(N > 0)P^{k-1}(X_1 = 0) \leq P(N \geq k) \leq P(N > 0)P^{k-1}(X_1 < y),$$

and then sum over k to obtain (25.13). To prove (25.14), fix an integer j and let $I_1 = \inf\{i > j : X_i \geq y\}$, $I_2 = \inf\{i > j : X_i > 0\}$. let $F = [X_{I_1} \geq \ell]$, and $G = [X_{I_2} \geq \ell + y]$. Then

$$E_j \cap G \subset E_j \cap [N^* = 0] \subset E_j \cap F,$$

and

$$\begin{aligned} P(G) &= \sum_{i=j+1}^{\infty} P(X_{j+1} = 0, \ldots, X_{i-1} = 0, X_i > 0, X_i \geq \ell + y) \\ &= \sum_{i=j+1}^{\infty} P^{i-j-1}(X_1 = 0)P(X_1 \geq \ell + y) \\ &= \frac{P(X_1 \geq \ell + y)}{P(X_1 > 0)}. \end{aligned}$$

Similarly,

$$P(F) = \frac{P(X_1 \geq \ell)}{P(X_1 \geq y)}.$$

So $P(E_j)P(G) \leq P(E_j \cap [N^* = 0]) \leq P(E_j)P(F)$. Sum over j to obtain (25.14). ■

We are now prepared to prove Blackwell's Renewal Theorem 25.2 in the case of non-negative, non-lattice displacements X. Recall $U(T) = U[0, T)$.

***Theorem 25.4** (Renewal Theorem for the Non-lattice Case).* Assume that the displacement X is a non-negative and non-lattice random variable. Then, for every $y > 0$,

$$\lim_{T\to\infty}[U(T+y) - U(T)] = \frac{y}{\mathbb{E}X_1},$$

with the convention that $1/\infty = 0$.

Proof. The plan is to show $D(y) \leq \frac{y}{\mathbb{E}X_1} \leq d(y)$. That is to show for all $y > 0$,

$$\tilde{D}(y) \leq \frac{1}{\mathbb{E}X_1} \leq \tilde{d}(y),$$

where $\tilde{D}(y) = D(y)/y$ and $\tilde{d}(y) = d(y)/y$. However, since

$$\sum_{j=0}^{N-1} U[a + \frac{jy}{N}, a + \frac{(j+1)y}{N}) = U[a, a+y),$$

one has $Nd(y/N) \leq d(y)$, and $ND(y/N) \geq D(y)$. Thus

$$\tilde{D}(y/N) \geq \tilde{D}(y), \quad \tilde{d}(y/N) \leq \tilde{d}(y),$$

and it therefore suffices to show that

$$\limsup_{y\to 0} \tilde{D}(y) \leq \frac{1}{\mathbb{E}X_1} \leq \liminf_{y\to 0} \tilde{d}(y). \tag{25.15}$$

For this, for $y > 0$ choose $\epsilon \in (0, y)$, and an arbitrary positive number M. By lemma 5, there are numbers c_i, d_i such that for $i = 0, 1, 2, \ldots, c_i \uparrow, d_i \uparrow$

$$y - \frac{\epsilon}{M} < c_{i+1} - c_i < y < d_{i+1} - d_i,$$

and

$$U[\alpha_n - d_i, \alpha_n - d_i + y) \to D(y), \quad U[\beta_n - c_i, \beta_n - c_i + y) \to d(y).$$

First consider $\tilde{d}(y)$. Write

$$1 = P(X_1 \geq \beta_n - c_0) + \sum_{i=0}^{q} P(N_i > 0, N_j = 0, j < i),$$

where $N_i = N(\beta_n - c_{i+1}, \beta_n - c_{i+1} + c_{i+1} - c_i)$ for $i < q$, and $N_q = N(0, \beta_n - c_q)$, with q chosen so that $c_q < \beta_n \leq c_{q+1}$. That is to say, for each fixed n, some S_k are sure to exceed $\beta_n - c_0$. To see this consider the partition of the non-negative half-line into adjacent intervals with endpoints

$$0 < \beta_n - c_q < \beta_n - c_{q-1} < \cdots < \beta_n - c_{i+1} - \beta_n - c_i < \cdots \beta_n - c_1 < \beta_n - c_0 < \infty.$$

Either $X_1 \geq \beta_n - c_0$ or there is a right-most subinterval of $[0, \beta_n - c_0)$ containing at least one S_k. Now considering this partition and first using (25.14) of Lemma 6, one has upon conditioning that

$$\begin{aligned}
P(N_i > 0, N_j = 0, j < i) &= P(N_j = 0, j < i | N_i > 0) P(N_i > 0) \\
&\leq P(N(\beta_n - c_i, \beta_n - c_i + c_i - c_0) = 0 | N(\beta_n - c_{i+1}, \beta_n - c_i) > 0) \\
&\quad \times P(N(\beta_n - c_{i+1}, \beta_n - c_i) > 0) \\
&\leq \frac{P(X_1 \geq c_i - c_0)}{P(X_1 > c_{i+1} - c_i)} P(N_i > 0),
\end{aligned} \tag{25.16}$$

and, using (25.13) of the same Lemma 6, $P(N_i > 0) \leq P(X_1 > 0) U_i$, where $U_i = \mathbb{E} N_i$. Thus, one has

$$1 \leq P(X_1 \geq \beta_n - c_0) + \sum_{i=0}^{M} \frac{P(X_1 \geq c_i - c_0) P(X_1 > 0) U_i}{P(X_1 > c_{i+1} - c_i)} + \sum_{i=M+1}^{\infty} \frac{P(X_1 \geq c_i - c_0)}{P(X_1 > c_{i+1} - c_i)}. \tag{25.17}$$

Letting $n \to \infty$

$$P(X_1 \geq y) \leq y\tilde{d}(y) P(X_1 > 0) \sum_{i=0}^{M} P(X_1 > iy - \epsilon) + \sum_{i=M+1}^{\infty} P(X_1 > i(y - \epsilon)). \tag{25.18}$$

Recall the elementary inequality for any non-negative random variable Z:

$$\mathbb{E} Z - 1 \leq \sum_{i=0}^{\infty} P(Z > i) - 1 \leq \mathbb{E} Z.$$

So, if $\mathbb{E} X_1$ is finite, then the upper bound in (25.18) converges and one has letting $M \to \infty$,

$$P(X_1 \geq y) \leq y\tilde{d}(y)\left(\frac{\mathbb{E} X_1 + \epsilon}{y} + 1\right), \tag{25.19}$$

and, letting $\epsilon \to 0$,

$$\frac{P(X_1 \geq y)}{P(X_1 > 0)} \leq \tilde{d}(y)\{\mathbb{E}X_1 + y\}, \tag{25.20}$$

so that $1 \leq \mathbb{E}X_1 \limsup_{y \to 0} \tilde{d}(y)$.

A similar argument applies to $\tilde{D}(y)$ as follows. First,

$$1 \geq \sum_{i=0}^{M-1} P(N_i^* > 0, N_j^* = 0, j < i),$$

where $N_i^* = N(\alpha_n - d_{i+1}, \alpha_n - d_{i+1} + d_{i+1} - d_i)$. Conditioning on $[N_i^* > 0]$ and using (25.14) and (25.13) of Lemma 6 in succession as before, one obtains

$$1 \leq \sum_{i=0}^{M-1} \frac{P(X_1 \geq d_{i+1} - d_0)P(X_1 \geq d_{i+1} - d_i)U_i^*}{P(X_1 > 0)}, \tag{25.21}$$

where $U_i^* = \mathbb{E}N_i^*$. Letting $n \to \infty$, with obvious replacements,

$$\frac{P(X_1 > 0)}{P(X_1 > y + \frac{\epsilon}{M})} \geq y\tilde{D}(y) \sum_{i=0}^{M-1} P(X_1 \geq (i+1)y + \epsilon). \tag{25.22}$$

Then,

$$\frac{P(X_1 > 0)}{P(X_1 \geq y)} \geq \tilde{D}(y)\{\mathbb{E}X_1 - y\}, \tag{25.23}$$

and therefore

$$1 \geq \mathbb{E}X_1 \limsup_{y \to 0} \tilde{D}(y). \tag{25.24}$$

If $\mathbb{E}X_1$ is infinite, then note that the lower bound in (25.22) diverges as $M \to \infty$, so that $\tilde{D}(y) = 0$ for all $y > 0$. ■

General Case $0 < \mathbb{E}X_1 \leq \infty$: Finally we drop the condition that the random variables be nonnegative and require only that the mean be positive (possibly infinite). We are now in a position to use these developments for the extension to Theorem 25.2. Blackwell's synthesis begins by consideration of successive new record highs of the random walk starting at zero, the so-called *ladder variables*. Let

$$N_1 = \inf\{n : S_n > 0\}, \quad N_{k+1} = \inf\{n : S_{N_k+n} - S_{N_k} > 0\}, k = 1, 2 \ldots, \tag{25.25}$$

and

$$Z_1 = S_{N_1}, \quad Z_k = S_{N_1+\cdots+N_k} - S_{N_1+\cdots+N_{k-1}}, k = 1, 2, \ldots. \tag{25.26}$$

Then $Z_1, Z_2, \ldots$ is an i.i.d. sequence of nonnegative (possibly defective) random variables with

$$S_{N_1+\cdots+N_k} = Z_1 + \cdots + Z_k, \quad k = 1, 2, \ldots. \tag{25.27}$$

The ladder points $(N_1 + \cdots + N_k, S_{N_1+\cdots+N_k})$, $k \geq 1$, comprise an i.i.d. sequence.

The random variables $N_1, N_1 + N_2, \ldots$ are referred to as *strict ascending ladder times* and the successive values $S_{N_1+\cdots+N_k}$ are the *ladder heights*. The qualifier "strict"refers to the strict inequality defining N_1 and the successive ladder times, while "ascending"depicts the direction of the inequality; for example, the definition of a descending ladder time would involve the inequality $\leq$ in place of $>$.

Let

$$V(t) = \#\{n : T_n := Z_1 + \cdots + Z_n \leq t\}, \quad v(t) = \mathbb{E}V(t), \tag{25.28}$$

and

$$R(t) = \#\{n < N_1 : -t \leq S_n \leq 0\}, \quad r(t) = \mathbb{E}R(t). \tag{25.29}$$

Lemma 7.

$$v(t) < \infty \quad \text{for all } t.$$

Proof. This follows from Stein's lemma (Corollary 11.5). ■

Proposition 25.5 (Wald's Formula). $\mathbb{E}N_1 < \infty$. Moreover, μ and $\mathbb{E}Z_1$ are both finite or both infinite, with

$$\mu\mathbb{E}N_1 = \mathbb{E}Z_1.$$

Proof. Without loss of generality one may assume that the displacements are bounded above, i.e., $P(X_1 \leq M) = 1$ for some number M. Otherwise one may truncate the displacements by $X_n^* = X_n \wedge M$, $n \geq 1$, and note that correspondingly, $N_1 \leq N_1^*$. Also, since $\mu > 0$, for M sufficiently large one has $\mu^* > 0$ as well. So the result for truncated displacements applies to non-truncated displacements as well. Now, for $T_n = Z_1 + \cdots + Z_n$, write

$$\frac{T_k}{k} = \frac{S_{N_1+\cdots+N_k}}{N_1 + \cdots + N_k} \frac{N_1 + \cdots + N_k}{k}, \tag{25.30}$$

and let $k \to \infty$. By the strong law of large numbers one has $\mathbb{E}Z_1 = \mu\mathbb{E}N_1$. Moreover, under the boundedness assumption on displacements, one has $Z_k, k \geq 1$, are bounded as well. In particular, $\mathbb{E}N_1 < \infty$ in this case, and in general (since $N_1 \leq N_1^*$). ■

Lemma 8.

$$r(t) < \infty \quad \text{for all } t \geq 0.$$

Proof. As an application of Wald's formula applied to $N_1 = \inf\{n : S_n > 0\}$, one has $\mathbb{E}N_1 < \infty$. The finiteness of $r(t)$ follows. ■

Proposition 25.6. For $a \in \mathbb{R}$, $y > 0$,

$$U[a, a+y) = \int_0^\infty \{r(t-a) - r(t-a-y)\}v(dt).$$

Proof. Let $\hat{N}_0 = \#\{n : 0 \leq n < N_1, a \leq S_n < a+y\}$,

$$\hat{N}_k = \#\{n : \sum_{j=0}^{k} N_j \leq n < \sum_{j=0}^{k} N_j + N_{k+1},\ a \leq S_n < a+y\}, \quad k \geq 1.$$

Then, recalling $S_{\sum_{j=0}^k N_j} = T_k$, writing $n = \sum_{j=0}^k N_j + m$, $0 \leq m \leq N_{k+1}$, one has on $[T_k = t]$,

$$[a \leq S_n < a+y] = [-(t-a) \leq S_{\sum_{j=1}^k N_j + m} - S_{\sum_{j=1}^k N_j} < -(t-a-y)].$$

Thus,

$$\mathbb{E}\{\hat{N}_k | T_k = t\} = r(t-a) - r(t-a-y), \tag{25.31}$$

so that

$$\mathbb{E}\hat{N}_k = \int_0^\infty \{r(t-a) - r(t-a-y)\}F_k(dt), \tag{25.32}$$

where $F_k(t) = P(T_k \leq t)$. The desired formula is obtained by summing over $k = 0, 1, 2, \ldots$, and using $v(t) = \sum_{k=0}^\infty F_k(t)$. ■

Proof of Part 2 of Blackwell's Renewal Theorem 25.2. Assume an arithmetic distribution with span $h > 0$. Using Proposition 25.6 one has the convolution formula

$$U[nh, (n+1)h) = \sum_{k=0}^\infty \tilde{r}(k-n)\tilde{v}(k) = \sum_{k=0}^\infty \tilde{r}(k)\tilde{v}(k+n), \tag{25.33}$$

where $\tilde{r}(k) = r(kh) - r((k-1)h)$ and $\tilde{v}(k) = v(kh) - v((k-1)h)$. Now,

$$\sum_{k=0}^{\infty} \tilde{r}(k) = \lim_{n\to\infty} r(n) = \mathbb{E}N_1 < \infty. \tag{25.34}$$

Also, $\lim_{n\to\infty} \tilde{v}(n) = \frac{h}{\mathbb{E}X_1}$, and $\lim_{n\to-\infty} \tilde{v}(n) = 0$. It follows, therefore, that

$$\lim_{n\to\infty} U[nh, (n+1)h) = \frac{h\mathbb{E}N_1}{\mathbb{E}Z_1}, \tag{25.35}$$

and

$$\lim_{n\to-\infty} U[nh, (n+1)h) = 0. \tag{25.36}$$

Now, Wald's equation completes the proof. ■

Proof of Part 1 of Blackwell's Renewal Theorem 25.2. Making a change of variable in Proposition 25.6, one has

$$U[a, a+y) = \int_0^\infty \{r(t) - r(t-y)\} v_a(dt), \tag{25.37}$$

where $v_a(\cdot) = v(a + \cdot)$.

For $M > 0$, write

$$U[a, a+y) = I_1(M, a, y) + I_2(M, a, y),$$

where

$$I_1(M, a, y) = \int_0^M \{r(t) - r(t-y)\} v_a(dt),$$

and

$$I_1(M, a, y) = \int_M^\infty \{r(t) - r(t-y)\} v_a(dt).$$

The previous theory in the non-negative, non-lattice case applies to the ladder variables $Z_n, n \geq 1$, to yield for every $y > 0$,

$$\lim_{t\to\infty} \{V(t+y) - V(t)\} = \frac{y}{\mathbb{E}Z_1}.$$

Using monotonicity of $r(t)$, it follows that for fixed M, y,

$$\lim_{a\to\infty} I_1(M,a,y) = \lim_{a\to\infty}[\int_0^M r(t)v_a(dt) - \int_0^{M-y} r(t)v_{a+y}(dt)]$$
$$= \frac{1}{\mathbb{E}Z_1}\int_{M-y}^{M} r(t)dt. \tag{25.38}$$

Similarly, $\lim_{a\to-\infty} I_1(M,a,y) = 0$. Next we show that for each fixed $y > 0$, one has, uniformly in a,

$$\lim_{M\to\infty} I_2(M,a,y) = 0. \tag{25.39}$$

Now,

$$I_2(M,a,y) = \sum_{n=0}^{\infty}\int_{M+ny}^{M+(n+1)y}\{r(t) - r(t-y)\}v_a(dt)$$
$$\le \sum_{n=0}^{\infty} r_1(M,n)\{v(a+M+(n+1)y) - v(a+M+ny)\},$$

where $r_1(M,n) = \sup_{M+ny<t<M+(n+1)y}\{r(t)-r(t-y)\}$. Again using the previous theory for the non-negative, non-lattice case, one has

$$\lim_{b\to\infty}\{V(b+y) - V(b)\} = \frac{y}{\mathbb{E}Z_1}.$$

So, for each y, there is a constant c_y such that for all M, a,

$$I_2(M,a,y) \le c_y\sum_{n=0}^{\infty} r_1(M,n).$$

Now,

$$\sum_{n=0}^{\infty} r_1(M,2n) \le \mathbb{E}N_1 - r(M),$$

and

$$\sum_{n=0}^{\infty} r_1(M,2n+1) \le \mathbb{E}N_1 - r(M),$$

where $r(M) \to \mathbb{E}N_1$ as $M \to \infty$. Therefore, for all a, one has

$$I_2(M, a, y) = |U[a, a + y) - I_1(M, a, y)| < \epsilon(M, y), \tag{25.40}$$

where $\epsilon(M, y) \to 0$ as $M \to \infty$ for fixed y, as claimed.

To see that this is sufficient to complete the proof of the theorem, observe that

$$|U[a, a + y) - \frac{y\mathbb{E}N_1}{\mathbb{E}Z_1}| \leq \epsilon(M, y) + |I_1(M, a, y) - \frac{1}{\mathbb{E}Z_1}\int_{M-y}^{M} r(t)dt|$$
$$+|\frac{1}{\mathbb{E}Z_1}\int_{M-y}^{M} r(t)dt - y\mathbb{E}N_1|. \tag{25.41}$$

Thus,

$$\limsup_{a\to\infty} |U[a, a + y) - \frac{y\mathbb{E}N_1}{\mathbb{E}Z_1}|$$
$$\leq \epsilon(M, y) + |\frac{1}{\mathbb{E}Z_1}\int_{M-y}^{M} r(t)dt - y\mathbb{E}N_1|. \tag{25.42}$$

Now, first letting $M \to \infty$, one then obtains

$$\lim_{a\to\infty} U[a, a + y) = \frac{y\mathbb{E}N_1}{\mathbb{E}Z_1}.$$

Combining this with Wald's theorem yields the theorem in the limit as $a \to \infty$. For the limit $a \to -\infty$, one has for all a

$$U[a, a + y) \leq \epsilon(M, y) + |I_1(M, a, y)|,$$

so that

$$\limsup_{a\to-\infty} U[a, a + y) \leq \epsilon(M, y).$$

Thus, letting $M \to \infty$, one has $\lim_{a\to-\infty} U[a, a + y) = 0$, as asserted. This completes the proof of Theorem 25.2 in the non-lattice case Part 1. ■

We can now consider an application to a more specialized context of renewal equations mentioned at the outset.

***Corollary 25.7** (Key Renewal Theorem: Lattice Case).* Assume that $\sum_{k=0}^{\infty} g(k)$ converges for the renewal equation

$$u(k) = g(k) + \sum_{j=0}^{k} u(k - j)q(j), k \geq 0, \quad u(k) = 0, k < 0,$$

where q is an arithmetic probability distribution on the integers with lattice span h, and mean μ. Then for each $k = 0, 1, 2\ldots,$

$$\lim_{n\to\infty} u(nh+k) = \frac{h}{\mu}\sum_{j=0}^{\infty} g(jh+k).$$

Proof. For simplicity take $h = 1, k = 0$. The general case is left to Exercise 1. In view of Proposition 25.1, one has

$$u(n) = \sum_{j=0}^{n} g(n-j)U(\{j\})) = \sum_{j=0}^{n} g(j)U(\{n-j\})), \quad n = 0, 1, \ldots, \tag{25.43}$$

where $U(\{m\}) = \sum_{n=0}^{\infty} Q^{*n}(\{m\}), m = 0, 1, \ldots$. Using the renewal theorem for U, it follows that for each fixed j, $g(j)U(\{n-j\}) \to g(j)/\mu$ as $n \to \infty$. Moreover, $|g(j)|U(\{n-j\}) \leq |g(j)|$, for each j and $\sum_{j=0}^{\infty}|g(j)| < \infty$. Thus the dominated convergence theorem can be applied to the sum on the right side of (25.43) to complete the proof. ∎

We now prove the key renewal theorem for the non-lattice case under a broad condition,[6] which is needed for an important application in Chapter 26 on ruin problems in insurance. First, a definition is required.

Definition 25.2. For a given function $g : [0,\infty) \to \mathbb{R}$, and $\delta_0 > 0$,let

$$m_n = \inf\{g(x) : (n-1)\delta \leq x \leq n\delta\}, \quad M_n = \sup\{g(x) : (n-1)\delta \leq x \leq n\delta\}$$

for $n = 1, 2, \ldots, 0 < \delta \leq \delta_0$. The function g is said to be directly Riemann integrable if, for all $0 < \delta \leq \delta_0$, (a) $\ell(\delta) = \delta\sum_n m_n$, and $L(\delta) = \delta\sum_n M_n$ converge absolutely, and (b) $L(\delta) - \ell(\delta) \to 0$ as $\delta \to 0$.

Here are five examples of directly Riemann integrable functions which are easy to check (Exercise 2): (1) A continuous function which vanishes outside a finite interval. (2) An indicator function of a finite interval. (3) If F is a distribution function on $[0,\infty)$ with finite mean, then $\overline{F}(x) = 1 - F(x)$ is directly Riemann integrable. (4) If g is directly Riemann integrable on $[0,\infty)$, then so is $x \to g(cx)$ for any $c > 0$. (5) A non-increasing integrable function on $[0,\infty)$.

Corollary 25.8 ***(Key Renewal Theorem: Non-lattice Case).*** Consider the renewal equation (25.1), where g is a directly Riemann integrable function on $[0,\infty)$ and Q is a non-lattice probability on $[0,\infty)$ with finite mean μ. Then the solution u of the renewal equation (25.1), satisfies

---

[6]The condition of direct Riemann integrability is due to Feller (1971), Chapter XI. Also see Ramasubramanian (2009), Chapter 3.

$$\lim_{x\to\infty} u(x) = (1/\mu)\int_{[0,\infty)} g(x)dx. \tag{25.44}$$

Proof. Assume without loss of generality that g is nonnegative (See Exercise 3). Let m and M be lower and upper approximations of g given by $m(x) = m_n$ for $(n-1)\delta \le x < n\delta$, and $M(x) = M_n$ for $(n-1)\delta \le x < n\delta (n = 1, 2, \dots)$. Then for each $x > 0$, writing $U(x)$ for $U[0, x)$ for convenience,

$$\int_{[0,x]} m(x-y)U(dy) \le \int_{[0,x]} g(x-y)U(dy) \le \int_{[0,x]} M(x-y)U(dy). \tag{25.45}$$

By a change of variables $y \to x - y$, one has

$$\int_{[0,x]} m(x-y)U(dy) = \sum_{n\ge 1} m_n[U(x-n\delta) - U(x-(n-1)\delta)]. \tag{25.46}$$

Here it may be noted that $U(y) = 0$ if $y < 0$. Since m is integrable and, for all n, $U(x-n\delta) - U(x-(n-1)\delta) \le C = C(\delta)$ for some constant C. By Blackwell's renewal theorem, one may use Lebesgue's dominated convergence theorem to get

$$\begin{aligned}\lim_{x\to\infty}\int_{[0,x]} m(x-y)U(dy) &= \lim_{x\to\infty}\sum_{n\ge 1} m_n[U(x-(n-1)\delta) - U(x-n\delta)]\\ &= (1/\mu)\sum_{n\ge 1} m_n\delta. \end{aligned}\tag{25.47}$$

Similarly,

$$\lim_{x\to\infty}\int_{[0,x]} M(x-y)U(dy) = (1/\mu)\sum_{n\ge 1} M_n\delta. \tag{25.48}$$

The last two sums are lower and upper Riemann sums for $\int_{[0,\infty)} g(x)dx$. As $\delta \downarrow 0$ these sums converge to the same limit by condition (b) in Definition 25.2. Hence the limit, as $x \to \infty$, of the middle expression in (25.45) equals $\int_{[0,\infty)} g(x)dx/\mu$. ■

Here is another example which will be useful in Chapter 26 on ruin problems in insurance.

Lemma 9. Let $f_1, f_2 : [0,\infty) \to (0,\infty)$ with f_1 non-decreasing and f_2 non-increasing such that $f_1 f_2$ is integrable, and $\lim_{\delta\to 0} c(\delta) = 1$, where $c(\delta) = \sup\{f_1(x+y)/f_1(x) : x \ge 0, 0 \le y \le \delta\}$. Then $g = f_1 f_2$ is directly Riemann integrable.

Proof. Using the notation in Definition 25.2, we have $M_n = \sup\{g(x) : (n-1)\delta \le x \le n\delta\} \le f_1(n\delta)f_2((n-1)\delta) \le c(2\delta)f_1((n-2)\delta)f_2((n-1)\delta))$ for $n = 2, 3, \dots$.

Note that $c(2\delta) \geq f_1((n-2)\delta + 2\delta))/f_1((n-2)\delta))$. Also, $f_1((n-2)\delta)f_2((n-1)\delta)) \leq f_1(x)f_2(x)$ for all x such that $(n-2)\delta \leq x \leq (n-1)\delta$, $n \geq 2$. Thus the upper approximation of the integral of g satisfies

$$\begin{aligned} L(\delta) &= \delta \sum_n M_n \leq \delta M_1 + \delta \sum_{n\geq 2} M_n \\ &\leq \delta \sup\{g(x) : 0 \leq x \leq \delta\} + c(2\delta) \int_{[0,\infty)} g(x)dx. \end{aligned} \tag{25.49}$$

It follows from the inequality involving $c(2\delta)$ in the first sentence of the proof that for all $n \geq 1$,

$$\begin{aligned} m_n &= \inf\{g(x) : (n-1)\delta \leq x \leq n\delta\} \\ &\geq f_1((n-1)\delta)f_2(n\delta) \\ &\geq (1/c(2\delta))f_1(n+1)\delta)f_2(n\delta). \end{aligned} \tag{25.50}$$

Since $f_((n+1)\delta)f_2(n\delta) \geq f_1(x)f_2(x)$ for $n\delta \leq x \leq (n+1)\delta$, it follows that

$$\ell(\delta) = \delta \sum_n m_n \geq (1/c(2\delta))) \int_{[0,\infty)} g(x)dx. \tag{25.51}$$

Letting $\delta \to 0$, it follows that g is directly Riemann integrable. ■

Other aspects of renewal theory can be formulated in terms of the renewal counting process as follows (also see Exercise 5).

***Theorem 25.9** (Elementary Renewal Theorem).* Let $N(t) = \sup\{n : S_n \leq t\}$. Then (i) $\mathbb{E}N(t)/t \to \frac{1}{\mu}$ as $t \to \infty$.

Proof. Apply Wald's formula (Proposition 25.5) to the inequality $t < S_{N(t)+1}$ to get $\liminf_{t\to\infty} \mathbb{E}N(t)/t \geq \frac{1}{\mu}$. For the reverse inequality apply Wald's formula to the truncated inter-renewals $\tilde{X}_j = X_j \wedge a$, to obtain $t \geq \mathbb{E}\tilde{T}_{\tilde{N}(t)} = \mathbb{E}\tilde{T}_{\tilde{N}(t)+1} - \mathbb{E}\tilde{X}_{\tilde{N}(t)+1} = \tilde{\mu}(\mathbb{E}\tilde{N}(t)+1) - \mathbb{E}\tilde{X}_{\tilde{N}(t)+1}$. Use $\mathbb{E}\tilde{N}(t) \geq \mathbb{E}N(t)$ and $\tilde{X} \leq a$ to get $\mathbb{E}N(t)/t \leq \frac{1}{\tilde{\mu}} + \frac{a-\tilde{\mu}}{\tilde{\mu}t}$. Finally, compute the limsup as $t \to \infty$, followed by monotone convergence theorem applied to $\tilde{\mu} = \mathbb{E}X_1 \wedge a$ as $a \to \infty$ to complete the proof. ■

Applications of the renewal theorem to problems in risk theory are provided in Chapter 26. Existence of the so-called *derivative martingale* associated with random cascades and branching random walks introduced in Chapter 21 provide another important area of application for Blackwell's notion of ladder random variables. The following example provides an illustrative application of the renewal theorem to a problem outside of probability theory; see Exercise 10 for another.

Example 3 *(Self-Similar Fractals).* An interesting application[7] of the renewal theorem involves the computation of various notions of fractal dimensions associated with self-similar compact sets, such as the Cantor set, that arise as limit points of iterated similarity contraction mappings. A *similarity contraction* map $S : \mathbb{R}^k \to \mathbb{R}^k$ is function of the form $S = rT$, where $T : \mathbb{R}^k \to \mathbb{R}^k$ is an isometry and $r \in (0, 1)$ a fixed parameter. We consider compact sets K that can be represented as follows. Let $S_1, \ldots, S_m$ be a given family of similarity contractions with respective parameters $r_1, \ldots, r_m$. Also let $C(m) = \{1, \ldots, m\}^\infty$. Then[8]

Lemma 10 *(Hutchinson).* For $\mathbf{j} \equiv (j_1, j_2, \ldots) \in \{1, 2, \ldots, m\}^\infty, x \in \mathbb{R}^k$, $\lim_{n\to\infty} S_{j_1} \cdots S_{j_n} x$ exists and does not depend on x. Moreover

$$K = \cup_{\mathbf{j}\equiv(j_1,j_2,\ldots)\in C(m)}\{\lim_{n\to\infty} S_{j_1} \cdots S_{j_n} x : x \in \mathbb{R}^k\}$$

is a compact subset of $\mathbb{R}^k$, and K is the unique compact set with the similarity invariance property: $K = S_1K \cup S_2K \cup \cdots \cup S_mK$.

Proof. For existence of K one may use a contraction mapping argument as follows. Since $S_1, \ldots, S_m$, are contraction maps and $\mathbb{R}^k$ is a complete metric space, each $S_{j_1\ldots j_m}$ has a unique fixed point, say $s_{j_1\ldots j_m}$. Let us show that for any $\mathbf{j} \in C(m)$, $\lim_{n\to\infty} s_{j_1\ldots j_n} := s_{\mathbf{j}}$ exists, and $K = \{s_{\mathbf{j}} : \mathbf{j} \in C(m)\}$ is a compact invariant set. Let $\lambda = \max_{1\le i,j\le m} ||s_i - s_j||$, and $R = \lambda/(1 - r)$, where $r = \max\{r_i : 1 \le i \le m\}$. Let $B(x, d)$ denote the closed ball of radius d centered at x. Then $\cup_{i=1}^m B(s_i, rR) \subset \cap_{j=1}^m B(s_j, R) := C$, say, since for $||s_i - x|| \le rR$, one has $||s_j - x|| \le \lambda + rR = \lambda + r\lambda/(1 - r) = R$. Note that if $x \in C$, then $||S_i x - s_i|| = ||S_i x - Ss_i|| \le r||x - s_i|| < ||x - s_i|| < R$, so that $S_iC \subset C, 1 \le i \le m$, and hence $C \supset S_{j_1}C \supset S_{j_1j_2}C \supset \cdots \supset S_{j_1\ldots j_n}C \supset \cdots$. That is, writing $C_{j_1} = S_{j_1}C, \ldots, C_{j_1j_2\cdots j_n}C$, one has $C \supset C_{j_1} \supset \cdots \supset C_{j_1\ldots j_n} \ldots$. But the fixed point $s_{j_1\ldots j_n} \in S_{j_1\ldots j_n}C$, and $\mathrm{diam}(S_{j_1\ldots j_n}C) \to 0$ as $n \to \infty$, and each $S_{j_1\ldots j_n}C$ is closed. Thus $\lim_{n\to\infty} s_{j_1\ldots j_n} = s_{\mathbf{j}}$k, say, exists. Now note that if $x \in \mathbb{R}^k, \mathbf{j} \in C(m)$, then $||S_{j_1\cdots j_n}x - s_{j_1j_2\cdots j_n}|| = ||S_{j_1\cdots j_n}x - S_{j_1\cdots j_n}s_{j_1j_2\cdots j_n}|| \le ||x - s_{j_1\cdots j_n}|| \to 0$, and $s_{j_1\cdots j_n} \to s_{\mathbf{j}}$ as $n \to \infty$. Thus the set K in the lemma equals $\{s_{\mathbf{j}} : \mathbf{j} \in C(m)\}$. Letting $K = \{s_{\mathbf{j}} : \mathbf{j} \in C(m)\}$, $S_i(s_{\mathbf{j}}) = s_{i\mathbf{j}}$, since $S_i(s_{\mathbf{j}}) \in S_i(\cap_{n=1}^\infty C_{j_1\ldots j_n}) = \cap_{n=1}^\infty C_{ij_1\ldots j_n} = \{s_{i\mathbf{j}}\}$. Thus $K = \cup_{i=1}^m S_iK$ is invariant.

It remains to prove that K is compact. For this give $C(m) = \{1, \ldots, m\}^\infty$ the product topology for the discrete topology on the factor spaces $\{1, \ldots, m\}$. Then $C(m)$ is compact. Define $\pi : C(m) \to K$ by $\pi(\mathbf{j}) = s_{\mathbf{j}}, \mathbf{j} \in C(m)$. Since $\mathrm{diam}(K)$ is bounded (K being a subset of C) it follows that π is continuous and hence that K is compact.

To prove uniqueness, suppose K is a compact set such that $K = S_1K \cup S_2K \cup \cdots \cup S_mK$. Let $S_{j_1\ldots j_n} = S_{j_1} \circ \cdots \circ S_{j_n}$, and $K_{j_1\ldots j_n} = S_{j_1\ldots j_n}K$. With this notation one has,

[7]Lalley (1988).

[8]Hutchinson (1981).

$$K = \cup_{i=1}^m S_i K = \cup_{i,j=1}^m S_i(S_j K) = \cup_{i,j=1}^m S_{ij} K = \cup_{i,j=1}^m K_{ij}. \tag{25.52}$$

Iterating by induction it follows that

$$K = \cup_{j_1,\ldots,j_n} K_{j_1\ldots j_n}.$$

Similarly, writing $K = \cup_{j_{n+1}=1}^m S_{j_{n+1}} K$ in the next iteration,

$$K_{j_1\ldots j_n} = \cup_{j_{n+1}=1}^m K_{j_1\ldots j_n j_{n+1}}.$$

Thus, $K \supset K_{j_1} \supset K_{j_1 j_2} \supset \cdots \supset K_{j_1 j_2 \ldots j_n} \supset \cdots$. Now, by compactness, $\lim_{n\to\infty} \operatorname{diam}(K_{j_1\ldots j_n}) = 0$, so the set $\cap_{n=1}^\infty K_{j_1\ldots j_n}$ is a singleton by completeness of $\mathbb{R}^k$, say $\{k_{j_1 j_2 \ldots}\}$. ■

Note that uniqueness fails among all similarity invariant sets since $\mathbb{R}^k$ is clearly invariant. Taking $k = 1, m = 2$ and $S_1 x = rx$, $S_2 x = rx + (1-r) = r(x + \frac{1-r}{r})$, $x \in \mathbb{R}$, one obtains generalized Cantor sets K, with $0 < r < 1/2$; the case $r = 1/3$ coincides with the definition of the Cantor set by removal of middle-thirds. The cases $1/2 \le r < 1$, yield $K = [0, 1]$. In any case, K_1, K_2 are, compact pairwise disjoint sets. The *similarity dimension* of K is the unique solution $d > 0$ to $r_1^d + r_2^d = 1$, i.e., $2r^d = 1, d = -\ln 2/\ln r$. For $\epsilon > 0$, a finite subset $F \subset K$ is said to be ϵ-separated if $|x - y| \ge \epsilon$ for all $x, y \in F, x \ne y$. The *packing function* of K is defined by

$$N(\epsilon) = \max\{|F| : F \subset K \text{ is } \epsilon - \text{separated}\}.$$

Proposition 25.10. [9] For each $\beta \in [0, -\ln r)$ there is a constant c_β, uniformly bounded in β, such that as $n \to \infty$,

$$N(e^{n\ln r + \beta}) \sim c_\beta e^{d(n\ln r + \beta)} = c_\beta 2^{-(n + \frac{\beta}{\ln r})},$$

where $d = -\ln 2/\ln r$, for $0 < r \le 1/2$.

Proof. That the similarity dimension $d = -\ln 2/\ln r > 0$ follows from the equation $2r^d = 1$. By compactness of K_1, K_2, there is a $\delta > 0$ such that $|x_1 - x_2| > \delta$for $x_1 \in$

[9]This example was selected for its simplicity, but is part of more involved applications of renewal theory developed by Lalley (1988) that also permits examples in which the images K_i, $1 \le i \le m$ are not disjoint sets.

$K_1, x_2 \in K_2$. Since K_1, K_2 are scaling-similar to K, a maximal ϵ-separated subset of K_i scales to a maximal ϵr^{-1}-separated subset of K. In particular, a maximal ϵ-separated subset of K_i has cardinality $N(\epsilon/r)$. Since $K_1 \cap K_2 = \emptyset$, one has for all $0 < \epsilon < \delta$,

$$N(\epsilon) = 2N(\epsilon/r).$$

More generally, for all $\epsilon > 0$, let

$$N(\epsilon) = 2N(\epsilon/r) + L(\epsilon), \tag{25.53}$$

where, since $\epsilon \to N(\epsilon)$ is a non-increasing integer valued function tending to one for large ϵ, $L(\epsilon) = N(\epsilon) - 2N(\epsilon/r)$ is a piecewise constant function with finitely many discontinuities. Moreover $L(\epsilon) = 0$ for $0 < \epsilon < \delta$. Define $u(x) = e^{-dx}N(e^{-x}), x > 0$. Then, writing $\epsilon = e^{-x}$, multiplying (25.53) by e^{-dx}, writing $\frac{1}{r} = e^{-\ln r}$ and using $2r^d = 1$, one has

$$u(x) = u(x + \ln r) + e^{-dx}L(e^{-x}), \quad x > 0. \tag{25.54}$$

$$u(x) = g(x) + \int_0^x u(x - y)Q(dy), \quad x > 0, \tag{25.55}$$

where $Q(dy) = \delta_{-\ln r}(dy)$ has mean $\mu = -\ln r$, and $g(x) = e^{-dx}L(e^{-x})$ is piecewise continuous with finitely many discontinuities, and compact support in $[0, \infty)$ since $L(x) = 0, x \in (0, \delta)$. In particular, g is directly Riemann integrable. Using a key renewal theorem, or more explicitly as in Example 1 with $0 \le \beta < h = -\ln r$,

$$\begin{aligned} u(nh - \beta) \;&=\; \sum_{j=0}^{[n-\frac{\beta}{h}]} g((n-j)h - \beta) = \sum_{k=0}^{n} g(kh - \beta) \\ &\to c_\beta := \sum_{k=0}^{\infty} g(kh - \beta). \end{aligned} \tag{25.56}$$

For uniform boundedness $c_\beta \le e^{dh}c_0$ note $g(kh - \beta) \le e^{dkh}e^{dh}L(e^{-kh})$. ∎

The Exercise 7 provides a similar illustration for a well-known fractal set K.

Exercises

1. Extend the key renewal theorem to the case of an arithmetic distribution with span $h > 1$, and for $k \geq 1$, and mean $\mu > 0$ by showing that $u(nh + k) \to \frac{h}{\mu}\sum_{j=0}^{\infty} g(jh + k)$ as $n \to \infty$.
2. Check that the five examples following Definition 25.2 are indeed directly Riemann integrable.
3. Show that it is sufficient to prove Corollary 25.8 for *non-negative* directly Riemann integrable g on $[0, \infty)$.
4. (*Waiting Time or Inspection Paradox*) The waiting time paradox, or inspection paradox, refers to the counterintuitive experience of longer average waits for arrivals relative to arbitrarily fixed times; e.g., arrivals of busses at a designated stop. Assume non-lattice renewal times with a positive density f and finite second moment. $\mathbb{E}Y_1^2 = \sigma^2 + \mu^2 < \infty$, $\mu = \mathbb{E}Y_1$.

   (a) Show that the residual life process $R(t) = S_{N(t)+1}, t \geq 0$, is a Markov process with unique invariant probability π having the complementary distribution for f as pdf $\pi(dt) = \frac{1}{\mu}(\int_t^{\infty} f(s))dt \equiv \frac{1}{\mu}F^c(t))dt, t > 0$.
   (b) Show that in steady state (a) $\mathbb{E}R(t) = \frac{1}{2}(\frac{\sigma^2}{\mu^2} + 1)\mu$, where $\frac{\sigma^2}{\mu^2}$ is the squared coefficient of variation[10] of Y_1, and
   (c) $\mathbb{E}R(t) > \mu$ if and only if $\frac{\sigma^2}{\mu^2} > 1$.

5. Assume $\mathbb{E}X_1 = \mu \in (0, \infty)$. For the renewal counting process $N(t) = \sup\{n : S_n \leq t\}$, show that $\frac{N(t)}{t} \to \frac{1}{\mu}$ as $t \to \infty$. [*Hint*: $S_{N(t)} \leq t < S_{N(t)+1}$.]
6. Show that $\varphi_n \to D(y)$ in probability as $n \to \infty$ and $\psi_n \to d(y)$ in probability as $n \to \infty$. [*Hint*: Use Chebyshev arguments along the following lines: Suppose $\varphi_n \not\to D(y)$, then there is an $\epsilon > 0$, such that $0 < \limsup_{n\to\infty} P(D(y) - \varphi_n > \epsilon) \leq \limsup_{n\to\infty} \mathbb{E}(D(y) - \varphi_n)/\epsilon \leq \frac{1}{\epsilon}(D(y) - \liminf_{n\to\infty} \mathbb{E}\varphi_n) \leq 0$.]
7. (*Sierpinski gasket*) Take $k = 2$ and $m = 3$ in Example 3. The Sierpinski gasket is the compact similarity invariant set $K \subset \mathbb{R}^2$ defined by the similarity contraction maps $S_1(x_1, x_2) = (\frac{x_1}{2}, \frac{x_2}{2})$, $S_2(x_1, x_2) = (\frac{1}{2} + \frac{x_1}{2}, \frac{x_2}{2})$, and $S_3(x_1, x_2) = (\frac{1}{4} + \frac{x_1}{2}, \frac{\sqrt{3}}{4} + \frac{x_2}{2})$, for $(x_1, x_2) \in \mathbb{R}^2$. For a geometric realization, iterate the following algorithm for each sub-triangle generated and not removed: Starting with a single solid equilateral triangle, subdivide into four congruent equilateral sub-triangles and remove the middle open triangle, i.e., K is the limit set. Use the renewal theorem to compute $\lim_{n\to\infty} e^{ndh} N(e^{-nh+\beta})$ where the lattice span $h = \ln 2$, $\beta \in (0, h)$, and $d > 0$ is the packing dimension.

[10] The precise form of the mean residual time differs a bit for renewal times having a density from integer renewal times, see Chapter 8, Exercise 12.

Chapter 26
Special Topic: Ruin Problems in Insurance

The ruin problem of insurance is another catalyst for many interesting methods in the historic development of probability. The basic question asks how the probability of eventual ruin of a company depends on the initial capital of the company. In standard models, one seeks to provide an answer in terms of the premium rate and the distribution of claims, be they relatively moderate or possibly catastrophically large. In the latter case, martingale theory proves useful. The first task for the asymptotic analysis of ruin probabilities is therefore to precisely delineate the roles of light- and heavy-tailed claim size distributions. Martingale theory proves useful for this analysis, especially for the light-tailed case. For the general analysis of ruin in the renewal model, also known as the Sparre–Andersen model, Blackwell's ingenious notion of *ladder heights and epochs*, together with his deep general renewal theorem, plays essential roles.

Throughout this chapter, we will identify a distribution function F on the real line $\mathbb{R}$ with its corresponding Lebesgue–Stieltjes measure F on the Borel σ-field of $\mathbb{R}$. That is, the same notation will be used interchangeably as a function and a measure.[1]

Lemma 1. Let F be the distribution function of a random variable $X > 0$, which is not bounded. That is, $F(0) = 0$, $F(x) < 1, \forall x \in (0, \infty)$. Write $\overline{G}(x) = 1 - G(x)$ for a distribution function G on $[0, \infty)$. Then $\liminf_{x\to\infty} \overline{F * F}(x)/\overline{F}(x) \geq 2$.

Proof. Let X_1, X_2 be independent with common distribution F. Then $P(X_1+X_2 > x) \geq P([X_1 > x] \cup [X_2 > x]) = P(X_1 > x) + P(X_2 > x) - P(X_1 > x)P(X_2 >$

[1] See BCPT, p. 228, for Lebesgue–Stieltjes measure.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_26

$x) = 2P(X_1 > x) - P^2(X_1 > x)$. Now divide the relations by $P(X_1 > x)$, and take the liminf as $x \to \infty$. ■

It is intuitively clear that the faster the tail probability of F decays, the larger is the left side in the above lemma. It is easy to check, for example, that if the tail decays exponentially, then one has $\overline{F * F}(x)/\overline{F}(x) \to \infty$ as $x \to \infty$. This motivates the following definition.

Definition 26.1. A distribution (function) F on $\mathbb{R}$, or a real-valued random variable X with distribution function F, is said to be *light-tailed* if for some $\theta > 0$, $\mathbb{E}\exp(hX) < \infty$ for all $h \in (-\theta, \theta)$, and F is *heavy-tailed* otherwise. If F is the distribution of a positive random variable, $F(0) = 0$, $F(x) < 1 \forall x > 0$, then F is *subexponential* if

$$\lim_{x\to\infty} \overline{F * F}(x)/\overline{F}(x) = 2. \tag{26.1}$$

The class of all subexponential distributions is denoted by $\mathcal{S}$.

Definition 26.2. A positive measurable function L on $(0, \infty)$ is said to be *slowly varying at infinity* if

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1, \quad \forall c > 0.$$

A non-negative function f is said to be *regularly varying with index* $\delta \geq 0$ if it is of the form $f(x) = x^{\delta} L(x)$, where L is slowly varying at infinity.

Proposition 26.1 ***(****Karamata Representation*[2]***).*** A positive measurable function L on $(0, \infty)$ is slowly varying at infinity if and only if from some $x_0 > 0$, it can be represented as

$$L(x) = d(x) \exp\{\int_{x_0}^{x} \frac{1}{t} g(t) dt\},$$

where g and d are measurable, $g(t) \to 0$ as $t \to \infty$, and $0 < d \equiv \lim_{x\to\infty} d(x) < \infty$.

Remark 26.1. In view of Lemma 1, it now follows that (26.1) is equivalent to the seemingly stronger property $\limsup_{x\to\infty}[\overline{F * F}(x)/\overline{F}(x)] = 2$.

Lemma 2. Suppose the distribution F of X is subexponential. Then (a) uniformly $\forall y > 0$ lying on a bounded set,

$$\lim_{x\to\infty} \overline{F}(x - y)/\overline{F}(x) = 1. \tag{26.2}$$

[2] For a proof, see Ramasubramanian (2009), pp. 73–76.

(b) The function $L(z) = F(\ln z)$ is slowly varying at infinity. (c) F is heavy-tailed.

Proof. (a) One has $\overline{F * F}(x) = 1 - F^{*2}(x) = 1 - \int_{[0,x]} F(x-t)F(dt) = 1 - \int_{[0,x]}[1 - \overline{F}(x-t)]F(dt) = \overline{F}(x) + \int_{[0,x]} \overline{F}(x-t)F(dt)$, so that, for $0 \leq y \leq x$, noting that $\overline{F}$ is a decreasing function, $\overline{F * F}(x)/\overline{F}(x) = 1 + \int_{[0,x]} \overline{F}(x-t)/\overline{F}(x)F(dt) = 1 + \int_{[0,y]} \overline{F}(x-t)/\overline{F}(x)F(dt) + \int_{[y,x]} \overline{F}(x-t)/\overline{F}(x)F(dt) \geq 1 + F(y) + [\overline{F}(x-y)/\overline{F}(x)](F(x) - F(y))$. Let $y > 0$ be fixed. Then letting $x \to \infty$ on both sides, one obtains $2 \geq 1 + F(y) + \limsup_{x\to\infty}\{\overline{F}(x-y)/\overline{F}(x)\}\{1 - F(y)\}$, or $\limsup_{x\to\infty}[\overline{F}(x-y)/\overline{F}(x)] \leq 1$. But, obviously, $\liminf_{x\to\infty} \overline{F}(x-y)/\overline{F}(x) \geq 1$. The uniformity of the inequalities for y in a bounded set is also immediate. Next consider (b). For $0 < a < 1$, write $y = -\ln(a), w = az, c = 1/a$. Then, for $0 < a < 1$, as $z \to \infty$,

$$L(az)/L(z) = F(\ln z - y)/F(\ln z) \to 1, \tag{26.3}$$

and, for $c > 1$, as $w \to \infty$,

$$L(cw)/L(w)(= L(z)/L(az)) \to 1, \tag{26.4}$$

establishing the slowly varying property of L. Finally consider (c). It follows from (b), writing $Z = \exp\{X\}$, that for every $h > 0$,

$$\begin{aligned} \mathbb{E}\exp\{hX\} &= \mathbb{E}Z^h \\ &\geq \mathbb{E}(Z^h \mathbf{1}_{[Z>z]}) \\ &\geq z^h P(Z > z) = z^h L(z) \to \infty, \end{aligned} \tag{26.5}$$

as $z \to \infty$, by a well-known property of slowly varying functions (see Exercise 1(b)). ■

The following is an important characterization of subexponentiality of F.

Proposition 26.2. $F \in \mathcal{S}$ if and only if

$$\lim_{x\to\infty} \overline{F^{*n}}(x)/\overline{F}(x) = n \quad \forall\, n = 2, 3, \ldots \tag{26.6}$$

The proof of this proposition follows from induction and the following useful lemma.[3]

[3] We follow Ramasubramanian (2009), pp. 87–89, 92–98, for several results that follow.

Lemma 3. Suppose $F \in \mathcal{S}$ and G is a distribution on $(0, \infty)$. If, for $0 < a < \infty$,

$$\lim_{x\to\infty} \overline{G}(x)/\overline{F}(x) = a, \tag{26.7}$$

then

$$\lim_{x\to\infty} \overline{F * G}(x)/\overline{F}(x) = 1 + a. \tag{26.8}$$

Proof. From the first few lines of the proof of Lemma 2 (a), one gets

$$\overline{F * G}(x)/\overline{F}(x) = 1 + (1/\overline{F}(x)) \int_{[0,x]} \overline{G}(x-y)F(dy). \tag{26.9}$$

Therefore, one needs to prove

$$\lim_{x\to\infty} (1/\overline{F}(x)) \int_{(0,x]} \overline{G}(x-y)F(dy) = a. \tag{26.10}$$

Given $\epsilon > 0$, let $\tilde{x} \equiv \tilde{x}(\epsilon)$ be such that $\overline{G}(u) \le (a+\epsilon)\overline{F}(u)\ \forall\ u \ge \tilde{x}$. This is possible due to (26.7). Then for $x > \tilde{x}$,

$$\begin{aligned}
&(1/\overline{F}(x)) \int_{[0,x]} \overline{G}(x-y)F(dy) \\
&= (1/\overline{F}(x)) \int_{(0,x-\tilde{x}]} \overline{G}(x-y)F(dy) + (1/\overline{F}(x)) \int_{[x-\tilde{x},x]} \overline{G}(x-y)F(dy) \\
&\le (a+\epsilon) \int_{(0,x-\tilde{x}]} (1/\overline{F}(x))\overline{F}(x-y)F(dy) + (1/\overline{F}(x))[F(x) - F(x-\tilde{x})] \\
&\le (a+\epsilon) \int_{[0,x]} (1/\overline{F}(x))\overline{F}(x-y)F(dy) \\
&\quad +(1/\overline{F}(x))[\overline{F}(x-\tilde{x}) - \overline{F}(x)].
\end{aligned} \tag{26.11}$$

Again, from the proof of Lemma 2, the last integral equals $\overline{F * F}(x)/\overline{F}(x) - 1$, which converges to 1 since $F \in \mathcal{S}$. Also, the second term in (26.11) converges to zero as $x \to \infty$, by Lemma 2. Hence

$$\limsup_{x\to\infty} (1/\overline{F}(x)) \int_{[0,x]} \overline{G}(x-y)F(dy) \le (a+\epsilon). \tag{26.12}$$

Next let x' be such that, given $\epsilon > 0$, $\overline{G}(u) \ge (a-\epsilon)\overline{F}(u) \forall u \ge x$. Then for $x > x'$ (see (26.11)),

$$(1/\overline{F}(x)) \int_{[0,x]} \overline{G}(x-y)F(dy) \ge (a-\epsilon)\frac{1}{\overline{F}(x)} \int_{(0,x-x']} \overline{F}(x-y)F(dy) \ge (a-\epsilon)F(x-x'), \tag{26.13}$$

using the fact that $(1/\overline{F}(x))\overline{F}(x-y) \geq 1$ for $y \geq 0$. Thus

$$\liminf_{x\to\infty}(1/\overline{F}(x))\int_{(0,x]}\overline{G}(x-y)F(dy) \geq a. \tag{26.14}$$

Now (26.10) follows from (26.12) and (26.14). ■

There is an alternative, but equivalent, definition of subexponentiality that is probably more intuitively appealing, as stated below.

Proposition 26.3. Let F be a distribution function on $(0, \infty)$.

$$F \in \mathcal{S} \iff P(S_n > x) = P(M_n > x)(1 + o(1)) \text{ as } x \to \infty, \forall n = 2, 3, \ldots, \tag{26.15}$$

where $S_n = X_1 + \cdots + X_n$, and $M_n = \max\{X_1, \ldots, X_n\}$, with $X_n, n \geq 1$, an i.i.d. sequence with common distribution F.

Proof. To see this note that for every $n = 2, 3, \ldots$, $P(S_n > x) = \overline{F^{*n}}(x)$, and

$$\begin{aligned} P(M_n > x) &= P(\cup_{j=1}^n [X_j > x]) \\ &= n\overline{F}(x) - (1/2)n(n-1)(\overline{F}^2(x)) + \cdots + (-1)^{n+1}\overline{F}^n(x) \\ &= n\overline{F}(x)(1 + o(1)) \text{ as } x \to \infty. \end{aligned} \tag{26.16}$$

Hence

$$\overline{F^{*n}}(x)/\overline{F}(x) = n\overline{F^{*n}}(x)/n\overline{F}(x) = n\{P(S_n > x)/P(M_n > x)\}(1 + o(1)) \to n$$

as $x \to \infty$. By Proposition 26.2, $F \in \mathcal{S}$ the converse is obtained by reversing these equalities. ■

The following result is extremely useful in proving integrability and interchange of summation and limits in a host of problems involving subexponential distributions.

Lemma 4 ***(Kesten's Lemma).*** If $F \in \mathcal{S}$, then for every $\epsilon > 0$, there exists a constant $K, 0 < K < \infty$, depending on ϵ, such that

$$\overline{F^{*n}}(x)/\overline{F}(x) \leq K(1 + \epsilon)^n, \forall x > 0, \forall n = 2, 3, \ldots \tag{26.17}$$

Proof. Let $X_1, X_2, \ldots$ be an i.i.d. sequence with a common distribution function F. For $S_n = \sum_{j=1}^n X_j, x_0 > 0, n \geq 1$, let

$$\alpha_n = \sup_{x > x_0} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)}.$$

Using subexponentiality, for any $\epsilon > 0$, one may choose x_0 such that for $x > x_0$,

$$P(S_2 > x, X_2 \le x) = \overline{F^{*2}}(x) - \overline{F}(x) \le (1 + \frac{\epsilon}{2})\overline{F}(x).$$

Using the definitions of α_n and x_0, one has, for $x > x_0$,

$$\begin{aligned} P(S_n > x, X_n \le x - x_0) &= \int_0^{x-x_0} P(S_{n-1} > x - y)F(dy) \\ &\le \alpha_{n-1} \int_0^{x-x_0} \overline{F}(x-y)F(dy) \\ &= \alpha_{n-1} P(S_2 > x, X_2 \le x - x_0) \\ &\le \alpha_{n-1}(1 + \frac{\epsilon}{2})\overline{F}(x). \end{aligned} \tag{26.18}$$

In addition,

$$P(S_n > x, X_n > x - x_0) \le \overline{F}(x - x_0) \le c\overline{F}(x), \tag{26.19}$$

for some $c > 0$. Here, using Lemma 2(a),

$$c = \sup_x \frac{\overline{F}(x - x_0)}{\overline{F}(x)} < \infty. \tag{26.20}$$

In view of (26.18) and (26.19), one has

$$\alpha_n \le \alpha_{n-1}(1 + \frac{\epsilon}{2}) + c, \quad n = 2, 3, \ldots$$

Thus, by induction, one has

$$\alpha_n \le \alpha_1(1 + \frac{\epsilon}{2})^{n-1} + c\sum_{j=0}^{n-2}(1 + \frac{\epsilon}{2})^j \le \alpha_1(1 + \frac{\epsilon}{2})^{n-1} + c[(1 + \epsilon/2)^{n-1} - 1]/\epsilon/2,$$

so that by definition of α_n, there exists a constant K_1 such that

$$\sup_{x > x_0} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \le K_1(1 + \frac{\epsilon}{2})^{n-1} \le K_1(1 + \epsilon)^n.$$

Also, since $\overline{F}(x)$ is decreasing on $[0, x_0]$ from $\overline{F}(0) = 1$, taking $K_2 = \frac{1}{\overline{F}(x_0)}$, one has

$$\sup_{x \le x_0} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \le K_2 \le K_2(1+\epsilon)^n.$$

The assertion follows by taking $K = K_1 \vee K_2$. ■

Definition 26.3. Given a distribution F on $(0, \infty)$ with a finite mean μ, the *integrated tail distribution* F_I is defined to be the distribution on $(0, \infty)$ having the density

$$f_I(x) = (1/\mu)\overline{F}(x), \; 0 < x < \infty. \tag{26.21}$$

The class of all distributions F such that $F_I \in \mathcal{S}$ is denoted by $\mathcal{S}_I$.

Recall that, using Fubini's theorem, one may check that $\int_{[0,\infty)} \overline{F}(t)dt = \mu$, so that (26.21) defines a probability density on $(0, \infty)$.

Remark 26.2. The precise asymptotic rate of the ruin probability is known generally only for claim size distributions F such that $F \in \mathcal{S}_I$, i.e., for F satisfying

$$\lim_{x\to\infty} [\overline{F_I * F_I}(x)/\overline{F_I}(x)] = 2. \tag{26.22}$$

A sufficient condition for this is the following[4]:

$$\lim_{x\to\infty} \int_{(0,x]} [\overline{F}(x-y)/\overline{F}(x)]\overline{F}(y)dy = 2\mu. \tag{26.23}$$

Definition 26.4. The class of all F on $(0, \infty)$ with a finite mean μ and satisfying (26.23) is denoted by $\mathcal{S}^*$.

Lemma 5. If $F \in \mathcal{S}^*$, then $F \in \mathcal{S}$ and and $F \in \mathcal{S}_I$ (i.e., $F \in \mathcal{S}$ and $F_I \in \mathcal{S}$).[5]

Definition 26.5. Let F be an absolutely continuous distribution on $(0, \infty)$ with density f. The *hazard function* q is defined by $q(x) = f(x)/\int_{(x,\infty)} f(t)dt (0 < x < \infty) = f(x)/\overline{F}(x)$. One says that q is *eventually decreasing* if q is decreasing on (a, ∞) for some $a > 0$.

Note that $q(x)$ is the density at x of the conditional distribution (of a random variable X with distribution F), given the event $[X > x]$. The following result is often useful in checking if $F \in \mathcal{S}^*$.

[4]See Ramasubramanian (2009), pp. 92, 93

[5]The proof may be found in Ramasubramanian (2009), pp. 96–98.

Lemma 6. Let F be absolutely continuous on $(0, \infty)$ with hazard function q. (a) If $\lim_{x\to\infty} xq(x) < \infty$, then $F \in \mathcal{S}^*$. (b) If (i) q is eventually decreasing to zero and (ii) $\int_{(0,\infty)} \exp\{xq(x)\}\overline{F}(x)dx < \infty$, then $F \in \mathcal{S}^*$.[6]

The following are simple examples of F belonging to the various classes introduced above. One can check the assertions by direct computation and/or using Lemma 6.

Example 1. Let F be exponential with density $f(x) = \lambda \exp\{-\lambda x\}\mathbf{1}_{(0,\infty)}(x)$. Then F is light-tailed.

Example 2. Let F have density $f(x) = cx^\delta \exp\{-\alpha x^\beta\}\mathbf{1}_{(0,\infty)}(x)$ for some $\alpha > 0$, $\beta > 0$, and $\delta \geq 0$. (a) If $\beta \geq 1$, then F is light-tailed. (b) If $0 < \beta < 1$, then $F \in \mathcal{S}^*$ (and, therefore, $F \in \mathcal{S}$ and $F \in \mathcal{S}_I$).

Example 3. If F has the density $f(x) = \delta(x^2 \ln^2 x)^{-1}\mathbf{1}_{(c,\infty)}(x)$ for some $\delta > 0$ and $c > 1$, then $F \in \mathcal{S}^*$.

Example 4. If F has the density $f(x) = \delta(x \ln^2 x)^{-1}\mathbf{1}(c, \infty)(x)$ for some $\delta > 0$ and $c > 1$, then the mean of F is infinite, so that F is heavy-tailed and does not belong to $\mathcal{S}_I$.

Let us now consider the *Sparre–Andersen model* of insurance. In this model, also known as the *general renewal model*, claims of (strictly positive) sizes $X_1, X_2, \dots$ arrive at random times $T_1 \leq T_2 \leq \cdots$, and a constant premium $c > 0$ per unit time is collected. The sequences $\{X_n : n \geq 1\}$ and $\{T_n : n \geq 1\}$ are assumed to be independent. It is also assumed that the inter-arrival times $A_i = T_i - T_{i-1} (i = 1, 2, \dots)$ are i.i.d., with $T_0 = 0$, and $\mathbb{E}A_i = \mu < \infty$. For an insurance company with an initial capital $u > 0$, the probability of ruin is defined by

$$\begin{aligned}\psi(u) &= P(\sum_{i=1}^n X_i > u + \sum_{i=1}^n cA_i, \text{ for some } n) \\ &= P(\sum_{i=1}^n Z_i > u \text{ for some } n). \qquad (26.24)\end{aligned}$$

Here we use the notation $Z_i = X_i - cA_i$. The common distribution of the i.i.d. sequence $\{Z_i : i \geq 1\}$ is assumed to satisfy the *Net Profit Condition* (NPC) defined by

$$\text{(NPC)} \qquad \mathbb{E}Z_i < 0. \qquad (26.25)$$

Note that if $\mathbb{E}Z_i$ is finite, and $\mathbb{E}Z_i \geq 0$, then by the strong law of large numbers, $\psi(u) = 1$ for all $u > 0$. To avoid the trivial case $\psi(u) = 0 \forall u > 0$, also assume

$$P(Z_i > 0) > 0. \qquad (26.26)$$

[6]The proof may be found in Ramasubramanian (2009), pp. 99–100.

A special case of the Sparre–Andersen model, in the case that the inter-arrival times $A_i, i \geq 1$, are i.i.d. exponential, is called the *Cramér–Lundberg model* . Both cases were considered in Chapter 11. The *light-tail case* is defined by assuming that the claim size distribution is light-tailed,

$$\mathbb{E}\exp\{qX_i\} < \infty \text{ for some } q > 0. \tag{26.27}$$

This implies that there exists $h, 0 < h \leq \infty$ such that

$$1 \leq m(q) := \mathbb{E}\exp\{qZ_i\} < \infty \text{ for } 0 \leq q < h, \quad \lim_{q \to h} m(q) = \infty. \tag{26.28}$$

Let us recall that under these assumptions for the general Sparre–Andersen model, it follows from martingale theory that one has the following Lundberg bound, see Proposition 11.8. There exists a unique $q = R > 0$ such that

$$m(q) = 1, \tag{26.29}$$

and

$$\psi(u) \leq \exp\{-Ru\}, \quad u > 0. \tag{26.30}$$

The parameter R is often referred to as the *Lundberg constant*. It will be seen from the derivation of the precise asymptotics of the ruin probability that the exponential rate provided by (26.30) cannot in general be improved upon. The true asymptotic rate will be shown to be given by $\psi(u) \sim d\exp\{-Ru\}$ for some constant $d \leq 1$, where R is the Lundberg constant; here, $\sim$ denotes that the ratio of its two sides converges to 1 as $u \to \infty$ (see Remark 26.4).

The following provides a basic result for the Cramér–Lundberg model.

Theorem 26.4 (Pollaczek–Khinchine Formula). Consider the Cramér–Lundberg model where A_1 is exponential with parameter λ, i.e., mean $1/\lambda$. Assume the NPC condition (26.25) and the condition (26.26). Then (a) one has the following renewal equation type relation satisfied by the survival probability $\phi(u) = 1-\psi(u)$, namely,

$$\phi(u) = \rho/(1+\rho) + 1/(1+\rho)\int_{(0,u]} \phi(u-x)F_{X,I}(dx), \tag{26.31}$$

and (b) its unique bounded solution is given by

$$\phi(u) = \rho/(1-\rho)\sum_{n=0}^{\infty}(1+\rho)^{-n}F_{X,I}^{*n}(u), \tag{26.32}$$

$$\psi(u) = (\rho/(1+\rho))\sum_{n=0}^{\infty}(1+\rho)^{-n}F_{X,I}^{*n}(u). \tag{26.33}$$

Here $\rho = c/(\lambda\mu) - 1$, and $\mu = \mathbb{E}X_1$.

Proof. Note that part (b) follows mimicking the proof of Proposition 25.1, Chapter 25, although $F_{X,I}/(1+\rho)$ is a defective probability measure. So we need to prove (26.31). Let $S_0 = 0$, $S_n = Z_1 + \cdots + Z_n (n \geq 1)$. Then

$$\begin{aligned}
\phi(u) &= P(S_n \leq u \text{ for all } n \geq 1) \\
&= P(Z_1 \leq u, S_n - Z_1 \leq u - Z_1 \text{ for all } n \geq 2) \\
&= \mathbb{E}(\phi(u - Z_1)\mathbf{1}_{[Z_1 \leq u]}) \\
&= \mathbb{E}(\phi(u - X_1 + cA_1)\mathbf{1}_{[X_1 - cA_1 \leq u]}) \\
&= \int_{(0,\infty)} (\int_{[0,ca+u]} \phi(u - x + ca)F_X(dx))\lambda \exp\{-\lambda a\}da \\
&= \int_{[u,\infty)} (\int_{[0,y]} \phi(y - x)F_X(dx))(\lambda/c) \exp\{-\lambda(y - u)/c\}dy \\
&= (\lambda/c) \exp\{\lambda u/c\} \int_{[u,\infty)} \exp\{-\lambda y/c\}(\int_{[0,y]} \phi(y - x)F_X(dx))dy.
\end{aligned}$$

Differentiating with respect to u, this yields

$$\begin{aligned}
\phi'(u) &= (\lambda/c)\phi(u) - (\lambda/c) \int_{[0,u]} \phi(u - x)F_X(dx) \\
&= (\lambda/c)\phi(u) - (\lambda/c)\{F_X(u)\phi(0) + \int_{[0,u]} \phi'(u - x)F_X(x)dx\},
\end{aligned}$$

on integration by parts, and using $F_X(0) = 0$. Integration of this over $[0, z]$ leads to

$$\begin{aligned}
&\phi(z) - \phi(0) \\
&= (\lambda/c) \int_{[0,z]} \phi(u)du - (\lambda/c) \int_{[0,z]} \{\phi(0)F_X(u) + \int_{[0,u]} \phi'(u - x)F_X(x)dx\}du \\
&= (\lambda/c) \int_{[0,z]} \phi(z - u)du - (\lambda/c)\phi(0) \int_{[0,z]} F_X(u)du - (\lambda/c) \int_{[0,z]} F_X(x) \\
&\quad \times \int_{[x,z]} \phi'(u - x)dudx \\
&= (\lambda/c) \int_{[0,z]} \phi(z - u)du - (\lambda/c) \int_{[0,z]} F_X(x)\phi(z - x)dx \\
&= (\lambda/c) \int_{[0,z]} \phi(z - u)(1 - F_X(u))du \\
&= (\lambda/c) \int_{[0,z]} \phi(z - u)\overline{F}_X(u)du.
\end{aligned} \tag{26.34}$$

Letting $z \uparrow \infty$, and using $\phi(\infty) = 1$ (Exercise 5), one obtains

$$\phi(0) = 1 - (\lambda/c) \int_{[0,\infty)} \overline{F}_X(u)du = 1 - \lambda\mu/c. \tag{26.35}$$

Since $\lambda\mu/c = 1/(1+\rho)$, $1-\lambda\mu/c = \rho/(1+\rho)$, the proof of (26.31) is complete.■

Remark 26.3. If $m_Y(q) \equiv \mathbb{E}\exp\{qY\} < \infty$ for some $q > 0$, for a non-negative random variable Y with $P(Y = 0) < 1$, then, given any $\gamma > 1$, there exists a unique $q_1 > 0$ such that $m_Y(q_1) = \gamma$ (Exercise 4).

Corollary 26.5 *(Asymptotic Ruin Probability: Light-Tailed Case).* (a) In addition to the hypothesis of Theorem 26.4, assume that the integrated tail distribution $F_{X,I}$ has a finite mgf $\int_{[0,\infty)} \exp\{qy\}F_{X,I}(dy)$ for some $q > 0$. Then there exists $\theta > 0$ such that

$$(1+\rho)^{-1} \int_{[0,\infty)} \exp\{\theta x\}F_{X,I}(dx) = 1. \tag{26.36}$$

(b) If, in addition to (a),

$$\nu = \int_{[0,\infty)} x(1+\rho)^{-1} \exp\{\theta x\}F_{X,I}(dx) < \infty, \tag{26.37}$$

then one has

$$\psi(u) \sim (1/\nu\theta)(1 - \lambda\mu/c)\exp\{-\theta u\} \tag{26.38}$$

as $u \to \infty$. Here the ratio of the two sides of the relation $\sim$ goes to 1 as $u \to \infty$.

Proof. Part (a) follows from Remark 26.3. For part (b), note that (see (26.31))

$$\begin{aligned}
&\psi(u)\\
&= 1 - \phi(u) = (1+\rho)^{-1} - (1+\rho)^{-1} \int_{(0,u]} (1 - \psi(u-x))F_{X,I}(dx)\\
&= (1+\rho)^{-1} \int_{(u,\infty)} F_{X,I}(dx) + (1+\rho)^{-1} \int_{(0,u]} \psi(u-x)F_{X,I}(dx),
\end{aligned} \tag{26.39}$$

so that

$$\exp\{\theta u\}\psi(u) = \exp\{\theta u\}(1+\rho)^{-1} \int_{(u,\infty)} F_{X,I}(dx) + \int_{[0,u]} \exp\{\theta(u-x)\}\psi(u-x)G(dx), \tag{26.40}$$

where $G(dx)$ is the probability measure $\exp\{\theta x\}F_{X,I}(dx)/(1+\rho)$. Hence $\tilde{\psi}(u) = \exp\{\theta u\}\psi(u)$ satisfies the renewal equation

$$\tilde{\psi}(u) = g(u) + \int_{[0,u]} \tilde{\psi}(u-x)G(dx). \tag{26.41}$$

Note that $g = f_1 f_2$, where $f_1(u) = \exp\{\theta u\}$ is increasing, and $f_2(u) = (1+\rho)^{-1} \int_{(u,\infty)} F_{X,I}(dx)$ is decreasing. Also,

$$\begin{aligned}\int_{[0,\infty)} f_1(u) f_2(u) du &= (1+\rho)^{-1} \int_{[0,\infty)} \exp\{\theta u\}[\int_{(u,\infty)} F_{X,I}(dx)]du \\ &= (1+\rho)^{-1} \int_{[0,\infty)} [\int_{[0,x)} \exp\{\theta u\}du]F_{X,I}(dx) \\ &= (1+\rho)^{-1}(1/\theta) \int_{[0,\infty)} (\exp\{\theta x\} - 1)F_{X,I}(dx) \\ &= (1/\theta)(1 - 1/(1+\rho)) = (1/\theta)(1 - \lambda\mu/c) < \infty. \end{aligned} \tag{26.42}$$

Hence g is directly Riemann integrable. By the Key Renewal Theorem (Corollary 25.8) in Chapter 25, one has

$$\lim_{u\to\infty} \tilde{\psi}(u) = \int_{[0,\infty)} g(x)dx/\nu = (1/\nu\theta)(1 - \lambda\mu/c). \tag{26.43}$$

■

Remark 26.4. On integration by parts, the integral on the left in (26.36) may be shown to equal $-\frac{1}{\theta} + \frac{1}{\theta^2\mu}[-1 + m(\theta)\frac{\lambda + c\theta}{\lambda}]$, where m is the mgf of $Z_i = X_i - cA_i$. It follows from Proposition 11.8 that there exists a unique $\theta = R > 0$ such that $m(\theta) = 1$. Substituting this above and in (26.36), one obtains $\theta = (\frac{c}{\lambda\mu} - 1)^{-1} = \rho^{-1}$ (Exercise 4). Under the additional assumption (b) in Corollary 26.5, it now follows that $\psi(u) \sim de^{-u}$, where $d = \frac{1}{\nu\theta}(1 - \frac{\lambda\mu}{c})$. (Also (26.36) holds with $q = \theta = R$.)

Example 5 *(Ruin Probability When Both Claim Size and Inter-arrival Time Are Exponential)****.*** In the Cramér–Lundberg model, if the claim size X also has the exponential distribution with parameter β, i.e., mean $1/\beta$, then it is not difficult to check that the renewal equation (26.39) is satisfied by

$$\psi(u) = \exp\{-Ru\}/(1+\rho), \tag{26.44}$$

where R is the Lundberg constant $R = \theta = \beta - \lambda/c$ in this case (Exercise 6).

We next turn to the general renewal model, that is, the Sparre–Andersen model. For this we follow the approach using *ascending ladder heights*.[7] For this define the first ascending ladder epoch τ_1,

$$[\tau_1 = n] = [S_j \leq 0 for 1 \leq j \leq n-1, S_n > 0] \quad (n = 1, 2, \dots), \tag{26.45}$$

recalling that $S_n = Z_1 + \cdots + Z_n (n = 1, 2, \dots)$, $S_0 = 0$. In view of the NPC, τ_1 is a defective random variable, with defect $1 - p$, where

$$0 < p = P(\tau_1 < \infty) < 1. \tag{26.46}$$

The second ascending ladder epoch τ_2 of the sequence $\{S_n : n \geq 1\}$ is defined on the set $[\tau_1 < \infty]$ as the first ladder epoch of the sequence $\{S_{\tau_1+n} - S_{\tau_1} : n = 0, 1, 2, \dots\}$. In this manner, one defines the $(j+1)$-th ascending ladder epoch τ_{j+1} on the event $[\tau_i < \infty$ for $i = 1, \dots, j]$ as the first ascending ladder epoch of the sequence $\{S_{\tau_j+n} - S_{\tau_j} : n = 0, 1, 2, \dots\}$. Note that allowing the value ∞ for $\tau_j (j = 1, 2, \dots)$, they are stopping times for the Markov process $\{S_n : n = 0, 1, \dots\}$. The first ascending ladder height L^1 and its distribution are defined by

$$[L^1 \in (0, x], \tau_1 = n] = [S_j \leq 0 \text{ for } 1 \leq j \leq n-1, S_n \in (0, x]], \tag{26.47}$$

$$H_n(dx) = P(L^1 \in dx, \tau_1 = n), \tag{26.48}$$

and

$$[L^1 \in (0, x]] = \cup_{n \geq 1} [L^1 \in (0, x], \tau_1 = n] = [L^1 \in (0, x], \tau_1 < \infty],$$

$$H(dx) = P(L^1 \in (dx), \tau_1 < \infty). \tag{26.49}$$

Hence L^1 is also a defective random variable with defect $1 - p$, being defined only on the event $[\tau_1 < \infty]$. Successive ascending ladder heights $L^{j+1} (j \geq 1)$ are similarly defined, namely, L^{j+1} is the first ascending ladder height of the sequence $\{S_{\tau_j+n} - S_{\tau_j} : n = 0, 1, 2, \dots\}$. Note that, by the strong Markov property, on the set $[\tau_i < \infty$ for $i = 1, \dots, j]$, the conditional distribution of (τ_{j+1}, L^{j+1}) given $\sigma\{(\tau_i, L^i) : i = 1, \dots, j\}$ is the same as the (defective) distribution of (τ_1, L^1). In particular, L^{j+1} is independent of $\{(\tau_i, L^i) : i = 1, \dots, j\}$, given $[\tau_i < \infty$ for $i = 1, \dots, j]$. Expressing the ruin probability (26.24) as

$$\psi(u) = P(M > u), \tag{26.50}$$

[7]Recall the ascending ladder height innovation introduced by Blackwell for the analysis of renewal of random walks; see Chapter 25.

where $M := \max\{0, \sup_{n\geq 1} S_n\} = \max\{0, S_1, S_2, \ldots, S_n, \ldots\}$ *a.s.*, one has

$$\begin{aligned}
\phi(u) &= 1 - \psi(u) = P(M \leq u) \\
&= P(M \leq u, \tau_1 = \infty) + \sum_{n\geq 1} P(M \leq u, \tau_i < \infty \text{ for } i = 1, \ldots, n, \tau_{n+1} = \infty) \\
&= (1-p) + \sum_{n\geq 1} P(M \leq u, \tau_i < \infty \text{ for } i = 1, \ldots, n)(1-p) \\
&= (1-p) + \sum_{n\geq 1} P(L^1 + \cdots + L^n \leq u, \tau_i < \infty \ldots \text{ for } i = 1, \ldots, n)(1-p) \\
&= (1-p)[1 + \sum_{n\geq 1} H^{*n}(u)], \qquad (26.51)
\end{aligned}$$

where $H^{*n}(u) = P(L^1 + \cdots + L^n \leq u, \tau_i < \infty \text{ for } i = 1, \ldots, n)$, and H is the distribution function of the defective random variable L^1 (see (26.49)). For the last equality in (26.51), note that the conditional distribution of L^n on the set $[\tau_n < \infty]$, given $\sigma\{\tau_j, L^j, 1 \leq j \leq n-1\}$, is $H(dx)$. Backward recursion then proves that the joint distribution of $L^j, 1 \leq j \leq n$ on the event $[\tau_i < \infty \text{ for } i = 1, \ldots, n]$, is the product measure $H(dx_1)\cdots H(dx_n)$. Because $H(dx)$ is defective, one cannot apply Blackwell's renewal theorem directly to compute the asymptotic value of

$$\phi(u) = 1 - \psi(u) = (1-p)\sum_{n\geq 0} H^{*n}(u). \qquad (26.52)$$

Here we denote $H^{*0}(dx) = \delta_0(dx)$, δ_0 being the Dirac measure (point mass) at 0.

First consider the light-tailed case (26.25)–(26.27). We will use[8] the i.i.d. random variables $Y_i, i \geq 1$, associated with $Z_i, i \geq 1$, having the corresponding distribution $F_{q^0}(dx) = \exp\{q^0 x\}F(dx)$, where F is the distribution function of Z_i and q^0 is as in (26.29). We will denote the ladder height probabilities of $\{Y_i : i \geq 1\}$ corresponding to the probabilities $H_{n,q^0}(dx)$, $H_{q^0}(dx)$ corresponding to (26.47) and (26.49). Note that $\mathbb{E}Y_i = \int_{\mathbb{R}} x F_{q^0}(dx) = \int_{\mathbb{R}} x \exp\{q^0 x\}F(dx) = m'(q^0) > 0$, since, as shown in the derivation of the Lundberg bound (26.29) with $R = q^0$, m is strictly increasing at q^0 $(> \tilde{q})$ (see the proof of Proposition 11.8). Therefore, the ascending ladder epochs and ascending ladder heights for the sequence $\{Y_i : i \geq 1\}$ are proper (non-defective) random variables. It is easy to check that $F^{*n}_{q^0}(dx) = \exp\{q^0 x\}F^{*n}(dx)$, since the joint distribution of $(Y_1, \ldots, Y_n)$ is $F_{q^0}(dy_1)F_{q^0}(dy_2)\cdots F_{q^0}(dy_n) = \exp\{q^0 y_1\}\exp\{q^0 y_2\}\cdots\exp\{q^0 y_n\}F(dy_1)F(dy_2)\cdots F(dy_n) = \exp\{q^0(y_1 + y_2 + \cdots + y_n)\}F(dy_1)F(dy_2)\cdots F(dy_n)$, which on integration, setting $x =$

[8] This follows the method of Feller (1971), Chapter XII.

$y_1 + y_2 + \cdots + y_n$, yields $\int_{\mathbb{R}} f(x) F_{q^0}^{*n}(dx) = \int_{\mathbb{R}} f(x) \exp\{q^0 x\} F^{*n}(dx)$ for every bounded measurable function f.

Lemma 7. Let $H_{q^0,n}$ and H_q be defined as in (26.49) and (26.50), but with Y_i's in place of Z_i's. Then

$$H_{q^0,n}(dx) = \exp\{q^0 x\} H_n(dx),\ H_{q^0}(dx) = \exp\{q^0 x\} H(dx). \tag{26.53}$$

Proof. One has, for every $x > 0$,

$$\begin{aligned}
&H_{n,q^0}(x) \\
&= P_{q^0}(S_j \leq 0 \text{ for } j = 1, .., n-1,\ S_n \in (0, x]) \\
&= \int \mathbf{1}_{x_1+\cdots+xj \leq 0 \text{ for } j=1,\ldots,n-1, x_1+\cdots+x_n \in (0,x]} \exp\{q^0(x_1 + \cdots + x_n)\} F(dx_1) \cdots F(dx_n) \\
&= \mathbb{E}(\mathbf{1}_{[S_j \leq 0 \text{ for } j=1,\ldots,n-1]} \mathbf{1}_{[S_n \in (0,x]]} \exp\{q^0 S_n)\}) = \mathbb{E}(\mathbb{E}(\ldots | S_n)) \\
&= \int \gamma_n(y) \exp\{q^0 y\} F^{*n}(dy),
\end{aligned}$$

where

$$\gamma_n(y) = \mathbb{E}(\mathbf{1}_{[S_j \leq 0\ rm for\ j=1,\ldots,n-1]} S_n)|_{[S_n=y]}.$$

This shows that $H_{n,q^0}(dy) = \exp\{q^0 y\}\gamma_n(y) F^{*n}(dy) = \exp\{q^0 y\} H_n(dy)$. Consequently, $H_{q^0}(x) = \sum_{n \geq 1} H_{n,q^0}(x) = \int_{(0,x]} \exp\{q^0 y\} H(dy)$, where $H_{q^0}(dx) = \exp\{q^0 x\} H(dx)$. It follows that the convolution $H_{q^0}^{*n}(dx)$ is $\exp\{q^0 x\} H^{*n}(dx)$. This completes the proof of the lemma. ■

As a consequence, one now has

$$H_{q^0}^{*n}(dx) = \exp\{q^0 x\} H^{*n}(dx),\ \ H^{*n}(dx) = \exp\{-q^0 x\} H_{q^0}^{*n}(dx). \tag{26.54}$$

Therefore, by (26.52),

$$\phi(dx) = (1-p) \sum_{n \geq 0} \exp\{-q^0 x\} H^{*n}(dx), \tag{26.55}$$

$$\begin{aligned}
\psi(u) &= (1-p) \sum_{n \geq 0} \int_{(u,\infty)} \exp\{-q^0 x\} H^{*n}(dx) \\
&= (1-p) \int_{(u,\infty)} \exp\{-q^0 x\} G_{q^0}(dx),
\end{aligned} \tag{26.56}$$

where $G_{q^0}(dx)$ is the renewal measure for H_{q^0} and, by Blackwell's renewal theorem,

$$G_{q^0}(dx) = \sum_{n\geq 0}(H^{*n}_{q^0}(dx) \sim b^{-1}dx, \tag{26.57}$$

where $b := \int_{(0,\infty)} xH_{q^0}(dx)$. Using this in (26.56), one arrives at the following theorem.

***Theorem 26.6** (Ruin in the Light-Tailed Sparre–Andersen Model).* Under the assumptions (26.25)–(26.27), letting $R = q^0$, the ruin probability is, asymptotically,

$$\psi(u) \sim (1-p)b^{-1}\int_{(u,\infty)} \exp\{-Rx\}dx = (1-p)\exp\{-Ru\}/bR, \ (u \to \infty). \tag{26.58}$$

We now turn to the case of *heavy-tailed claim sizes*. In particular, the claim size does not have a finite moment generating function $\mathbb{E}\exp\{qX_i\}$ for any $q > 0$, so that

$$m(q) = \mathbb{E}\exp\{qZ_i\} = \infty \ \forall q > 0.$$

In this case, we use the distribution $\tilde{H} = H/p$ to normalize the defective ladder height distribution H. Continuing with the notation $\overline{G}(u) = 1 - G(u)$ for a distribution function G, rewrite (26.52) as

$$\begin{aligned}
\psi(u) &= 1-(1-p)\sum_{n\geq 0} H^{*n}(u) \\
&= 1-(1-p)\sum_{n\geq 0} p^n\tilde{H}^{*n}(u) \\
&= 1-(1-p)\sum_{n\geq 0} p^n(1-\overline{\tilde{H}^{*n}}(u)) \\
&= (1-p)\sum_{n\geq 0} p^n\overline{\tilde{H}^{*n}}(u) \\
&= (1-p)\sum_{n\geq 1} p^n\overline{\tilde{H}^{*n}}(u)
\end{aligned} \tag{26.59}$$

for $u > 0$. Note that $\tilde{H}^{*0} = \delta_0$. Assume that $\tilde{H}$ is subexponential. Then $\tilde{H}$ is heavy-tailed and has the important characterization by Proposition 26.2 For every $n \geq 2$,

$$\overline{\tilde{H}^{*n}}(u)/\overline{\tilde{H}}(u) \to n \text{ as } u \to \infty. \tag{26.60}$$

One may now substitute (26.60) inside the summation in (26.59) to get

$$\begin{aligned}
\psi(u)/\overline{\tilde{H}}(u) &= (1-p)\sum_{n\geq 1} p^n \overline{\tilde{H}^{*n}}(u)/\overline{\tilde{H}}(u) \\
&= (1-p)[p + \sum_{n\geq 2} p^n \overline{\tilde{H}^{*n}}(u)/\overline{\tilde{H}}(u)] \\
&\to (1-p)[p + \sum_{n\geq 2} np^n] \\
&= (1-p)\sum_{n\geq 1} np^n = p/(1-p), \text{ as } u \to \infty. \qquad (26.61)
\end{aligned}$$

The justification of the interchange of the order of taking the limit (26.60) and the summation over n, in (26.61), is provided by Kesten's Lemma 4.

Theorem 26.7 *(Ruin Under Heavy-Tailed Claim Size)***.** In addition to (26.25) and (26.26), assume that the normalized ladder height distribution $\tilde{H}$ is subexponential. Then

$$\psi(u)/\overline{\tilde{H}}(u) \to p/(1-p) \text{ as } u \to \infty, \qquad (26.62)$$

where $p = H(\infty) = P(\tau_1 < \infty)$ is the total mass of the ascending ladder height distribution H.

Since H is less tractable analytically, and perhaps computationally as well, the following result is useful. For this, denote by F_X the distribution of the claim size X with mean μ and by $F_{X,I}$ the integrated tail distribution of the claim size X, with density

$$f_{X,I}(x) = (d/dx)F_{X,I}(x) = (1/\mu)\int_{(x,\infty)} F_X(t)dt, \quad x \in (0,\infty). \qquad (26.63)$$

The following useful result is stated here without proof.[9]

Corollary 26.8. Assume (26.25) and (26.26) and that the integrated claim size distribution $F_{X,I}$ is subexponential. Then, writing μ, λ for the means of the claim size X_i and the inter-arrival time A_i, respectively, one has

$$\psi(u)/\overline{F}_{X,I}(u) \to \mu/(c\lambda - \mu) \text{ as } u \to \infty. \qquad (26.64)$$

Note that (26.62) and (26.64) imply, in particular, the following:

$$\overline{\tilde{H}}(u))/\overline{F}_{X,I}(u) \to \mu/(c\lambda - \mu)]/p(1-p), \text{ as } u \to \infty, \qquad (26.65)$$

[9]See Rolski et al. (1999), Theorem 6.5.11 for the proof.

and

$$\overline{H}(u)/\overline{F_{X,I}}(u) \to \mu/[(c\lambda - \mu)p], \text{ as } u \to \infty. \tag{26.66}$$

Example 6. In the general Sparre–Andersen model with the claim size having the Pareto distribution with density

$$f(x, k, \beta) = \beta k^{\beta} x^{-\beta-1} \mathbf{1}_{(k,\infty)}(x),$$

it follows from (26.64) that

$$\psi(u) = (c/\lambda\mu - 1)^{-1} k^{\beta-1}/u^{\beta-1}(1 + o(1)) \text{ as } u \to \infty. \tag{26.67}$$

Exercises

1. (a) Prove that the function represented in Proposition 26.1 is slowly varying at infinity. [*Hint*: It is enough to show that (i) $\forall a > 1$, $\lim_{x\to\infty} \int_x^{ax} \frac{1}{t} g(t)dt = 0$, and (ii) $\forall 0 < a < 1$, $\lim_{x\to\infty} \int_{ax}^{x} \frac{1}{t} g(t)dt = 0$. Use the fact that, given any $\delta > 0$, however small, there exists $t_1 > 0$ such that $|g(t)| < \delta$ for all $t > t_1$.]
   (b) (i) Assuming the Karamata Representation Theorem 26.1, prove that for all $a > 0$, $x^a L(x) \to \infty$ and $x^{-a} L(x) \to 0$ as $x \to \infty$. (ii) Show that $L(x) = o(\ln x)$ as $x \to \infty$.
2. Suppose that the claim sizes are fixed amounts $X_i = a > 0$ for all $i \geq 1$ in the Cramér–Lundberg model. Show that the Lundberg constant R is a decreasing function of a. [*Hint*: Calculate R using (26.29).]
3. Determine heavy-/light-tailed properties of each of the following distributions: (i) lognormal, (ii) log-Cauchy, (iii) Gamma, and (iv) Weibull distribution with pdf $f(x) = \frac{\theta x^{\theta-1}}{\beta^{\theta}} \exp\{-(\frac{x}{\beta})^{\theta}\}$, $x \geq 0, \theta, \beta > 0$.
4. Verify the computation in Remark 26.4.
5. Complete the proof of (26.35).
6. (i) Verify the formula for the ruin probability for exponential claims and inter-arrivals in Example 5. [*Hint*: If F_X is exponential, then $F_{X,I} = F_X$. Use (26.39).]
   (ii) Do the same for the case in which the claim size distribution is a convex combination of two exponential distributions, say having pdf $\frac{1}{2}\beta_1 e^{-\beta_1 x} + \frac{1}{2}\beta_2 e^{-\beta_2 x}$, $x > 0$, where $\beta_1, \beta_2 > 0$.
7. Which of the claim size distributions in Examples 1–4 satisfy the hypothesis of Theorem 26.7, and what are the asymptotic ruin probabilities? [*Hint*: Examples 1 and 2. Use (26.64).]

Chapter 27
Special Topic: Fractional Brownian Motion and/or Trends: The Hurst Effect

This chapter involves efforts to understand a phenomenon first reported in connection with the Nile River data by Hurst (1951), referred to as the *Hurst phenomena*, and identified as an anomaly by Feller (1951). The history is rich by way of consideration of various possible scenarios ranging from heavy tails, long range dependence, or climatic trends. The chapter is concluded with a brief survey of mathematical aspects of the *fractional Brownian motion* model by Mandelbrot and Wallis (1968) motivated by this application, as well as in other contexts such as mathematical finance[1] and economics. This includes its extension to a random field model indexed by k-dimensional Euclidean space.

Let $\{Y_n : n = 1, 2, \ldots\}$ be a sequence of random variables representing the annual flows into a reservoir. Let $S_n = Y_1 + \cdots + Y_n$, $n \geq 1$, $S_0 = 0$. Also let $\overline{Y}_N = N^{-1}S_N$ denote the mean flow rate over a span of N years. A variety of complicated natural processes (e.g., sediment deposition, erosion, etc.) constrain the life and capacity of a reservoir. However a particular design parameter analyzed extensively by hydrologists, based on an idealization in which water usage and

[1] An interesting construction of Rogers (1997) shows that the fractional Brownian motion model is not arbitrage-free unless it is Brownian motion. That is, unless it is Brownian motion, it is not a semi-martingale, and consequently there cannot exist an equivalent martingale measure; see Chapter 23 for the financial math implications.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_27

natural loss would occur at an annual rate estimated by $\overline{Y}_N$ units per year, is the (dimensionless) statistic defined by

$$\frac{R_N}{D_N} := \frac{M_N - m_N}{D_N}, \tag{27.1}$$

where

$$M_N := \max\{S_n - n\overline{Y}_N; n = 0, 1, \ldots, N\} \tag{27.2}$$

$$m_N := \min\{S_n - n\overline{Y}_N : n = 0, 1, \ldots, N\},$$

$$D_N := \left[\frac{1}{N}\sum_{n=1}^{N}(Y_n - \overline{Y}_N)^2\right]^{\frac{1}{2}}, \quad \overline{Y}_N = \frac{1}{N}S_N. \tag{27.3}$$

Here S_n is the total inflow for the first n years and $n\overline{Y}_N$ the losses at the assumed rate $\overline{Y}_N$. The hydrologist Hurst (1951) stimulated much interest in the possible behaviors of R_N/D_N for large values of N. On the basis of the data analyzed for regions of the Nile River, Hurst published a finding that plots of $\log(R_N/D_N)$ versus $\log N$ are linear with slope $H \approx 0.75$. Feller (1951) was soon to show this to be an anomaly relative to the standard statistical framework of i.i.d. flows $Y_1, \ldots, Y_n$ having finite second moment. The precise form of Feller's analysis is as follows.

Let $\{Y_n\}_{n=1}^{\infty}$ be an i.i.d. sequence with

$$\mathbb{E}Y_n = d, \qquad \mathrm{Var}\, Y_n = \sigma^2 > 0. \tag{27.4}$$

First consider that, by the central limit theorem, $S_N = Nd + O_p(N^{\frac{1}{2}})$ in the sense that $(S_N - Nd)/\sqrt{N}$ is, for large N, distributed approximately like a Gaussian random variable with mean zero and variance σ^2. If one defines

$$\widetilde{M}_N = \max\{S_n - nd : 0 \le n \le N\},$$

$$\widetilde{m}_N = \min\{S_n - nd : 0 \le n \le N\},$$

$$\widetilde{R}_N = \widetilde{M}_N - \widetilde{m}_N, \tag{27.5}$$

then by the functional central limit theorem (FCLT)

$$\left(\frac{\widetilde{M}_N}{\sigma\sqrt{N}}, \frac{\widetilde{m}_N}{\sigma\sqrt{N}}\right) \Longrightarrow (\widetilde{M}, \widetilde{m}), \quad \frac{\widetilde{R}_N}{\sigma\sqrt{N}} \Longrightarrow \widetilde{R} \qquad \text{as } n \to \infty, \tag{27.6}$$

where $\Longrightarrow$ denotes convergence in distribution of the sequence on the left to the distribution of the random variable(s) on the right. Here

$$\begin{aligned} \widetilde{M} &:= \max\{B_t : 0 \le t \le 1\}, \\ \tilde{m} &:= \{B_t : 0 \le t \le 1\}, \\ \widetilde{R} &:= \widetilde{M} - \tilde{m}, \end{aligned} \tag{27.7}$$

with $\{B_t\}_{t\ge 0}$ a standard Brownian motion starting at zero. It follows that the magnitude of $\tilde{R}_N$ is $O_p(N^{\frac{1}{2}})$. Under these circumstances, therefore, one would expect to find a fluctuation between the maximum and the minimum of partial sums, centered on the mean, over a period N to be of the order $N^{\frac{1}{2}}$. To see that this still remains for $R_N/(\sqrt{N}D_N)$ in place of $\widetilde{R}_N/(\sqrt{N}\sigma)$, first note that, by the strong law of large numbers applied to Y_n and Y_n^2 separately, we have with probability 1 that $\overline{Y}_N \to d$, and

$$D_N^2 = \frac{1}{N}\sum_{n=1}^{N} Y_n^2 - \overline{Y}_N^2 \to \mathbb{E}Y_1^2 - d^2 = \sigma^2, \qquad \text{as } N \to \infty. \tag{27.8}$$

Therefore, with probability 1, as $N \to \infty$,

$$\frac{R_N}{\sqrt{N}D_N} \sim \frac{R_N}{\sqrt{N}\sigma}, \tag{27.9}$$

where "$\sim$" indicates "asymptotic equality" in the sense that the ratio of the two sides goes to 1 as $N \to \infty$. This implies that the asymptotic distributions of the two sides of (27.9) are the same. Next notice that, with $\Rightarrow$ denoting convergence in distribution,

$$\begin{aligned} \frac{M_N}{\sigma\sqrt{N}} &= \max_{0\le n\le N}\left(\frac{(S_n - nd) - n(\overline{Y}_N - d)}{\sigma\sqrt{N}}\right) \\ &= \max_{0\le t\le 1}\left(\frac{S_{[Nt]} - [Nt]d}{\sigma\sqrt{N}} - \frac{[Nt]}{N}\left(\frac{S_N - Nd}{\sigma\sqrt{N}}\right)\right) \\ &\Rightarrow \max_{0\le t\le 1}(B_t - tB_1) := M \end{aligned} \tag{27.10}$$

and

$$\frac{m_N}{\sigma\sqrt{N}} = \min_{0\le t\le 1}\left(\frac{S_{[Nt]} = [Nt]d}{\sigma\sqrt{N}} - \frac{[Nt]}{N}\left(\frac{S_N - Nd}{\sigma\sqrt{N}}\right)\right) \Rightarrow \min_{0\le t\le 1}(B_t - tB_1) := m,$$

and

$$\left(\frac{M_N}{\sigma\sqrt{N}}, \frac{m_N}{\sigma\sqrt{N}}\right) \Rightarrow (M, m). \tag{27.11}$$

Therefore,

$$\frac{R_N}{D_N\sqrt{N}} \sim \frac{R_N}{\sigma\sqrt{N}} \Rightarrow M - m = R, \tag{27.12}$$

where R is a strictly positive random variable. Once again then, R_N/D_N, the so-called *rescaled adjusted range* statistic, is of the order of $O_p(N^{\frac{1}{2}})$.

We will say that the *Hurst exponent* is H if $R_N/(D_N N^H)$ converges in distribution to a nonzero real-valued random variable as N tends to infinity. In particular, this includes the case of convergence in probability to a positive constant. The basic problem raised by Hurst is to identify circumstances under which one may obtain an exponent $H > \frac{1}{2}$, representing the so-called *Hurst effect*. The next major theoretical result following Feller was again somewhat negative, though quite insightful. Specifically, Moran (1964) considered the case of i.i.d. random variables $Y_1, Y_2, \dots$ having "heavy tails" in their distribution. In this case, the re-scaling by D_N serves to compensate for the increased fluctuation in R_N to the extent that cancellations occur *resulting again in* $H = \frac{1}{2}$.

In the remainder of this chapter, we will consider two theories that provide $H > \frac{1}{2}$. In the one cited earlier, $H > \frac{1}{2}$ is shown to occur under a stationary but strongly correlated model having moments of all orders. In the other theory,[2] $H > \frac{1}{2}$ is shown to occur for independent but nonstationary flows having finite second moments. In particular, it will be shown that under an appropriately slow trend superimposed on a sequence of i.i.d. random variables, the Hurst effect appears.

Let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence with $\mathbb{E}X_n = d$ and $\operatorname{Var} X_n = \sigma^2$ as above, and let $f(n)$ be an arbitrary real-valued function on the set of positive integers. We assume that the observations Y_n are of the form

$$Y_n = X_n + f(n). \tag{27.13}$$

The partial sums of the observations are

$$\begin{aligned} S_n &= Y_1 + \cdots + Y_n = X_1 + \cdots + X_n + f(1) + \cdots + f(n) \\ &= S_n^* + \sum_{j=1}^{n} f(j), \qquad S_0 = 0, \end{aligned} \tag{27.14}$$

[2]Bhattacharya et al. (1983).

where

$$S_n^* = X_1 + \cdots + X_n. \tag{27.15}$$

Introduce the notation D_N^* for the standard deviation of the X-values $\{X_n : 1 \le n \le N\}$,

$$D_N^{*2} := \frac{1}{N}\sum_{n=1}^{N}(X_n - \overline{X}_N)^2. \tag{27.16}$$

Then, writing $\overline{f}_N = \sum_{n=1}^{N} f(n)/N$,

$$\begin{aligned}
D_N^2 &:= \frac{1}{N}\sum_{n=1}^{N}(Y_n - \overline{Y}_N)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}(X_n - \overline{X}_N)^2 + \frac{1}{N}\sum_{n=1}^{N}(f(n) - \overline{f}_N)^2 \\
&\quad + \frac{2}{N}\sum_{n=1}^{N}(f(n) - \overline{f}_N)(X_n - \overline{X}_N) \\
&= D_N^{*2} + \frac{1}{N}\sum_{n=1}^{N}(f(n) - \overline{f}_N)^2 + \frac{2}{N}\sum_{n=1}^{N}(f(n) - \overline{f}_N)(X_n - \overline{X}_N).
\end{aligned} \tag{27.17}$$

For convenience, write

$$\begin{aligned}
\mu_N(n) &:= \sum_{j=1}^{n}(f(j) - \overline{f}_N), \quad \mu_N(0) = 0, \\
\Delta_N &:= \max_{0\le n\le N}\mu_N(n) - \min_{0\le n\le N}\mu_N(n).
\end{aligned} \tag{27.18}$$

Also write

$$\begin{aligned}
M_N &= \max_{0\le n\le N}\{S_n - n\overline{Y}_N\} = \max_{0\le n\le N}\left\{S_n^* - n\overline{X}_N + \mu_N(n)\right\}, \\
m_N &= \min_{0\le n\le N}\{S_n - n\overline{Y}_N\} = \min_{0\le n\le N}\left\{S_n^* - n\overline{X}_N + \mu_N(n)\right\}, \\
M_N^* &= \max_{0\le n\le N}\{S_n^* - n\overline{X}_N\}, \quad m_N^* = \min_{0\le n\le N}\{S_n^* - n\overline{X}_N\}, \\
R_N &= M_N - m_N, \\
R_N^* &= M_N^* - m_N^*.
\end{aligned} \tag{27.19}$$

Observe that

$$M_N \le \max_{0\le n\le N} \mu_N(n) + \max_{0\le n\le N} (S_n^* - n\overline{X}_N),$$
$$m_N \ge \min_{0\le n\le N} \mu_N(n) + \min_{0\le n\le N} (S_n^* - n\overline{X}_N), \tag{27.20}$$

and

$$M_N \ge \max_{0\le n\le N} \mu_N(n) + \min_{0\le n\le N} (S_n^* - n\overline{X}_N),$$
$$m_N \le \min_{0\le n\le N} \mu_N(n) + \max_{0\le n\le N} (S_n^* - n\overline{X}_N). \tag{27.21}$$

From (27.20), one gets $R_N \le \Delta_N + R_N^*$, and from (27.21), $R_N \ge \Delta_N - R_N^*$. In other words,

$$|R_N - \Delta_N| \le R_N^*. \tag{27.22}$$

In the same manner,

$$|R_N - R_N^*| \le \Delta_N. \tag{27.23}$$

It remains to estimate D_N and Δ_N.

Lemma 1. If $f(n)$ converges to a finite limit, then D_N^2 converges to σ^2 with probability 1.

Proof. In view of (27.17), it suffices to prove

$$\frac{1}{N}\sum_{n=1}^{N}(f(n)-\overline{f}_N)^2+\frac{2}{N}(f(n)-\overline{f}_N)(X_n-\overline{X}_N) \to 0 \qquad \text{as } N\to\infty. \tag{27.24}$$

Let α be the limit of $f(n)$. Then

$$\frac{1}{N}\sum_{n=1}^{N}(f(n)-\overline{f}_N)^2 = \frac{1}{N}\sum_{n=1}^{N}(f(n)-\alpha)^2 - (\overline{f}_N-\alpha)^2. \tag{27.25}$$

Now if a sequence $g(n)$ converges to a limit θ, then so do its arithmetic means $N^{-1}\sum_1^N g(n)$, $N \ge 1$. Applying this to the sequences $(f(n)-\alpha)^2$ and $f(n)$, observe that (27.25) goes to zero as $N \to \infty$. Next

$$\frac{1}{N}\sum_{n=1}^{N}(f(n)-\overline{f}_N)(X_n-\overline{X}_N) = \frac{1}{N}\sum_{n=1}^{N}(f(n)-\alpha)(X_n-d)-(\overline{f}_N-\alpha)(\overline{X}_N-d). \tag{27.26}$$

The second term on the right clearly tends to zero as N increases. Also, by the Schwarz inequality,

$$\left|\frac{1}{N}\sum_{n=1}^{N}(f(n)-\alpha)(X_n-d)\right| \le \frac{1}{N}\left(\sum_{n=1}^{N}(f(n)-\alpha)^2\right)^{\frac{1}{2}}\left(\sum_{n=1}^{N}(X_n-d)^2\right)^{\frac{1}{2}}.$$

By the strong law of large numbers, $N^{-1}\sum_{n=1}^{N}(X_n-d)^2 \to \mathbb{E}(X_1-d)^2=\sigma^2$ and the arithmetic means $N^{-1}\sum_{n=1}^{N}(f(n)-\alpha)^2$ go to zero as $N\to\infty$, since $(f(n)-\alpha)^2\to 0$ as $n\to\infty$. ■

From (27.12), (27.22), and Lemma 1, we get the following result.

Theorem 27.1. If $f(n)$ converges to a finite limit, then, for every $H>\frac{1}{2}$,

$$\left|\frac{R_N}{D_N N^H}-\frac{\Delta_N}{D_N N^H}\right| \to 0 \qquad \text{in probability as } N\to\infty. \tag{27.27}$$

In particular, the Hurst effect with exponent $H>\frac{1}{2}$ holds if and only if, for some positive number c',

$$\lim_{N\to\infty}\frac{\Delta_N}{N^H}=c'. \tag{27.28}$$

Example 1. Take

$$f(n)=\alpha+c(n+m)^{\beta} \quad (n=1,2,\ldots), \tag{27.29}$$

where $\alpha, c, m,$ and β are parameters, with $c\neq 0$, $m\ge 0$. The presence of m indicates the *starting point* of the trend, namely, m units of time before the time $n=0$. Since the asymptotics are not affected by the particular value of m, we assume henceforth that $m=0$ without essential loss of generality. For simplicity, also take $c>0$. The case $c<0$ can be treated in the same way.

First let $\beta<0$. Then $f(n)\to\alpha$, and Theorem 27.1 applies. Recall that

$$\Delta_N=\max_{0\le n\le N}\mu_N(n)-\min_{0\le n\le N}\mu_N(n), \tag{27.30}$$

where

$$\begin{aligned}\mu_N(n)&=\sum_{j=1}^{n}(f(j)-\overline{f}_N) \qquad \text{for } 1\le n\le N,\\ \mu_N(0)&=0.\end{aligned} \tag{27.31}$$

Notice that, with $m = 0$ and $c > 0$,

$$\mu_N(n) - \mu_N(n-1) = c\left(n^\beta - \frac{1}{N}\sum_{j=1}^{N} j^\beta\right) \tag{27.32}$$

is positive for $n < (N^{-1}\sum_{j=1}^{N} j^\beta)^{1/\beta}$ and is negative or zero otherwise. This shows that the *maximum* of $\mu_N(n)$ is attained at $n = n_0$ given by

$$n_0 = \left[\left(\frac{1}{N}\sum_{j=1}^{N} j^\beta\right)^{1/\beta}\right], \tag{27.33}$$

where $[x]$ denotes the *integer part* of x. The minimum value of $\mu_N(n)$ is *zero*, attained at $n = 0$ and $n = N$. Thus,

$$\Delta_N = \mu_N(n_0) = c\sum_{k=1}^{n_0}\left(k^\beta - \frac{1}{N}\sum_{j=1}^{N} j^\beta\right). \tag{27.34}$$

By a comparison with a Riemann sum approximation to $\int_0^1 x^\beta dx$, one obtains

$$\frac{1}{N}\sum_{j=1}^{N} j^\beta = N^\beta \sum_{j=1}^{N}\left(\frac{j}{N}\right)^\beta \frac{1}{N} \sim \begin{cases} (1+\beta)^{-1}N^\beta & \text{for } \beta > -1 \\ N^{-1}\log N & \text{for } \beta = -1 \\ N^{-1}\sum_{j=1}^{\infty} j^\beta & \text{for } \beta < -1. \end{cases} \tag{27.35}$$

By (27.33) and (27.35),

$$n_0 \sim \begin{cases} (1+\beta)^{-1/\beta}N & \text{for } \beta > -1 \\ N/\log N & \text{for } \beta = -1 \\ \left(\sum_1^\infty j^\beta\right)^{1/\beta} N^{1/(-\beta)} & \text{for } \beta < -1. \end{cases} \tag{27.36}$$

From (27.34)–(27.36), it follows that

$$\Delta_N \sim \begin{cases} cn_0\left(\frac{n_0^\beta}{1+\beta} - \frac{N^\beta}{1+\beta}\right) \sim c_1 N^{1+\beta}, & -1 < \beta < 0, \\ cn_0(n_0^{-1}\log n_0 - N^{-1}\log N) \sim c\log N, & \beta = -1, \\ c\sum_{j=1}^{\infty} j^\beta = c_2, & \beta < -1. \end{cases} \tag{27.37}$$

Here $c_1 = c(-\beta)(1+\beta)^{-2-1/\beta}$, and c_2 is a positive constant depending only on β. Now consider the following cases.

Case 1: $-\frac{1}{2} < \beta < 0$. In this case, Theorem 27.1 applies with $H(\beta) = 1 + \beta > \frac{1}{2}$. Note that, by Lemma 1, $D_N \sim \sigma$ with probability 1. Therefore,

$$\frac{R_N}{D_N N^{1+\beta}} \to c_1 > 0 \qquad \text{in probability as } N \to \infty, \text{ if } -\frac{1}{2} < \beta < 0. \tag{27.38}$$

Case 2: $\beta < -\frac{1}{2}$. Use inequality (27.23), and note from (27.37) that $\Delta_N = o_p(N^{\frac{1}{2}})$. Dividing both sides of (27.23) by $D_N N^{\frac{1}{2}}$, one gets, in probability as $N \to \infty$,

$$\frac{R_N}{D_N N^{\frac{1}{2}}} \sim \frac{R_N^*}{D_N N^{\frac{1}{2}}} \sim \frac{R_N^*}{\sigma N^{\frac{1}{2}}} \qquad \text{if } \beta < -\frac{1}{2}. \tag{27.39}$$

But $R_N^*/\sigma N^{\frac{1}{2}}$ converges in distribution to R by (27.11). Therefore, the Hurst exponent is $H(\beta) = \frac{1}{2}$.
Case 3: $\beta = 0$. In this case, the Y_n are i.i.d. Therefore, as proved at the outset, the Hurst exponent is $\frac{1}{2}$.
Case 4: $\beta > 0$. In this case, Lemma 1 does not apply, but a simple computation yields

$$D_N \sim c_3 N^{\beta} \qquad \text{with probability 1 as } N \to \infty, \text{ if } \beta > 0. \tag{27.40}$$

Here $c_3 = \beta/(\beta + 1)$. Combining (27.40), (27.37), and (27.22), one gets

$$\frac{R_N}{N D_N} \to c_4 \qquad \text{in probability as } N \to \infty, \text{ if } \beta > 0, \tag{27.41}$$

where c_4 is a positive constant. Therefore, $H(\beta) = 1$.
Case 5: $\beta = -\frac{1}{2}$. In this case, one considers the process $\{Z_N(s)\}_{0 \le s \le 1}$ defined by

$$Z_N\left(\frac{n}{N}\right) = \frac{S_n - n\overline{Y}_N}{\sqrt{N} D_N} \qquad \text{for } n = 1, 2, \ldots, N,\ Z_N(0) = 0,$$

and linearly interpolated between n/N and $(n+1)/N$. Observe that $Z_N(n/N) = (S_n^* - n\overline{X}_N)/\sqrt{N} D_N - \mu_N(n/N)/\sqrt{N} D_N$ and that the polygonal process corresponding to $(S_n^* - n\overline{X}_N)/\sqrt{N} D_N$ converges in distribution to a Brownian bridge $\{B_s^*\}_{0 \le s \le 1}$, by the FCLT (Theorem 17.2). On the other hand, for $0 < t \le 1$,

$$\begin{aligned}
\mu_N([Nt]) &= c \sum_{j=1}^{[Nt]} j^{-\frac{1}{2}} - c \frac{[Nt]}{N} \sum_{j=1}^{N} j^{-\frac{1}{2}} \\
&= c\sqrt{N} \left\{ \sum_{j=1}^{[Nt]} \left(\frac{j}{N}\right)^{-\frac{1}{2}} \cdot \frac{1}{N} - \frac{[Nt]}{N} \sum_{j=1}^{N} \left(\frac{j}{N}\right)^{-\frac{1}{2}} \cdot \frac{1}{N} \right\} \sim 2c\sqrt{N}(\sqrt{t} - t).
\end{aligned} \tag{27.42}$$

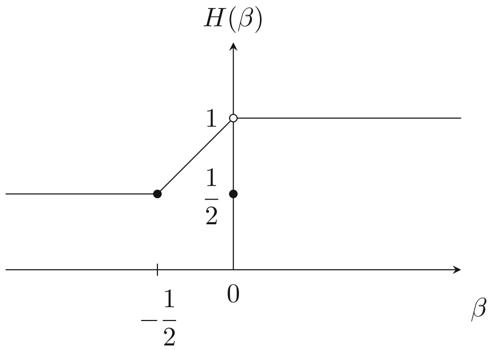


Fig. 27.1 The Hurst exponent as a Function of the Trend Rate β

Thus $\{Z_N(s)\}_{0\le s\le 1}$ converges in distribution to $\{Z_s := B_s^* + 2c\sqrt{s}(1-\sqrt{s})\}_{0\le s\le 1}$, and the asymptotic distribution of $R_N/(\sqrt{N}D_N)$ is the *non-degenerate* distribution of

$$\max_{0\le s\le 1} Z_s^* - \min_{0\le s\le 1} Z_s^*.$$

In particular, $H(-\frac{1}{2}) = \frac{1}{2}$.

The graph in Figure 27.1 of $H(\beta)$ versus β summarizes the results of the preceding cases 1 through 5.

For purposes of data analysis, note that (27.38) implies

$$\log \frac{R_N}{D_N} - \left[\log c' + (1+\beta)\log N\right] \to 0 \qquad \text{if } -\frac{1}{2} < \beta < 0. \tag{27.43}$$

In other words, for large N, the plot of $\log R_N/D_N$ against $\log N$ should be approximately linear with slope $H = 1+\beta$, if $-\frac{1}{2} < \beta < 0$.

This concludes the analysis of the statistic R_N/D_N for data in which there is a time varying trend.

For an alternative,[3] explanation of the occurrence of $H > \frac{1}{2}$ suppose that Y_1, $Y_2, \ldots$ is a stationary Gaussian sequence of random variables with mean zero and covariances $\gamma(k) = \mathbb{E}\{Y_n Y_{n+k}\}$, $k = 0, 1, 2, \ldots, \sigma^2 := \gamma(0)$. By *stationarity* it is meant that the distribution of $\{Y_n\}_{n=1}^{\infty}$ is invariant under time shifts, i.e., it equals the distribution of $\{Y_{n+k}\}_{n=1}^{\infty}$ for every $k = 1, 2, \ldots$. One may note that there is no loss in generality in centering the observations for the statistics in (27.16) and (27.19) when the mean is constant. We will consider the case of a slow decay of correlations of the form

$$\gamma(k) \sim ck^{-\theta} \qquad \text{as } k \to \infty \tag{27.44}$$

[3]See Mandelbrot and Wallis (1968).

for some $0 < \theta \leq 1$, $c > 0$. A proof that such covariance functions exist will be given below.

With $H \equiv H(\theta) = \frac{2-\theta}{2} \in (1/2, 1]$, let

$$Y_t^{(N)} = \frac{S_{[Nt]}}{N^H}, \quad t \geq 0, \tag{27.45}$$

and let $\{\tilde{Y}_t^{(N)}\}_{t\geq 0}$ be the corresponding polygonal process obtained by linear interpolation of the values $\frac{S_{[Nt]}}{N^H}$, $t = 0, \frac{1}{N}, \frac{2}{N}, \ldots$. Then the finite-dimensional distributions of $\{Y_t^{(N)}\}_{t\geq 0}$ (and $\{\tilde{Y}_t^{(N)}\}_{t\geq 0}$) are Gaussian with mean zero and for $t_1 < t_2$, one has as $N \to \infty$

$$\begin{aligned}
&\mathbb{E}\left(Y_{t_2}^{(N)} - Y_{t_1}^{(N)}\right)^2 \\
&\quad\equiv N^{2H+\theta-2}\mathbb{E}\left(Y_{t_2}^{(N)} - Y_{t_1}^{(N)}\right)^2 \\
&\quad= \frac{[N(t_2-t_1)]\sigma^2}{N^{2-\theta}} + \frac{2}{N^{2-\theta}} \sum_{k=1}^{[N(t_2-t_1)]-1} [N(t_2-t_1)-k]\gamma(k) + o(1) \\
&\quad= o(1) + 2\sum_{k=1}^{[N(t_2-t_1)]-1} \left[t_2 - t_1 - \frac{k}{N}\right]\left(\frac{k}{N}\right)^{-\theta} \frac{\gamma(k)}{k^{-\theta}} \cdot \frac{1}{N} \\
&\quad\to c'(t_2-t_1)^{2-\theta}\int_0^1 (1-x)x^{-\theta}dx = c''(t_2-t_1)^{2-\theta}.
\end{aligned} \tag{27.46}$$

From (27.46), it follows that $\mathrm{Var}(Y_t^{(N)}) \to c''t^{2-\theta}$ and $\mathrm{Cov}(Y_{t_1}^{(N)}, Y_{t_2}^{(N)}) \to \frac{1}{2}c''\{t_2^{2-\theta} + t_1^{2-\theta} - (t_2-t_1)^{2-\theta}\} = f(t_1,t_2)$, say $(t_1 < t_2)$. Therefore, the mean vector and covariance matrix of the (Gaussian) finite-dimensional distributions converge. This establishes convergence of the Gaussian finite-dimensional distributions.

Since, as in (17.6), for $\delta > 0$, using Chebyshev's inequality, one has

$$\begin{aligned}
&P\left(\sup_{0\leq t\leq T}\left|Y_t^{(N)} - \tilde{Y}_t^{(N)}\right| > \delta\right) \\
&\qquad\leq P\left(\frac{|Y_m|}{N^H} > \delta \qquad \text{for some } m = 1, 2, \ldots, [NT]+1\right) \\
&\qquad\leq ([NT]+1)\frac{\mathbb{E}|Y_1|^r}{N^{rH}\delta^r} = o(1) \qquad \text{as } N \to \infty,
\end{aligned} \tag{27.47}$$

for any choice of $r > \frac{1}{H} = \frac{2}{2-\theta}$, the limit distributions of $\{Y_t^{(N)}\}_{t\geq 0}$ and $\{\tilde{Y}_t^{(N)}\}_{t\geq 0}$ coincide. The existence of a limit distribution on the space $C[0,\infty)$ follows by a tightness computation (Exercise 2). The corresponding limit process is a mean zero

Gaussian process $\{B_t^{(H)} : t \geq 0\}$ having continuous sample paths with covariance $\mathbb{E}\{B_t^{(H)} B_s^{(H)}\} = \frac{1}{2}\{s^{2H} + t^{2H} - |t-s|^{2H}\}$ $(2H = 2 - \theta > 1)$ referred to as a *fractional Brownian motion*. The case $\theta = 1$ is Brownian motion.

The functional limit theorem (weak convergence) above extends[4] to non-Gaussian stationary sequences $\{Y_n\}_{n=1}^{\infty}$ with correlations of the form (27.44) as well. To conclude this chapter, let us calculate the Hurst exponent for this model. We have, in the same manner as (27.10) and (27.11) but with the N^H scaling, that as $N \to \infty$,

$$\left(\frac{M_N}{\sigma N^H}, \frac{m_N}{\sigma N^H}\right) \Longrightarrow \left(\max_{0 \leq t \leq 1} (B_t^{(H)} - tB_1^{(H)}), \min_{0 \leq t \leq 1} (B_t^{(H)} - tB_1^{(H)})\right). \tag{27.48}$$

Also since $\overline{Y}_N = \frac{S_N}{N}$ has mean zero and since $\mathbb{E}\overline{Y}_N^2 \equiv (N^{2H}/N^2)\mathbb{E}(Y_1^{(N)})^2 \to 0$ by (27.46), one has $\overline{Y}_N \to 0$ in probability as $N \to \infty$. Similarly, $\mathbb{E}Y_i^2 = \sigma^2$, and $\mathbb{E}\{(Y_i^2 - \sigma^2)(Y_{i+k}^2 - \sigma^2)\} = 2\gamma^2(k)$, implies that $D_N^2 = \frac{1}{N}\sum_{i=1}^N (Y_i - \overline{Y}_N)^2 = \frac{1}{N}\sum_{i=1}^N Y_i^2 - (\overline{Y}_N)^2 \to \sigma^2$ in probability as $N \to \infty$, since

$$\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^N Y_i^2\right) = \frac{\mathbb{E}(Y_1^2 - \sigma^2)^2}{N} + \frac{2}{N^2}\sum_{i=1}^{N-1}\sum_{k=1}^{N-i} \mathbb{E}\{(Y_i^2 - \sigma^2)(Y_{i+k}^2 - \sigma^2)\} \to 0$$

as $N \to \infty$.

In the case $0 < \theta < 1$, the power law decay of correlations is often referred to as a *long range dependence* in this context.[5] However this is only a reflection of the scaling by N^H and non-Brownian limit. For Brownian motion ($\theta = 1$), the latter is an *independent* sequence.

Remark 27.1. The model (27.13), (27.14) is that of a random walk with a *nonlinear growth in mean*. It has been shown that if there is a trend $\mathbb{E}Y_n = f(n)$, which decays slowly, as is the case in Example 1 with $f(n) = O(n^\beta)$ and $-\frac{1}{2} < \beta < 0$, then the Hurst effect is manifested. It should be noted that the assumption "$\{X_n\}_{n=1}^{\infty}$ is i.i.d." in this case can be relaxed to "$\{X_n\}_{n=1}^{\infty}$ is *stationary and weakly dependent*."[6] That is, all results derived for the model (27.13) and Example 1 carry over to the case of a stationary sequence $\{X_n\}_{n=1}^{\infty}$ having finite second moments for which both the strong law of large numbers and the FCLT hold under the same scaling as in the i.i.d. case. Observe that the only properties of the sequence $\{X_n\}_{n=1}^{\infty}$ that have been made use of are as follows: as $N \to \infty$, (1) $\overline{X}_N \to \mathbb{E}X_1$ a.s., $\frac{1}{N}\sum_{j=1}^N X_j^2 \to \mathbb{E}X_1^2$ a.s. and (2) the polygonal process $\{\tilde{X}_t^{(H)}\}_{t \geq 0}$ obtained by linear interpolation of

[4]See Samorodnitsky and Taqqu (2016) for comprehensive treatment of central limit theory under long range dependence.

[5]Also see Samorodnitsky (2006).

[6]See Bradley (2007) for comprehensive treatment of limit theorems under weak dependence.

$\tilde{X}^{(N)}_{n/N} := (S^*_n - nd)/\sqrt{N}$ $(n = 0, 1, \ldots)$ converges in distribution to a Brownian motion $\{X_t\}_{t\geq 0}$ with a positive diffusion coefficient. Thus the Hurst effect shows up for "weakly dependent" $\{Y_n\}_{n=1}^{\infty}$ with slowly decaying mean, as well as for "strongly dependent" stationary $\{Y_n\}_{n=1}^{\infty}$.

The fractional Brownian motion naturally extends to a random field as follows.

Definition 27.1. The fractional Brownian random field with exponent $H \in (0, 1)$ is the mean zero Gaussian process $\{B^{(H)}_x : x \in \mathbb{R}^k\}$ having continuous sample paths with covariance

$$\Gamma_H(x, y) = \mathbb{E}\{B^{(H)}_x B^{(H)}_y\} = \frac{1}{2}\{|x|^{2H} + |y|^{2H} - |x - y|^{2H}\}, \quad x, y \in \mathbb{R}^k.$$

The case $H = \frac{1}{2}$ defines a random field referred to as *multiparameter Brownian motion*.[7] However, for this definition to serve as a starting point, one needs to check that (i) $\Gamma_H(x, y)$ is positive-definite[8] and (ii) there is a version with continuous paths. This second point can easily be obtained as a consequence of the Kolmogorov–Chentsov condition and is left as Exercise 1.

Lemma 2. Fix $0 < H < 1$, and define

$$\gamma_H(s, r) = |s - r|^{H-\frac{1}{2}}\mathrm{sign}(s - r) + |r|^{H-\frac{1}{2}}\mathrm{sign}(r), \quad s, r \in \mathbb{R}.$$

Then $\gamma_H(t, \cdot) \in L^2(\mathbb{R})$ for each t and

$$\langle \gamma_H(t, \cdot), \gamma_H(s, \cdot)\rangle = c_H(|s|^{2H} + |t|^{2H} - |s - t|^{2H}),$$

for a positive constant c_H.

Proof. First note that $|1 - u|^{H-\frac{1}{2}}\mathrm{sign}(u)$ and $|u|^{H-\frac{1}{2}}\mathrm{sign}(u)$ are locally square-integrable in neighborhoods of $u = 0$ and $u = 1$, $\gamma_H(0, r) = 0$ and $\gamma_H^2(r, t) = |t|^{2H-1}\gamma_H^2(1, \frac{r}{t})$, $t \neq 0$. Moreover, for $H < 1$,

$$\begin{aligned}\int_{-\infty}^{-2} \gamma_H^2(1, u)du &+ \int_2^{\infty} \gamma_H^2(1, u)du \\ &= 2\int_2^{\infty} \gamma_H^2(1, u)du = 2\int_2^{\infty} \{u^{H-\frac{1}{2}} - (u - 1)^{H-\frac{1}{2}}\}^2 du\end{aligned}$$

[7] Lévy (1945).

[8] Historically the positive-definiteness problem was solved for multiparameter Brownian motion by Schoenberg (1938) and more generally for $0 < H < 1$ by Gangolli (1967). The simpler proof here is given in Ossiander and Waymire (1989).

$$= 2\int_2^\infty \big(\int_{u-1}^u (H - \frac{1}{2}) v^{H-\frac{3}{2}} dv\big)^2 du$$
$$\leq 2(H - \frac{1}{2})^2 \int_2^\infty \int_{u-1}^u v^{2H-3} dv du$$
$$\leq 2(H - \frac{1}{2})^2 \int_2^\infty (u-1)^{2H-3} du < \infty. \tag{27.49}$$

Thus, for each t, $\gamma_H(t, \cdot) \in L^2(\mathbb{R})$, and $||\gamma_H(t, \cdot)||_2^2 = 2c_H|t|^{2H}$, where $c_H = \frac{1}{2}||\gamma_H(1, \cdot)||_2^2 > 0$. So, $\gamma_H(t, \cdot) - \gamma_H(s, \cdot) \in L^2(\mathbb{R})$, and

$$||\gamma_H(t, \cdot) - \gamma_H(s, \cdot)||_2^2 = \int_{\mathbb{R}} \{|t - r|^{H-\frac{1}{2}} \mathrm{sign}(t - r) - |s - r|^{H-\frac{1}{2}} \mathrm{sign}(s - r)\}^2 dr$$
$$= \int_{\mathbb{R}} \{|t - s - y|^{H-\frac{1}{2}} \mathrm{sign}(t - s - y) + |y|^{H-\frac{1}{2}} \mathrm{sign}(y)\}^2 dy$$
$$= ||\gamma_H(t - s, \cdot)||_2^2 = 2c_H|t - s|^{2H}. \tag{27.50}$$

The assertion follows since

$$\langle\gamma_H(t, \cdot), \gamma_H(s, \cdot)\rangle = \frac{1}{2}\{||\gamma_H(t, \cdot)||_2^2 + ||\gamma_H(s, \cdot)||_2^2 - ||\gamma_H(t, \cdot) - \gamma_H(s, \cdot)||_2^2\}.$$

■

Proposition 27.2. For $H \in (0, 1)$, the kernel $\Gamma_H(x, y), x, y \in \mathbb{R}^k$ is positive-definite.

Proof. Transforming to spherical coordinates, one has

$$\int_{S^{k-1}} \int_{\mathbb{R}^k} \gamma_H(\langle x, \theta\rangle, r)\gamma_H(\langle y, \theta\rangle, r) dr d\theta$$
$$= \int_{S^{k-1}} \frac{1}{2} c_H \{|\langle x, \theta\rangle|^{2H} + |\langle y, \theta\rangle|^{2H} - |\langle x - y, \theta\rangle|^{2H}\} d\theta$$
$$= C_{k,H}\{|x|^{2H} + |y|^{2H} - |x - y|^{2H}\}, \tag{27.51}$$

where $C_{k,H} = \frac{1}{2} c_H \int_{S^{k-1}} |\langle\varphi, \theta\rangle|^{2H} d\theta$ does not depend on $\varphi \in S^{k-1}$. Thus, one has

$$\sum_{1\leq i,j\leq m} \Gamma_H(x_i, x_j) = \sum_{1\leq i,j\leq m} c_i c_j \int_{S^{k-1}} \int_{\mathbb{R}^k} \gamma_H(\langle x_i, \theta\rangle, r)\gamma_H(\langle x_j, \theta\rangle, r) dr d\theta$$
$$= \int_{S^{k-1}} \int_{\mathbb{R}^k} \big(\sum_{i=1}^m c_i \gamma_H(\langle x_i, \theta\rangle, r)\big)^2 dr d\theta \geq 0.$$

■

Exercises

1. Use the Kolmogorov–Chentsov condition to show the existence of a continuous version of the fractional Brownian random field. [*Hint*: Check that $B_x^{(H)} - B_y^{(H)}$ is Gaussian with mean zero and variance $\mathbb{E}|B_x^{(H)} - B_y^{(H)}|^2 = |x-y|^{2H}$ from which the higher order moments follow as $\mathbb{E}|B_x^{(H)} - B_y^{(H)}|^{2k} = \frac{(2k)!}{k!2^k}|x-y|^{2kH}, k \geq 1$. Choose $k > \frac{1}{2H}$.]
2. Prove tightness of the Gaussian polygonal process $\{Y_t^{(N)}\}_{t\geq 0}$. [*Hint*: It is sufficient to show that there are positive numbers α, β, and M such that $\mathbb{E}|Y_t^{(N)} - Y_s^{(N)}|^{\alpha} \leq M|t-s|^{1+\beta}$. Make computations similar to those for the 4th-moment proof[9] of the FCLT.]
3. Show that $\mathbb{E}(B_t^{(H)} - B_s^{(H)})^2 = |t-s|^{2H}$.
4. (*Self-similarity of Fractional Brownian Motion*) Show that the distribution of the fractional Brownian motion with exponent H is invariant under the transformation $W_t = \lambda^{-H} B_{\lambda t}^{(H)},\ t \geq 0$, for any fixed but arbitrary $\lambda > 0$.
5. Use the Kolmogorov–Chentsov theorem to show that on any finite time interval, fractional Brownian motion with the Hurst exponent H is a.s. H older continuous of order θ for any $0 < \theta < H$.
6. Suppose that $Y_1, \ldots, Y_n$ is an i.i.d. sequence with symmetric stable distribution having characteristic function $\mathbb{E}e^{i\xi Y_1} = e^{-|\xi|^{\alpha}}, \xi \in \mathbb{R}$, for an exponent $0 < \alpha \leq 2$. In this exercise, we compute the expected (unadjusted) range of the sums as a function of n.

   (i) Show that $n^{-\frac{1}{\alpha}} S_n := n^{-\frac{1}{\alpha}}(Y_1 + \cdots + Y_n) =^{\text{dist}} Y_1$ for each $n \geq 1$. Assume that $1 < \alpha \leq 2$ in the remainder of exercises (ii)–(iv).
   (ii) Show that $\mathbb{E}|Y_1| < \infty$ and a.s. $\frac{S_n}{n} \to \mathbb{E}Y_1 = 0$.
   (iii) Let $c = \mathbb{E}\max\{0, Y_1\} > 0$. Show that $\mathbb{E}S_n^+ = n^{\frac{1}{\alpha}}c$. Let $M_n = \max\{0, S_1, \ldots, S_n\}$, and show $\mathbb{E}M_n = c\sum_{k=1}^{n} k^{\frac{1}{\alpha}-1} \sim c\alpha n^{\frac{1}{\alpha}}$ as $n \to \infty$.
   (iv) Let $m_n = \min\{0, S_1, \ldots, S_n\}$, and show $\mathbb{E}(M_n - m_n) = 2\mathbb{E}M_n \sim 2c\alpha n^{\frac{1}{\alpha}}$ as $n \to \infty$. [Hint: Use scaling and Spitzer (1956) combinatorial lemma in (iii).]

[9] See BCPT, p. 152.

Chapter 28
Special Topic: Incompressible Navier–Stokes Equations and the Le Jan–Sznitman Cascade

The three-dimensional incompressible Navier–Stokes equations are nonlinear partial differential equations formulated in an effort to embody the basic physics of fluid flow in accordance with Newton's laws of motion (See Landau and Lifshitz (1987)). The equations are involved in models of fluid flow with applications ranging from modeling ocean currents in oceanography to blood flow in medicine, among many others. However, understanding the existence and/or uniqueness of smooth solutions to these equations, when the initial data is smooth, presents one of the great unsolved mathematical challenges of the twentieth and twenty-first centuries (Ladyzhenskaya (2003), Fefferman (2006)). The main goal of the present chapter is to present a probabilistic cascade model of Le Jan and Sznitman (1997) and its subsequent extensions, in which solutions may be represented as expected values of a certain vector product over a random tree, provided the expectations exist.

Assuming constant (unit) mass density, the equations take the form of four nonlinear partial differential equations in four unknowns of a scalar pressure p and the three scalar components of velocity $v = (v_1, v_2, v_3)$. In vector form, the incompressible Navier–Stokes equations in free space are

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = \nu \Delta v - \nabla p + g, \;\; \nabla \cdot v = 0, \;\; v(0^+, x) = v_0(x), x \in \mathbb{R}^3, \tag{28.1}$$

where $\nu > 0$ is a positive (viscosity) parameter, $v_0 : \mathbb{R}^3 \to \mathbb{R}^3$ is the initial velocity data, $\nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3})$, $\Delta = \nabla \cdot \nabla = \sum_{j=1}^3 \frac{\partial^2}{\partial x_j^2}$, and $g : [0, \infty) \times \mathbb{R}^3 \to \mathbb{R}^3$ is an external forcing. It is to be emphasized that the incompressibility condition

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8_28

$\nabla \cdot v = 0$ is to hold at all points in time and space as part of the system of equations, including for the initial data. For convenience, we will also assume the forcing to be incompressible. In coordinate form, the system of equations involves four unknown scalar coordinates, v_1, v_2, v_3 and the pressure p, governed by the following four scalar equations:

$$\frac{\partial v_j}{\partial t} + \sum_{k=1}^{3} v_k \frac{\partial v_j}{\partial x_k} = \nu \sum_{k=1}^{3} \frac{\partial^2 v_j}{\partial x_k^2} - \frac{\partial p}{\partial x_j} + g_j, \quad \sum_{k=1}^{3} \frac{\partial v_k}{\partial x_k} = 0, \;\; j = 1, 2, 3, \tag{28.2}$$

in addition to the given three coordinates of initial velocity.

The left-hand side of the equation is the acceleration of the fluid at time-space location (t, x) when measured in the Eulerian coordinate system. That is, one views the (unit mass density) fluid from the perspective of fluid parcels passing by the location x at time t. In particular, the nonlinear term is an artifact of the frame of reference, as opposed to modeling assumptions, and is therefore intrinsic to the equations of motion. The right-hand side describes the dispersive (linearized stress-strain relations on the fluid), the pressure, and the external forces acting on the fluid that produce the flow in accordance with Newton's "F= MA."The equation $\nabla \cdot v = 0$ is the incompressibility condition for the flow, assuming constant fluid mass density.

Remark 28.1. As a matter of first principles, the velocity field at time t and spatial location x is $v(t, x)$. The position in $\mathbb{R}^3$ of a fluid parcel at time t originating at $x(0) = a$, denoted by $x(t, a)$, is subject to the instantaneous velocity field $v(t, x(t, a))$. The acceleration is, according to the chain rule, given as $\frac{d}{dt} v(t, x(t, a)) = \frac{\partial v}{\partial t} + v \cdot \nabla v$. If one ignores viscous forces, then the force per unit mass $\frac{\partial v}{\partial t} + (v \cdot \nabla) v$ is a conservative force and, therefore, must be the gradient of a potential function. This potential is referred to as the *pressure*. The equations take the form

$$\frac{\partial v}{\partial t} + (v \cdot \nabla) v = -\nabla p, \quad \nabla \cdot v = 0, \;\; v(0^+, x) = v_0(x), x \in \mathbb{R}^3, \tag{28.3}$$

and are called the *Euler equations*. The inclusion of the term $\nu\Delta$ in the Navier–Stokes equations is an attempt to model and resolve intrinsic viscous forces.

The question of whether given any smooth initial velocity field $v(0, x) = v_0(x)$, there is (in a suitable sense) a unique smooth solution to (28.1) for all $t > 0$ and all $x \in \mathbb{R}^3$ remains a major unsolved problem for mathematics and physics. Furthermore, to solve this problem,[1] "One must leave both the choice of the phase space and the class of generalized solutions to the researcher without prescribing to him" infinite smoothness or some other smoothness of solutions. The only requirement needed is indeed that the uniqueness theorem must hold in the chosen class of generalized solutions." To the point, it is known that a weak solution

[1] This quote is attributed to Ladyzhenskaya (2003).

exists[2] for the incompressible Navier–Stokes equations, given smooth initial data, that is unique for as long as it remains a smooth solution. The lack of uniqueness of a smooth solution would imply the existence of a non-smooth solution for the given initial data. However the uniqueness problem has remained open in the absence of a small initial data constraint. In this regard, the Le Jan–Sznitman cascade[3] provides a novel probabilistic framework in which to address representations of solutions for these equations as expected values of certain naturally associated multiplicative cascades, especially when relaxed to the *Le Jan–Sznitman cascade without coin tossing*[4]

Remark 28.2. Historically,[5] the two-dimensional incompressible Navier–Stokes equations were shown to have unique globally stable weak solutions[6] on the torus (i.e., for bounded domain with periodic boundary conditions). In three dimensions, similar global existence results for weak solutions were also obtained by Leray[7] and, later by Hopf[8] but in contrast to 2d, uniqueness and smoothness remain open in 3d.

The Le Jan–Sznitman cascade originates with a probabilistic representation of solutions to a Fourier-transformed version of (28.1) in integrated or the so-called *mild* form. Specifically, denote the Fourier transform of a vector-valued function $f \equiv (f_1, f_2, f_3) : \mathbb{R}^3 \to \mathbb{R}^3$ by $\hat{f} = (\hat{f}_1, \hat{f}_2, \hat{f}_3)$, where

$$\hat{f}_j(\xi) = (2\pi)^{-\frac{3}{2}} \int_{\mathbb{R}^3} e^{-i\xi\cdot x} f_j(x)dx, \quad \xi \in \mathbb{R}^3, j = 1, 2, 3,$$

when it exists.[9]

---

[2]Leray proved existence of a global weak solution for any initial data in $L^2(\mathbb{R}^3)$, i.e. finite energy. Hopf (1951) reached the same conclusion for a bounded domain with homogeneous Dirichlet boundary and finite energy conditions. The existence of a global weak solution for any smooth initial data was shown by Lemarié-Rieusset (2002). Their uniqueness and smoothness remain open in 3d.

[3]Le Jan and Sznitman (1997).

[4]The Le Jan–Sznitman cascade can be relaxed to permit possible explosion of the associated branching process as a route to non-uniqueness in Dascaliuc et al. (2015). However the full treatment exceeds the space limitations on this chapter and will not be presented in this volume beyond remarks and an example illustration of ideas at the end of this chapter.

[5]See Ladyzhenskaya (2003) for a more detailed overview of the historical developments.

[6]See Chapter 4 of the PhD thesis of Leray (1933).

[7]Leray (1934): Sur le mouvement d'un liquide visqueux emplissant l'espace, *Acta. Math.* **63** loc cit.

[8]Hopf (1951).

[9]See BCPT Chap. VI for a comprehensive treatment of the Fourier transform. It may be noted that the definition of Fourier transform in the present chapter is a standard variant in analysis on the version usually encountered in probability as the characteristic function, and however the proofs of properties are unchanged. The essential differences are in the complex conjugate of $e^{i\xi\cdot x}$ and integration with respect to $\frac{1}{(2\pi)^{\frac{3}{2}}}dx$ in place of dx. This latter factor is often selected to make the

First note that taking Fourier transforms, one sees that incompressibility is the condition that the vector field $\hat{v}(t, \xi)$ is orthogonal to ξ at all times. On the other hand, the Fourier transform of the pressure term $i\hat{p}(\xi)\xi$ is in the direction of $i\xi$. In particular the term involving pressure may be removed by applying a linear projection $\pi_{\xi^\perp} w := w - (e_\xi \cdot w)e_\xi,\ e_\xi = \frac{\xi}{|\xi|},\ (\xi \neq 0),\ w \in \mathbb{R}^3$, to the equation in the direction orthogonal to ξ, referred to as a *Leray projection*. Assuming incompressible forcing terms, $\hat{g}$ is invariant under this projection, as is $\hat{v}$. Secondly, by integration by parts, the Fourier transforms of derivatives are Fourier multipliers and the Fourier transform of a product results in a convolution. This projection, together with multiplying by $e^{-\nu|\xi|^2 t}$ and integrating, yields the *mild form* of the equations as follows (for $j = 1, 2, 3$):

$$\frac{\partial v_j}{\partial t} + \sum_{k=1}^{3}(2\pi)^{\frac{3}{2}} \int_{\mathbb{R}^3} \hat{v}_k(t, \xi - \omega)(-i\omega_k)\widehat{\pi_{\xi^\perp} v}_j(t, \omega)d\omega$$
$$= -\nu|\xi|^2 \hat{v}_j(t, \xi) + \hat{g}_j(t, \xi). \tag{28.4}$$

Equivalently,

$$\frac{\partial}{\partial t}(e^{\nu|\xi|^2 t}\hat{v}_j(t, \xi)) = \sum_{k=1}^{3} e^{\nu|\xi|^2 t}(2\pi)^{\frac{3}{2}}\{\int_{\mathbb{R}^3} \hat{v}_k(t, \xi - \omega)(i\omega_k)\widehat{\pi_{\xi^\perp} v}_j(t, \omega)d\omega\}$$
$$+ e^{\nu|\xi|^2 t}\hat{g}_j(t, \xi), \quad j = 1, 2, 3. \tag{28.5}$$

Integrating with respect to t, one has

$$\hat{v}_j(t, \xi)$$
$$= e^{-\nu|\xi|^2 t}\hat{v}_0(\xi) + \sum_{k=1}^{3}(2\pi)^{\frac{3}{2}} \int_0^t e^{-\nu|\xi|^2(t-s)}$$
$$\times \{\int_{\mathbb{R}^3} \hat{v}_k(s, \xi - \omega)(i\omega_k)\widehat{\pi_{\xi^\perp} v}_j(s, \omega)d\omega\}ds$$
$$+ \int_0^t e^{-\nu|\xi|^2(t-s)}\hat{g}_j(s, \xi)ds, \quad j = 1, 2, 3. \tag{28.6}$$

Fourier transform an isometry on the $L^2(\mathbb{R}^3, dx)$-space. The minus sign stems from a historical connection with computing coefficients in Fourier series.

In vector form, one has, for $\xi \in \mathbb{R}^3\backslash\{0\}$,

$$\hat{v}(t,\xi) = e^{-\nu|\xi|^2 t}\hat{v}_0(\xi) + \int_0^t e^{-\nu|\xi|^2(t-s)} \int_{\mathbb{R}^3} |\xi|(2\pi)^{\frac{3}{2}}\hat{v}(s,\xi-\omega) \otimes_{e_\xi} \hat{v}(s,\omega)d\omega$$
$$+ \int_0^t e^{-\nu|\xi|^2(t-s)}\hat{g}(s,\xi)ds, \quad \xi\cdot\hat{v}(t,\xi) = 0,\ \hat{v}(0^+,x) = \hat{v}_0(\xi), \tag{28.7}$$

where for $\xi \in \mathbb{R}^3\backslash\{0\}$, $\otimes_{e_\xi}$ is the (nonassociative) vector product defined by

$$w \otimes_{e_\xi} z = i(e_\xi \cdot z)\pi_{\xi^\perp} w, \quad w, z \in \mathbb{C}^3, \tag{28.8}$$

where $e_\xi = \frac{\xi}{|\xi|}, \xi \neq 0$. The scaling of ξ by $|\xi|$ in this product is responsible for the factor $|\xi|$ appearing in (28.7). From a probabilistic perspective, it is convenient to recognize the exponentially distributed holding time $P(T_\xi > t) = e^{-\nu|\xi|^2 t}$ in an initial state $\xi \neq 0$, with pdf $\nu|\xi|^2 e^{-\nu|\xi|^2 s}$, $s \geq 0$. In particular, with a temporary focus on the time parameter, introducing (and removing) another factor of $|\xi|$, and changing variables $s \to t-s$, one obtains the equivalent form:

$$\hat{v}(t,\xi) = e^{-\nu|\xi|^2 t}\hat{v}_0(\xi) + \int_0^t \nu|\xi|^2 e^{-\nu|\xi|^2 s}\{\frac{1}{2}\int_{\mathbb{R}^3} \frac{2(2\pi)^{\frac{3}{2}}}{\nu|\xi|}$$
$$\hat{v}(t-s,\omega) \otimes_{e_\xi} \hat{v}(t-s,\xi-\omega)d\omega + \frac{1}{2}\frac{2\hat{g}(t-s,\xi)}{\nu|\xi|^2}\}ds. \tag{28.9}$$

The "fair coin toss" probability $\frac{1}{2}$ is introduced and removed by a factor of two. In the suggested expected value representation of the equation, this randomizes between twice the (rescaled) forcing term, including zero forcing, and twice the branched product term. To accommodate spatial averaging, suppose that $h : \mathbb{R}^3\backslash\{0\} \to (0,\infty)$ is a positive measurable function such that

$$h * h(\xi) \leq C|\xi|h(\xi), \quad \xi \neq 0, \tag{28.10}$$

for some constant $C > 0$. Note that replacing h by h/C is equivalent to the standardization $C = 1$. Such functions satisfying (28.10) are referred to as *majorizing kernels* in this context.[10] An important example (Exercise 1) for its scaling properties is given by

$$h(\xi) = \frac{1}{\pi^3|\xi|^2}, \quad \xi \neq 0. \tag{28.11}$$

[10] See Bhattacharya et al. (2003).

This kernel will be referred to as the *dilogarithmic kernel* for reasons provided in Exercise 2. Another is the *Bessel kernel* defined by

$$h(\xi) = \frac{1}{2\pi|\xi|}e^{-|\xi|}, \quad \xi \neq 0. \tag{28.12}$$

Each of the examples was included in the original development by Le Jan and Sznitman, where, notably, one has equality in (28.10). However a variety of additional examples[11] may be constructed that naturally include, for example, the parametric family (Exercise 5)

$$h_\beta^{(\alpha)}(\xi) = |\xi|^{\beta-2}e^{-\alpha|\xi|^\beta}, \xi \neq 0, 0 \leq \beta \leq 1, \alpha > 0. \tag{28.13}$$

A given majorizing kernel can be used to rescale by

$$u(t,\xi) = \frac{1}{h(\xi)}\hat{v}(t,\xi), \quad \xi \neq 0. \tag{28.14}$$

Given a majorizing kernel h, let

$$m(\xi) = \frac{2(2\pi)^{\frac{3}{2}} h * h(\xi)}{\nu|\xi|h(\xi)}, \quad \varphi(t,\xi) = \frac{2\hat{g}(t,\xi)}{\nu|\xi|^2 h(\xi)}, \xi \neq 0. \tag{28.15}$$

Then, multiplying (28.7) by $\frac{1}{h(\xi)}$, one has the following:

$$\begin{aligned} u(t,\xi) = e^{-\nu|\xi|^2 t}u_0(\xi) + \int_0^t \nu|\xi|^2 e^{-\nu|\xi|^2 s}\{\frac{1}{2}m(\xi)\int_{\mathbb{R}^3\times\mathbb{R}^3} u(t-s,w_1) \\ \otimes_{e_\xi} u(t-s,w_2)H(d\omega_1\times d\omega_2|\xi) + \frac{1}{2}\varphi(t-s,\xi)\}ds, \end{aligned} \tag{28.16}$$

where for each $\xi \neq 0$, and for any bounded measurable function $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$,

$$\int_{\mathbb{R}^3\times\mathbb{R}^3} f(\omega_1,\omega_2)H(d\omega_1\times d\omega_2|\xi) = \int_{\mathbb{R}^3} f(\omega_1,\xi-w_1)\frac{h(\omega_1)h(\xi-\omega_1)}{h*h(\xi)}d\omega_1. \tag{28.17}$$

Now (28.16) may be viewed as an equation in expected values of a recursively defined branching cascade as follows: Starting with wave number $\xi \neq 0$ and time horizon $t \geq 0$, the process holds for an exponentially distributed time T_ξ with parameter $\nu|\xi|^2$. On the event $[T_\xi > t]$, the cascade terminates with value $u_0(\xi)$, but on the event $[T_\xi \leq t]$, in the remaining time $t - T_\xi$, one tosses a fair coin κ_θ, independently of the value of T_ξ, which, if (tail) $[\kappa_\theta = 0]$, determines a terminal outcome of $\varphi(T_\xi,\xi)$ for the process, while, if (head) $[\kappa_\theta = 1]$, renews this process

[11] Bhattacharya et al. (2003).

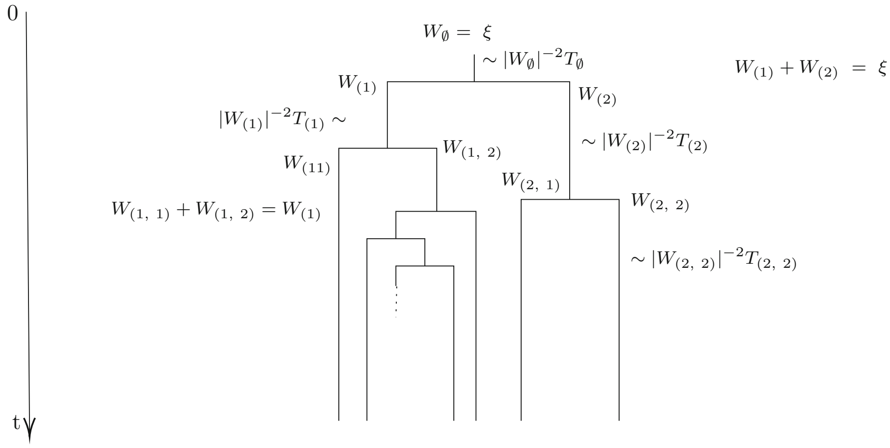


Fig. 28.1 Le Jan–Sznitman Cascade

from a pair of new wave numbers (W_1, W_2), $W_1 + W_2 = \xi$, generated according to the distribution $H(d\omega_1 \times d\omega_2|\xi)$, with respective holding times T_{W_1}, T_{W_2}, conditionally independent and exponentially distributed given (W_1, W_2). Since the *critical* branching process terminates at all fair coin tosses that result in tail, the branching process cannot grow indefinitely. Thus, iterating this process produces an a.s. finite nested set of terminal values of a finite random tree, denoted $\tau_\theta(t, \xi)$. Let $X(t, \xi)$ denote the nested (nonassociative) $\otimes_\xi$ product of terminal values obtained for this cascade by time t. In view of the recursive definition of $X(t, \xi)$, one may write (with $\xi_\theta = \xi$)

$$X(t,\xi) = \begin{cases} u_0(\xi_\theta), & T_{\xi_\theta} \geq t \\ \varphi(t - T_{\xi_\theta}, \xi_\theta), & T_{\xi_\theta} < t, \kappa_\theta = 0 \\ m(\xi_\theta)X^{(1)}(t - T_{\xi_\theta}, W_1) \otimes_{\xi_\theta} X^{(2)}(t - T_{\xi_\theta}, W_2), & T_{\xi_\theta} < t, \kappa_\theta = 1 \end{cases}, \tag{28.18}$$

where the root wave number $\xi_\theta = \xi$ and subsequent holding times, coin tosses, etc. are updated according to this recursion rule to define $X^{(1)}$ and $X^{(2)}$ by the same rules at the updated root wave numbers W_1 and W_2, respectively, as depicted in Figure 28.1. $u(t, \xi) = \mathbb{E}_\xi X(t, \xi)$ solves (28.16), provided that the indicated expected value exists .

The following theorem provides a global representation as an expected value of a vector product over an a.s. finite tree for sufficiently small initial data relative to the selected majorizing kernel.

First let us note that an extension of Fourier transform to *tempered distributions*[12] is achieved as follows. A function $\varphi \in C(\mathbb{R}^k)$ is said to have *rapid*

[12]See Reed and Simon (1972) for a more comprehensive treatment of the theory of distributions in analysis.

decay if $\sup_{x\in\mathbb{R}^k}(1+|x|^n)|\varphi(x)| < \infty$ for $n = 1, 2, \ldots$. Here multi-indices $\alpha, \beta \in \{0, 1, 2, \ldots\}^k$ are used to define powers and orders of derivatives as $x^\alpha = (x_1, \ldots, x_k)^\alpha = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$, and $\partial^\beta = \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}} \cdots \frac{\partial^{\beta_k}}{\partial x_k^{\beta_k}}$.

The *Schwartz space* $\mathcal{S}$ is a vector space consisting of all (test) functions $\varphi \in C(\mathbb{R}^\infty)$ such that φ and all of its partial derivatives have rapid decay. In particular, $C_c^\infty(\mathbb{R}^k) \subset \mathcal{S} \subset L^2(\mathbb{R}^k)$. So $\mathcal{S}$ is a dense subspace of L^2. Defining $||\varphi||_{n,\beta} = \sup_{x\in\mathbb{R}^k}|(1+|x|)^n\partial^\beta\varphi(x)|$, one may write $\mathcal{S} = \{\varphi \in C(\mathbb{R}^\infty) : ||\varphi||_{n,\beta} < \infty \text{ for all } n, \beta\}$. In general, $||\varphi||_{\alpha,\beta} = \sup_{x\in\mathbb{R}^k}|x^\alpha\partial^\beta\varphi(x)|$ provides a family of semi-norms for $\mathcal{S}$; the failure to be a norm is due to the fact that $||\varphi||_{\alpha,\beta} = 0$ does not imply $\varphi = 0$ (unless $\alpha = \beta = 0$). The function

$$d(\varphi, \psi) = \sum_{\alpha,\beta} 2^{-|\alpha|-|\beta|} \frac{||\varphi-\psi||_{\alpha,\beta}}{1+||\varphi-\psi||_{\alpha,\beta}}, \tag{28.19}$$

defines a metric for $\mathcal{S}$ such that $\varphi_n \to \varphi$ as $n \to \infty$ if and only if $||\varphi_n - \varphi||_{\alpha,\beta} \to 0$ for all α, β. In particular $(\mathcal{S}, d)$ is a complete metric space, and $(C_c^\infty(\mathbb{R}^k), d)$ is a dense subspace.

The space $\mathcal{S}'$ of *tempered distributions* is the dual space of $\mathcal{S}$ consisting of all continuous linear functionals f on $\mathcal{S}$; continuous in the sense that if $\varphi_n \to \varphi$ in $(\mathcal{S}, d)$, then $f(\varphi_n) \to f(\varphi)$ as $n \to \infty$. Since the Fourier transforms of test functions $\varphi \in \mathcal{S}$ are well-defined as L^2-functions, the definition of the Fourier transform of a tempered distribution $f \in \mathcal{S}'$ extends to be the tempered distribution $\hat{f}$ given by $\hat{f}(\varphi) = f(\hat{\varphi})$.

Let $\mathcal{F}_{h,T}$ denote the Banach space completion of $\{v \in \mathcal{S}' : ||v||_{\mathcal{F}_{h,T}} = \sup_{\xi\neq 0, 0\le t\le T} \frac{|\hat{v}(t,\xi)|}{h(\xi)} < \infty\}$, where $\mathcal{S}'$ denotes the space of tempered distributions on $\mathbb{R}^3$.

Theorem 28.1. [13] Let h be a normalized majorizing kernel. Let $0 < T \le +\infty$ be fixed but arbitrary, and assume that $||v_0||_{\mathcal{F}_{h,T}} \le (\sqrt{2\pi})^3\nu/2$, and $||(-\Delta)^{-1}g||_{\mathcal{F}_{h,T}} \le (\sqrt{2\pi})^3\nu^2/4$. Then (28.7) has a unique solution in the ball $\mathcal{B}_0(0, R)$ centered at 0 of radius $R = (\sqrt{2\pi})^3\nu/2$ in the space $\mathcal{F}_{h,T}$. Moreover, the Fourier transform is given by

$$\hat{v}(t, \xi) = h(\xi)\mathbb{E}X(t, \xi), \quad \xi \in \mathbb{R}^3\backslash\{0\}.$$

Proof. Recall, $m(\xi) = \frac{2(2\pi)^{\frac{3}{2}} h*h(\xi)}{\nu|\xi|h(\xi)}$, $u(t, \xi) = \frac{\hat{v}(t,\xi)}{h(\xi)}$, and $\varphi(t, \xi) = \frac{2\hat{g}(t,\xi)}{\nu|\xi|^2h(\xi)}$. Under the hypothesis of the theorem, one has $|u_0(\xi)| \le 1$, and $|\varphi(t, \xi)| \le 1$ for all $t, \xi \neq 0$, so that from (28.18), one has $|X(t, \xi)| \le 1$ a.s. In particular $\mathbb{E}X(t, \xi)$ exists. Moreover, taking expected values and conditioning on T_θ show that $u(t, \xi)$ is a solution. Uniqueness is proven as follows. Suppose that $w(t, \xi)$ is another global

[13] See Bhattacharya et al. (2003).

solution with $|w(t,\xi)| \leq Rh(\xi)$. Without loss of generality, one may replace h by $h_\nu = c_\nu h$, where $c_\nu = R = \nu(2\pi)^{\frac{3}{2}}$ and define $\gamma(t,\xi) = c_\nu/h_\nu(\xi)$. Then $\sup_{\xi\neq 0, 0\leq t\leq T} |\gamma(t,\xi)| \leq 1$. Define a truncation of the cascade at time t, denoted $\tau_\theta(t)$, by

$$\tau^{(n)}(t,\xi_\theta) = \{v \in \tau_\theta(t) : |v| \leq n\}, \quad n = 0, 1, \ldots \tag{28.20}$$

Let $Y(\tau_\theta^{(n)}(t,\xi_\theta))$ be the recursively defined random functional given by

$$Y(\tau_\theta^{(n)}(t,\xi_\theta)) = \begin{cases} w_0(\xi_\theta), & T_\theta \geq t \\ \varphi(\xi_\theta, t - T_\theta) & T_\theta < t, \kappa_\theta = 0 \\ m(\xi_\theta) Y(\tau_{(1)}^{(n-1)}(t - T_\theta, W_1)) \otimes_{\xi_\theta} Y(\tau_{(2)}^{(n-1)}(t - T_\theta, W_2)), & T_\theta < t, \kappa_\theta = 1 \end{cases}, \tag{28.21}$$

for $n = 1, 2, \ldots$, where $w_0(\xi) = u_0(\xi)/h_\nu(\xi)$, $\varphi(t,\xi) = 2\hat{g}(t,\xi)/(\nu|\xi|^2 h_\nu(\xi))$, $m(\xi) = 2h_\nu * h_\nu(\xi)/)\nu(2\pi)^{\frac{3}{2}}|\xi|h_\nu(\xi) \leq 1$, and

$$Y(\tau_\theta^{(0)}(t,\xi_\theta)) = \begin{cases} w_0(\xi_\theta), & T_\theta \geq t \\ \varphi(\xi_\theta, t - T_\theta) & T_\theta < t, \kappa_\theta = 0 \\ m(\xi_\theta)\gamma(t - T_\theta, \xi_{(1)}) \otimes_{\xi_\theta} \gamma(t - T_\theta, \xi_{(2)}), & T_\theta < t, \kappa_\theta = 1 \end{cases}. \tag{28.22}$$

Observe that since $w(t,\xi)$ is assumed to be a solution, it follows from the equation that $\gamma(t,\xi) = \mathbb{E}_\xi Y(\tau_\theta^{(1)}(t,\xi))$. Moreover, this extends by induction by conditioning on $\mathcal{F}_n = \sigma(T_v, \xi_v, \kappa_v : |v| \leq n)$ to yield

$$\gamma(t,\xi) = \mathbb{E}_\xi Y(\tau_\theta^{(n)}(t,\xi)), \quad n = 0, 1, 2, \ldots \tag{28.23}$$

In fact, $Y(\tau_\theta^{(n)}(t,\xi)), n = 01, 2, \ldots$, is a martingale with respect to this filtration (Exercise 4). Specifically, one has constant expected values

$$\begin{aligned} &\mathbb{E}_{\xi_\theta=\xi} Y(\tau_\theta^{(n+1)}(t,\xi)) \\ &= w_0(\xi)e^{-\nu|\xi|^2 t} + \frac{1}{2}\int_0^t \nu|\xi|^2 e^{-\nu|\xi|^2 s}\varphi(t-s,\xi)ds \\ &\quad + m(\xi)\mathbb{E}_{\xi_\theta=\xi}\{Y(\tau_{(1)}^{(n)}(t - T_\theta, W_1)) \otimes_{\xi_\theta} Y(\tau_{(2)}^{(n)}(t - T_\theta, W_2))|T_\theta \leq t, \kappa_\theta = 1\} \\ &= w_0(\xi)e^{-\nu|\xi|^2 t} + \frac{1}{2}\int_0^t \nu|\xi|^2 e^{-\nu|\xi|^2 s}\varphi(t-s,\xi)ds + m(\xi)\frac{1}{2}\int_0^t \nu|\xi|^2 e^{-\nu|\xi|^2 s} \\ &\quad \cdot \int \mathbb{E}_{\xi_{(1)}} Y(\tau_{(1)}^{(n)}(t-s,\xi_{(1)})) \otimes_{e_\xi} \mathbb{E}_{\xi_{(2)}} Y(\tau_{(2)}^{(n)}(t-s,\xi_{(2)}))ds. \end{aligned} \tag{28.24}$$

Now observe that

$$Y(\tau_\theta^{(0)}, \xi) = X(t,\xi) \quad \text{on} \quad [\tau_\theta^{(0)} = \tau_\theta(t,\xi)]. \tag{28.25}$$

and more generally, since the terms $\gamma(t - R_v, \xi_v)$ appear in Y only at truncated vertices,

$$Y(\tau_\theta^{(n)}, \xi) = X(t, \xi) \quad \text{on} \quad [\tau_\theta^{(n)} = \tau_\theta(t, \xi)]. \tag{28.26}$$

Thus, since

$$\mathbb{E}|Y(\tau_\theta^{(n)}(t, \xi))| \leq 1, n = 0, 1, 2, \ldots, \tag{28.27}$$

and

$$\mathbb{E}|X(t, \xi)| \leq 1, \tag{28.28}$$

we have

$$\begin{aligned}|\gamma(t, \xi) - \mathbb{E}X(t, \xi)| &= |\mathbb{E}Y(\tau_\theta^{(n)}(t, \xi)) - X(t, \xi)\mathbf{1}[\tau_\theta^{(n)}(t, \xi) \neq \tau_\theta(t, \xi)]| \\ &\leq 2P(\tau_\theta^{(n)}(t, \xi) \neq \tau_\theta(t, \xi) \to 0 \text{ as } n \to \infty, \end{aligned} \tag{28.29}$$

completing the proof of global existence and uniqueness for small initial data. ∎

The simple example[14] below is introduced to illustrate the potential value[15] for analysis of non-uniqueness as the result of modifying the Le Jan–Sznitman cascade by eliminating the coin toss and consequently permit possible explosion. By eliminating the coin toss in the Le Jan–Sznitman cascade, one may ignore (exploding) paths of finite length in the definition of the cascade and retain the expected value representation assuming integrability. In this regard, it is insightful to consider the explosion/non-uniqueness problem for mean-field cascades in which the wave numbers are replaced by fixed constants $\alpha > 0$ as in the following example.

Example 1 ***(α-Riccati Equation).*** Fix a parameter $\alpha > 0$, and consider the initial value problem

$$\frac{du}{dt} = -u(t) + u^2(\alpha t), \quad u(0) = u_0. \tag{28.30}$$

In mild form, one has

$$u(t) = u_0 e^{-t} + \int_0^t e^{-s} u^2(\alpha(t - s))ds, \quad t \geq 0. \tag{28.31}$$

[14] This example is related to self-similar solutions to a mean-field version of the Navier–Stokes and/or Burgers equations in Dascaliuc et al. (2018) and Dascaliuc et al. (2019), respectively.

[15] In another direction, an expected value representation naturally suggests possible numerical Monte-Carlo schemes, e.g., see Ramirez (2006).

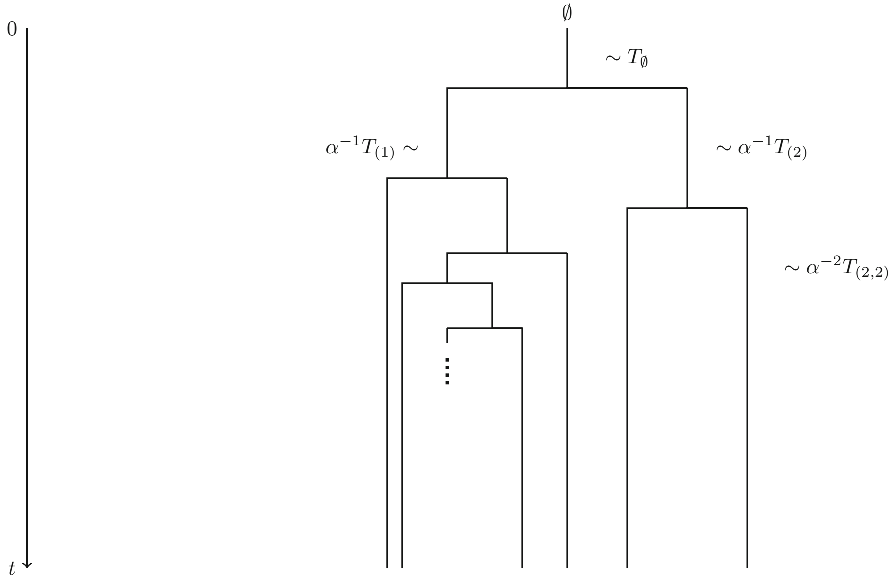


Fig. 28.2 α-Riccati Cascade

The stochastic cascade associated with this equation is depicted in Figure 28.2. In the case $\alpha = 1$, this is the well-known Riccati differential equation, thus the name α-Riccati.[16]

As with the Navier–Stokes equations, we wish to consider the uniqueness problem for global solutions. Adopting a probabilistic perspective on (28.31), let T_0 denote a mean one exponentially distributed random variable. If the event $[T_0 > t]$ occurs, then the process terminates with value u_0. On the event $[T_0 \leq t]$, the process branches and repeats in the remaining time $t - T_0$ with a pair of independent exponential holding times T_1 and T_2, each having mean α. To code this cascade, let $\{T_v : v \in \mathbb{T} = \cup_{n=0}^{\infty}\{1, 2\}^n\}$ be a family of i.i.d. mean one exponentially distributed random variables. Define the evolutionary process

$$V^{(\alpha)}(t) = \left\{v \in \mathbf{T} : \sum_{j=0}^{|v|-1} \alpha^{-j} T_{v|j} \leq t < \sum_{j=0}^{|v|} \alpha^{-j} T_{v|j}\right\}, \quad t \geq 0. \tag{28.32}$$

[16]The uniqueness problem for this equation with initial data $u_0 = 0$ was analyzed by Athreya (1985) and more generally for arbitrary initial data by Dascaliuc et al. (2018) as a multiplicative cascade of the type described in this chapter.

Accordingly, $V^{(\alpha)}$ is a continuous time jump Markov process (Exercise 7) taking value in the (countable) space $\mathcal{E}$ of *evolutionary sets* defined inductively by $V \in \mathcal{E}$ if and only if V is a *finite* subset of $\mathbf{T} = \cup_{n=0}^{\infty}\{1,2\}^n$, such that

$$V = \begin{cases} \{\theta\} & \text{if } \#V = 1, \\ W\backslash\{w\} \cup \{<w1>, <w2>\} & \text{some } W \in \mathcal{E},\ \#W = \#V - 1,\ w \in W,\ \text{else.} \end{cases}$$

Remark 28.3. Although $V^{(\alpha)}$ is a Markov process on $\mathcal{E}$, the cardinality functional $\#V^{(\alpha)}$ is *not* generally Markov, exceptions being $\alpha = \frac{1}{2}, 1$. When $\alpha = 1$, $\#V^{(\alpha)}$ is the classical Yule process, and so it is obviously Markov, while the case $\alpha = \frac{1}{2}$ is also special since it is the Poisson process (Exercise 7).

Let $N^{(\alpha)}(t) = |V^{(\alpha)}(t)| \equiv \#V^{(\alpha)}(t), t \geq 0$, denote the cardinality of the set $V^{(\alpha)}(t)$.

Theorem 28.2. Let $\underline{X}(t) = u_0^{N^{(\alpha)}(t)} 1[S \geq t], t \geq 0$, where

$$S = \inf_{\mathbf{s}\in\{1,2\}^\infty} \sum_{j=0}^{\infty} \alpha^{-j} T_{\mathbf{s}|j}.$$

If $\underline{u}(t) = \mathbb{E}\underline{X}(t) < \infty$, then u is a solution to (28.31). Moreover, if u is any solution to (28.31) with $u(0) = u_0$, then $\underline{u} \leq u$.

Proof. Since $N^{(\alpha)}(0) = 1$, $S \geq 0$, one has $\underline{X}(0) = u_0$ a.s. So $\underline{u}(0) = u_0$. Observe that

$$\underline{X}(t) = \begin{cases} 0 & \text{if } S < t \\ u_0 & \text{if } S \geq T_0 \geq t \\ \underline{X}^{(1)}(\alpha(t - T_0))\underline{X}^{(2)}(\alpha(t - T_0)) & \text{if} T_0 < t \leq S, \end{cases}$$

where $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent and distributed as $\underline{X}$. The equation for $\underline{u}$ follows by an expected value of this decomposition, conditioning on T_0 in the last case. To prove minimality of $\underline{u}$, assume $u(t)$ is a global solution. Define the following sequences of stochastic processes:

$$X_0(t) = 0, \qquad X_n(t) = \begin{cases} u_0, & T_\theta \geq t \\ X_{n-1}^{(1)}(\alpha(t - T_\theta))\, X_{n-1}^{(2)}(\alpha(t - T_\theta)), & T_\theta < t \end{cases}, \quad n \in \mathbb{N}, \tag{28.33}$$

and

$$Y_0(t) = u(t), \quad Y_n(t) = \begin{cases} u_0, & T_\theta \geq t \\ Y_{n-1}^{(1)}(\alpha(t - T_\theta))\, Y_{n-1}^{(2)}(\alpha(t - T_\theta)), & T_\theta < t \end{cases}, \quad n \in \mathbb{N}. \tag{28.34}$$

In the above, $X_n^{(1)}$ and $X_n^{(2)}$ are i.i.d. as X_n, same for Y. More explicitly,

$$X_n(t) = u_0^{N_n(t)} 0^{M_n(t)},$$

where $N_n(t)$ is the number of paths v with $|v| < n$ that cross t and $M_n(t)$ is the number of paths of length $|v| = n$ that survive by time t. Also,

$$Y_n(t) = u_0^{N_n(t)} \prod_{\substack{|v|=n, \\ v \text{survives by } t}} u(\tau_v)\ ,$$

where $\tau_v = \alpha\,(\tau_{v|k} - T_{v|k})$ with $k = |v| - 1$ and $\tau_\theta = t$.

Clearly, $X_n(t) \leq Y_n(t)$ a.s. Moreover, since $X_n(t)$ is eventually monotone (constant) in n if $S \geq t$ and $X_n(t) = 0$ in $S < t$, we see that

$$\lim_{n\to\infty} X_n(t) = \underline{X}(t).$$

Also, using induction on n, $\mathbb{E}(Y_n(t)) = u(t)$ for all $n \in \mathbb{N}$. Thus, by Fatou's lemma, $\mathbb{E}(\underline{X}(t)) \leq u(t)$, which proves the assertion on minimality of $\underline{u}$. ■

The random variable S denotes the *explosion time* for the branching cascade. A simple illustration of its role in the non-uniqueness is illustrated by the following corollary.

Corollary 28.3. Suppose that $u_0 = 1$. Then $u(t) \equiv 1$ for all $t \geq 0$ is a solution, as is $\underline{u}(t) = P(S \geq t), t \geq 0$. Moreover, $u = \underline{u}$ is the unique global solution if and only if $\alpha \leq 1$.

Proof. It is clear by inspection that $u \equiv 1$ is a solution, and $\underline{u}$ is the minimal solution. The corollary is resolved by showing that $P(S \geq t) < 1$ if and only if $\alpha > 1$. In fact, let $L = \sup_{w\in\{1,2\}^{\mathbb{N}}} \sum_{k=0}^{\infty} \frac{T_{w|k}}{\alpha^k} = \lim_{n\to\infty} \max_{|v|=n} \sum_{k=0}^{n} \frac{T_{v|k}}{\alpha^k}$ be the length of the longest path. Then, $S \leq L$ a.s. since S is the length of the shortest path. Moreover, for $0 \leq \alpha \leq 1$, $L = S = \infty$ a.s. since for any path $\mathbf{s} \in \{1, 2\}^\infty$, with probability one,

$$\sum_{j=0}^{\infty} \alpha^{-j} T_{\mathbf{s}|j} \geq \sum_{j=0}^{\infty} T_{\mathbf{s}|j} = \infty,$$

by, for example, the strong law of large numbers. On the other hand, for $\alpha > 1$, note that the sequence $L_n = \max_{|v|=n} \sum_{j=0}^{n} \alpha^{-j} T_{v|j}$, $n \geq 1$, may be bounded iteratively by

$$L_{n+1} \leq L_n + \Theta_{n+1} \leq T_\theta + \sum_{n=1}^{\infty} \Theta_n,$$

where $\Theta_n = \alpha^{-n} \max\{T_n^{(1)}, \ldots, T_n^{(2^n)}\}$, where $T_n^{(j)}$ are i.i.d. mean one exponential random variables. Fix a sequence θ_n, $n \geq 1$, to be determined, and consider

$$\begin{aligned}
\mathrm{Prob}(\Theta_n > \theta_n) &= 1 - \mathrm{Prob}(\Theta_n \leq \theta_n) \\
&= 1 - (1 - e^{-\theta_n \alpha^n})^{2^n} \\
&\leq e^{n \ln 2 - \theta_n \alpha^n} = e^{-n},
\end{aligned} \tag{28.35}$$

for $\theta_n = n(\ln 2 + 1)\alpha^{-n}$. Thus, using Borel–Cantelli lemma, one has with probability one $\Theta_n \leq \theta_n$ for all but finitely many n, and therefore $\sum_{n=1}^{\infty} \Theta_n < \infty$ a.s. Thus, $S \leq L < \infty$ a.s. follows since $L = \lim_{n\to\infty} L_n \leq T_\theta + \sum_{n=1}^{\infty} \Theta_n$. ■

The event $[L < \infty]$, referred to as *hyper-explosion*, is thus equivalent to explosion $[S < \infty]$. The non-uniqueness for small ($u_0 = 0$) initial data is related to the hyper-explosive property. Various other interesting properties of this model are explored in the exercises.

Exercises

1. Verify that $h(\xi) = \frac{1}{\pi^3 |\xi|^2}, \xi \neq 0$, and $h(\xi) = \frac{1}{2\pi|\xi|} e^{-|\xi|}, \xi \neq 0$, are standardized majorizing kernels.
2. This exercise assumes the majorizing kernel $h(\xi) = \frac{1}{\pi^3 |\xi|^2}, \quad \xi \neq 0$.

   (a) Show that $-\int_1^t \frac{\ln(1-u)}{u} du = \sum_{k=1}^{\infty} \frac{t^k}{k^2}$ for $|t| < 1$. The latter is broadly referred to as Euler's dilogarithm series, or Spence's function, denoted $Li_2(t)$.
   (b) Show that under the density of the magnitude, the ratio $\frac{|W_1|}{|\xi|}$ (or $\frac{|W_2|}{|\xi|}$) is given by $\frac{2}{\pi^2} \ln |\frac{1+r}{1-r}| \frac{dr}{r}$.
   (c) Show that the density in (b) is symmetrically distributed about the identity of the multiplicative group $(0, \infty)$.
   (d) Show that the angle between W_1 and W_2 is uniformly distributed on $\{(\theta_1, \theta_2, \theta_3) \in (0, \pi) \times (0, \pi) \times (0, 2\pi) : \theta_1 + \theta_2 < \pi\}$, where θ_1, θ_2 are co-latitudes, and θ_3 longitude, respectively.

3. Show that uniqueness of a global solution to the genealogical Navier–Stokes equations with initial data $|u_0|$ as an expected value over the genealogical cascade (without coin tossing) implies global uniqueness for incompressible Navier–Stokes equations for the initial data u_0 as an expected value over the Le Jan–Sznitman cascade without coin tossing.
4. Show that $Y(\tau_\theta^{(n)}(t,\xi)), n = 0, 1, 2, \ldots,$ is a martingale.
5. Show that the functions $h_\beta^\alpha, 0 \le \beta \le 2, \alpha > 0$, defined in (28.13), provide a parametric family of majorizing kernels linking the Bessel and dilogarithmic majorizing kernels.
6. (*Non-explosive Yule*) It was shown in Exercise 13 of Chapter 5 that the Yule branching process is non-explosive. Complete the following steps for an alternative technique to prove non-explosion for a Yule process with exponentially distributed lifetimes with intensity parameter $\lambda > 0$. Let $\zeta = \lim_{n\to\infty} \min_{|v|=n} \sum_{j=0}^n \frac{1}{\lambda} T_{v|j}$, where $T_u, u \in \cup_{n=0}^\infty \{1,2\}^n$ is a binary tree-indexed family of mean one exponential random variables.

   (i) Use Fatou's lemma to show for arbitrary $\theta > 0$,

   $$\mathbb{E}e^{-\theta\zeta} \le \liminf_n \mathbb{E}e^{\{-\min_{|v|=n}\sum_{j=0}^n \frac{\theta}{\lambda} T_{v|j}\}}.$$

   (ii) Show that $\mathbb{E}e^{-\theta\zeta} \le \liminf_n 2^n \mathbb{E}\prod_{j=0}^n e^{-\frac{\theta}{\lambda}T_{v|j}}$. [*Hint*: Bound minimal term by sum over $|v| = n$.]

   (iii) Show that by selecting $\theta > \lambda \ln 2$, $\mathbb{E}e^{-\theta\zeta} = 0$, and therefore $\zeta = \infty$ a.s.

7. (a) Show that the evolutionary set process[17] $V^{(\alpha)}(t), t \ge 0$, defined by (28.32) is a Markov process.
   (b) Show that the cardinality $|V^{(1)}(t)|$ has a geometric distribution in the case $\alpha = 1$, referred to as the Yule process.
   (c) Show that $|V^{(\frac{1}{2})}(t)|$ is a Poisson process in the case $\alpha = 1/2$. [Hint: Theorem 15.3]
8. (a) Define a genealogical gauge $a_\alpha(V) = \sum_{v\in V} \alpha^{|v|}$ for evolutionary sets $V \in \mathcal{E}$. In particular $a_1(V) \equiv |V|$. Show that $A(t) = e^{(2\alpha-1)t} a_\alpha(V(t)), t \ge 0$, is a positive martingale.
   (b) Show that A is uniformly integrable if and only if $\alpha \in (\alpha_c, 1]$, where α_c is the unique solution to $\alpha_c \ln \alpha_c = \alpha_c - \frac{1}{2}$ in $(0, 1]$. In particular $\alpha_c \approx 0.187$. [*Hint*: Use the Neveu–Chauvin inequality[18].]
   (c) In the case $\alpha = 1$, show that $\lim_{t\to\infty} e^t |V^{(1)}(t)| = A_1$ is exponentially distributed.[19]
9. Consider a modification of the α-Riccati model in which at each generation a pair of frequency values (α, β) is generated. For a path v, the holding time is

[17] This exercise is largely based on Dascaliuc et al. (2017).

[18] See BCPT, (p. 40).

[19] This result can be traced back to Kendall (1966).

exponential with frequency $\alpha^{l(v)}\beta^{r(v)}$, where $l(v) = |v| - r(v) = |\{j \le |v| : v_j = 1\}$. Show that the tree is (i) non-explosive if $0 < a, b \le 1$, (ii) explosive but not hyper-explosive if $0 < a \le 1 < b$, and (iii) hyper-explosive if $a, b > 1$.

10. (*Fisher-KPP Equation*) The Fisher–KPP equation is the one-dimensional scalar equation

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial t^2} + v - v^2, \quad t > 0, v(0,x) = v_0(x),\ x \in \mathbb{R}.$$

(a) Show that v solves Fisher–KPP if and only if $v = 1 - u$ solves $\frac{\partial u}{\partial t} = \frac{\partial^2 v}{\partial x^2} + u^2 - u$.

(b) Show[20] for the Fourier transform of $\hat{u}(t, \xi)$, satisfying the equivalent equation (a), that for a positive function h such that $h * h(\xi) = (1 + \xi^2)h(\xi)$, $\xi \in \mathbb{R}$, one has that $\chi(t, \xi) = \frac{\hat{u}(t,\xi)}{h(\xi)}$ satisfies

$$\begin{aligned}\chi(t, \xi) \\ &= e^{(1+\xi^2)t}\chi_0(\xi) + \int_0^t \int_{\mathbb{R}} (1 + \xi^2)e^{-(1+\xi^2)s}\chi(t-s, \eta)\chi(t-s, \xi - \eta)\frac{h(\eta)h(\xi - \eta)}{(1+\xi^2)h(\xi)}d\eta ds.\end{aligned}$$

(c) Show the existence of a suitable function h. [*Hint*: Consider the ordinary differential equation governing the inverse Fourier transform of h.]

[20] For a development of this theory from the perspective of stochastic explosion, see Dascaliuc et al. (2021a) and Dascaliuc et al. (2021b).

References

Agresti A (1975) On the extinction times of varying and random environment branching processes. J Appld Probab 12(1):39–46.

Aidékon E, Shi Z (2014) The Seneta–Heyde scaling for the branching random walk. Ann Probab 42:959–993.

Albrecher H, Constantinescu C, Thomann E (2012) Asymptotic results for renewal risk models with risky investments. Stoch Proc Their Appl 122(11):3767–3789.

Aldous D (1991) The continuum random tree I. Ann Probab 19(1):1–28.

Aldous D (1993) The continuum random tree III. Ann Prob 21(1):248–289.

Arratia R, Goldstein LM, Kochman F (2019) Size bias for one and all. Probab Surv 16:1–61.

Athreya KB (1985) Discounted branching random walks. Adv Appld Prob 17:53–66.

Barral J, Jin X (2014) On exact scaling log-infinitely divisible cascades. Prob Thry Rel Fields 160:521–565.

Barral J, Rhodes R, Vargas V (2012) Limiting laws of supercritical branching random walks. Compt Rend Math 350(9):535–538.

Barral J, Kupiainen A, Nikula M (2014) Critical Mandelbrot cascades. Commun Math Phys 325:685–711.

Benzi R, Biferale L, Ciliberto S, Struglia MV, Tripiccione R (1996) Generalized scaling in fully developed turbulence. Physica D 96:162.

Bhattacharya R, Waymire E (2016) A basic course in probability theory. Springer universitext series, 2nd edn. Springer, Berlin. (ERRATA: http://sites.science.oregonstate.edu/~waymire/)

Bhattacharya R, Gupta VK, Waymire E (1983) The Hurst effect under trends. J. Appl Probab 20(3):649–662.

Bhattacharya R, Chen L, Dobson S, Guenther R, Orum C, Ossiander M, Thomann E, Waymire E (2003) Majorizing kernels and stochastic cascades with applications to incompressible Navier–Stokes equations. Trans Amer Math Soc 355(12):5003–5040.

Biggins JD (1976) The first- and last-birth problems for a multitype age-dependent branching process. J Appl Probab 8(3):446–459.

Biggins JD (1977) Martingale convergence in the branching random walk. J Appl Prob 14:25–37.

Biggins JD, Kyprianou AE (2004) Measure change in multitype branching. Adv Appl Probab 36(2):544–581.

Biggins J (2010) Branching Out, Probability in Honor of Sir John Kingman. Cambridge University Press, 112–133.

Billingsley P (1968) Convergence of probability measures. Wiley, New York.

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8

Billingsley P (1983) The singular function of bold play: Gambling theory provides natural examples of a function with anomalous differentiability properties. Amer Sci 71(4):392–397.
Black F, Scholes M (1973) The pricing of options and corporate liabilities. J Political Economy 81:637–654.
Blackwell D (1948) A renewal theorem. Duke Math J 15:145–151.
Blackwell D (1953) Extension of a renewal theorem. Pacific J Math 3:315–320.
Blumenthal R (1957) An extended Markov property. Trans Amer Math Soc 85:52–72.
Bolthausen E (1989) A note on the diffusion of directed polymers in a random environment. Commun Math Phys 123(4):529–534.
Bolthausen E (1991) On directed polymers in a random environment. Lect Notes Monogr Ser 18:41–47.
Bradley RC (2007) Introduction to strong mixing conditions, vol 1–3. Kendrick Press, Heber City.
Brémaud P (1981) Point processes and queues: martingale dynamics. Springer, Berlin.
Budaev VP (2008) Generalized scale invariance and log-Poisson statistics for turbulence in the scrape-off-layer plasma in the T-10 Tokamak. Plasma Phys Reprts 34(10):799–814.
Burd G, Waymire E (2000) Independent random cascades on Galton-Watson trees. Proc Amer Math Soc 128(9):2753–2761.
Chavarria GR, Baudet C, Benzi R, Ciliberto S (1995) Hierarchy of the velocity structure functions in fully developed turbulence. J Phys II 5:485.
Chen X (2015) A necessary and sufficient condition for the nontrivial limit of the derivative martingale in a branching random walk. Adv Appl Prob 47:741–760.
Chentsov NN (1956) Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the 'heuristic' approach to the Kolmogorov-Smirnov tests. Theor Probab Appl 1:140–144.
Christensen S (2012) Generalized Fibonacci numbers and Blackwell's renewal theorem. Stat Prob Lett 82(9):1665–1668.
Chung KL (1976) Excursions in Brownian motion. Ark Mat 14(1–2):155–177.
Cox JC, Ross SA, Rubinstein M (1979) Option pricing: A simplified approach. J Financial Econ. 7(3):229–263.
Dascaliuc R, Michalowski N, Thomann E, Waymire E (2015) Symmetry breaking and uniqueness for the incompressible Navier–Stokes equations. Chaos An Interdiscip J Nonlinear Sci 25(7):16. https://doi.org/10.1063/1.4913236
Dascaliuc R, Michalowski M, Thomann E, Waymire E (2017) A delayed Yule process. Proc Amer Math Soc 146:1335–1346.
Dascaliuc R, Thomann E, Waymire E (2018) Stochastic explosion and non-uniqueness for α-Riccati equation. J Math Anal Appl 476(1):53–85.
Dascaliuc R, Michalowski N, Thomann E, Waymire E (2019) Complex burgers equation: A probabilistic perspective. In: Sidoravicius V (eds) Sojourns in probability theory and statistical physics–I. Springer proceedings in mathematics & statistics, vol 298. Springer, Singapore, pp 138–170.
Dascaliuc R, Pham T, Thomann E, Waymire E (2021a) Doubly stochastic yule cascades part 1: The explosion problem in the time-reversible case arXiv:2103.06912.
Dascaliuc R, Pham T, Thomann E, Waymire E (2021b) doubly stochastic yule cascades part 2: the explosion problem in the non-reversible case. arXiv:2107.13182.
Dey P, Waymire E (2015) On normalized multiplicative cascades under strong disorder. Elect J Probab 20(32):1–13.
Doeblin W (1938) Exposé de la Théorie des Chaînes simples constants de Markoff á un nombre fini d'États. Revue Math de l'Union Interbalkanique 2:77–105.
Doob JL (1953) Stochastic processes. Wiley, New York.
Dubrulle B (1994) Intermittency in fully developed turbulence: log-Poisson statistics and generalized scale covariance. Phys Rev Lett 73:959.
Durrett R, Liggett TM (1983) Fixed points of the smoothing transformation. Z Wahr verw Gebiete 54:275–301.

Dvoretsky A, Erdos P (1951) Some problems on random walk in space. In: Proceedings of the second Berkeley symposium on mathematical statistics and probability, pp 353–367.
Dynkin EB, Yushekivic A (1956) Strong Markov processes. Teor Ver 1:149–155.
Fefferman CL (2006) Existence and smoothness of the Navier–Stokes equation. In: Millennium prize problems. Clay Mathematics Institute, Cambridge, pp 57–67.
Feller W (1951) The asymptotic distribution of the range of sums of independent random variables. Ann Math Statist 22:427–432.
Feller W (1968, vol 1, 3rd edn/1971, vol 2, 2nd edn) An introduction to probability theory and its applications. Wiley, New York.
Ferguson T (1989) Who solved the secretary problem? Statist Sci 4(3):282–289.
Fields E (2020) Poisson processes and linear programs for allocating shared vehicles. Notices AMS 67(11):1804–1805.
Fisher RA (1937) The wave of advance of advantageous genes. Ann Eugenics 7(4):353–369.
Frisch U (1991) From global scaling, á la Kolmogorov, to local multifractal scaling in fully developed turbulence. Proc B Soc Lond A 434:89–99.
Gangolli R (1967) Positive definite kernels on homogeneous spaces and certain stochastic processes related to Levy's Brownian motion of several parameters. Ann Inst Henri Poincare 3(B):121–225.
Gantert N, Höfelsauer T (2019) Large deviations for the maximum of a branching random walk. arXiv:1802.03960v5.
Gikhman II, Skorokhod AV (1969) An introduction to the theory of random processes. W. B. Saunders, Philadelphia.
Guivarc'h Y (1990) Sur une extension de la notion de loi semi-stable. Ann Inst H Poincaré Probab Statist 26(2):261–285.
Gupta VK, Waymire E (1993) A statistical analysis of mesoscale rainfall as a random cascade. J Amer Meteor Soc 32:252–267.
Hammersley JM (1974) Postulates for sub-additive processes. Ann Prob 2:652–680.
Hardy GH, Wright EM (1938) An introduction to the theory of numbers 1st edn. Clarendon Press, Oxford.
Hartman P, Wintner A (1941) On the law of the iterated logarithm. Amer J Math 63:169–176.
Hofstad van der R, Keane M (2008) An elementary proof of the hitting time theorem. Amer Math Monthly 115(8):753–756.
Holley R, Liggett TM (1981) Generalized potlatch and smoothing processes. Z Wahr verw Gebiete 55:165–195.
Hopf E (1951) Uber die Anfangswertaufgabe f ur die hydrodynamischen Grundgleichungen. Math Nachr 4:213–231.
Hurst HE (1951) Long term storage capacity of reservoirs. Trans Amer Soc Civil Eng 116:770–779.
Hutchinson J (1981) Fractals and self-similarity. Indiana U Math J 30:713–747.
Itô K, McKean HP (1963) Brownian motions on the half-line. Ill J Math 7:181–231.
Janson S (2007) Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas. Probab Surv 4:80–145.
Johnson T, Waymire E (2011) Tree polymers in the infinite volume limit at critical strong disorder. J Appl Probab 48(3):885–891.
Kahane JP (1974) Sur le modéle de turbulence de Benoit Mandelbrot. C R Acad Sci 278:621–623.
Kahane J-P (1989) Random multiplications, random coverings, and multiplicative chaos. In: Berkson E, Tenny Peck N, Jerry Uhl J (eds) Proceedings of the special year in modern analysis. London mathematical society lecture note series, vol 137. Cambridge U Press, Cambridge, pp 196–255.
Kahane JP, Peyrère J (1976) Sur certaines martingales de Benoit Mandelbrot. Adv Math 22(2):131–145.
Kemperman JHB (1950) The general one-dimensional random walk with absorbing barriers. Thesis, Excelsior, The Hague.
Kendall D (1966) Branching processes since 1873. J London Math Soc1(1):385–406.

Kesten H, Ney P, Spitzer F (1966) The Galton-Watson process with mean one and finite variance. Theory Probab Appl XI(4):513–540.

Khinchine A (1924) Uber einen Satz der Wahrschinlichkeitsrechnung. FundamentaMath 6:9–20.

Khintchine A (1933) Der Satz vom iterierten logarithmus. In: Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik und Ihrer Grenzgebiete, vol 4. Springer, Berlin.

Kingman JFC (1975) The first birth problem for an age-dependent branching process. Ann Prob 3:790–801.

Kolmogoroff AN (1929) Uber der das iterierten Logarithmus. Math Ann 101:126–135.

Kolmogorov AN (1938) Zur l osung einer biologischen aufgabe. Comm Math Mech Chebyshev Univ Tomsk 2:1–12.

Kolmogorov AN (1941) The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. Dokl. Akad. Nauk SSSR 30:301.

Kolmogorov AN (1956) On skorokhod convergence. Theory Probab Appl 1(2):215–222.

Kolmogorov AN (1962) A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. J Fluid Mech 13:82.

Kolmogorov A, Petrovskii I, Piskunov N (1937) A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. In: Tikhomirov VM (ed) Selected works of AN Kolmogorov I. Kluwer, Alphen aan den Rijn, 1991, pp 248–270 (trans. Volosov VM from Bull. Moscow Univ., Math. Mech. 1, pp 1–25).

Konstantopoulos T (1995) Ballot theorems revisited. Statist Probab Lett 24:331–338.

Kovchegov Y, Zaliapin I (2020) Random self-similar trees: A mathematical theory of Horton laws. Probab Surveys, Institute of Math Stat 17:1–213.

Kyprianou AE (1998) Slow variation and uniqueness of solutions to the functional equation in the branching random walk. J Appl Probab 35(4):795–801.

Ladyzhenskaya OA (2003) Sixth problem of the millennium: Navier–Stokes equations, existence and smoothness 2003. Russian Math Surv 58(2):25–286.

Lalley SP (1988) The packing and covering functions of some self-similar fractals. Indiana U Math J 37(3):699–709.

Lalley S, Selker T (1987) A conditional limit theorem for the frontier of a branching Brownian motion. Ann Probab 15(3):1052–1061.

Le Cam L (1960a) An approximation theorem for the Poisson binomial distribution. Pacific J Math 10:1181–1197.

Le Cam L (1960b) A stochastic description of precipitation. In: Neyman J (ed)Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, vol III, pp165–186.

Le Gall J-F (2005) Random trees and applications. Probab Surv 2:245–311.

Le Jan Y, Sznitman AS (1997) Stochastic cascades and 3-dimensional Navier–Stokes equations. Prob Theory Rel Fields 109:343–366.

Leray J (1933) Étude de diverses équations intégrales non linéaires et de quelques problémes que pose l'hydrodynamique. J Math Pures Appl 12(9):1–82.

Leray J (1934) Sur le mouvement d'un liquide visqueux emplissant l'espace. Acta Math 63:193–248.

Lévy P (1937) Théorie de l'addition des variables aléatores. Gauthier-Villars, Paris.

Lévy P (1939) Sur certains processus stochastiques homogènes". Compositio Math 7:283–339.

Lévy P (1945) Sur le mouvement Brownien dependant de plusieurs parametres. C R Acad Sei Paris 220:420–422.

Lindvall T (1992) Lectures on the coupling method. Wiley, New York.

Lyons R (1997) A simple path to Biggins' martingale convergence for branching random walk. In: Athreya K, Jagers P (eds) Classical and modern branching processes. Springer, New York, pp 217–221.

Lyons R, Peres Y (2016) Probability on trees and networks. Cambridge series in statistical and probabilistic mathematics, vol 42. Cambridge University Press, New York.

Lyons R, Pemantle R, Peres Y (1995) Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. Ann Probab 23(3):1125–1138.

Mandelbrot B (1974a) Intermittent turbulence in self similar cascades: divergence of high moments and dimension of the carrier. J Fluid Mech 62:331–358.

Mandelbrot B (1974b) Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire: quelques extensions. C R Acad Sci Paris 278:355–358.

Mandelbrot B, Wallis J (1968) Noah, Joseph, and operational hydrology. Water Resources Res 4:909–918.

Miles Jr, EP (1960) Generalized Fibonacci numbers and associated matrices. Amer Math Monthly 67:745–752.

Molchan GM (1997) Turbulent cascades: limitations and a statistical test of the lognormal hypothesis. Phys. Fluids9:2387.

Moran PAP (1964) On the range of cumulative sums. Ann Inst Statist Math 16:109–112.

Neveu J (1965) Mathematical foundations of the calculus of probability. San Francisco, Holden-Day.

Neveu J, Pitman J (1989) Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion. In: Seminaire de probabilites, vol XXIII. Springer, Berlin, pp 239–247.

Newburgh R, Peidle J, Rueckner W (2006) Einstein, Perrin, and the reality of atoms: 1905 revisited. Am J Phys 74:478–481.

Obukhov A (1962) Some specific features of atmospheric turbulence. J Fluid Mech 13:77.

Ossiander M, Waymire E (1989) Certain positive-definite kernels. Proc Amer Math Soc 107(2):487–492.

Ossiander M, Waymire E (2002) On estimation theory for multiplicative cascades. Sankhyā Indian J Statist 64 Series A, Pt.2:323–343.

Politano H, Pouquet A (1995) Model of intermittency in magnetohydrodynamic turbulence. Phys Rev E 52:636.

Perrin J (1913) Les Atomes, (French) JFM 44.0913.12 Paris: Alcan. XVI + 296 S. 16mo.

Perrin J (1929) Atoms, Ox Bow Press, (English translation), reprinted 1990, 295 pages.

Peyrière J (1974) Turbulence et dimension de Hausdorff. C R Acad Sci Paris 278:567–569.

Peyriére J (1977) Calculus de dimensions de Hausdorff. Duke Math J 44:591–601.

Pitman J (1997) Enumerations of trees and forests related to branching processes and random walks. Technical preport No. 482, Department of Statistics, University of California, Berkeley, pp 1–25.

Pitman J (2002) Combinatorial Stochastic Processes, Ecole d'Eté de Probabilitiés de Saint-Flour XXXII-2002, Springer Lecture notes No. 1875. Springer, Berlin.

Ramasubramanian S (2009) Lectures on insurance models. American Mathematical Society, Providence.

Ramirez JM (2006) Multiplicative cascades applied to PDEs (two numerical examples). J Comp Phy 214(1):122–136.

Reed MC, Simon B (1972) Functional analysis, vol 1. Methods of mathematical physics. Academic Press, New York.

Richardson LF (1922) Weather prediction by numerical process. Cambridge University Press, Cambridge.

Rogers LCG (1997) Arbitrage with fractional Brownian motion. Math Fin 7(1):95–105.

Rolski T, Schmidli H, Schmidt V, Teugels JL (1999) Stochastic processes for insurance and finance. Wiley, Chester.

Salas HD, Poveda G, Mesa O (2017) Testing the beta-lognormal model in amazonian rainfall fields using the generalized space q-entropy. Entropy Appl Environ Water Eng 19(12):685.

Samorodnitsky G (2006) Long range dependence. Foundations Trends Stoch. Syst 1(3):163–257.

Schoenberg IJ (1938) Metric spaces and positive definite functions. Trans Amer Math Soc 44:522–536.

She Z-S, Levesque (1994): Universal scaling laws in fully developed turbulence. Phys Rev Lett 72:336.

She Z, Waymire E (1995) Phys Rev Lett 74(2):262–265.

Shi Z (2012) Branching random walks: Ecole D'Ete de probabilites de saint-flour XLII. Springer, New York.
Sirao T, Nisida T (1952) On some asymptotic properties concerning Brownian motion. Nagoya Math J 4:97–101.
Skorokhod A (1956) Teoria, vol 1 (trans. Theory of probability and applications, it appears as "Limit Theorems for Stochastic Processes"), vol 1, pp 261–290.
Skorokhod's AV (1965) Studies in the theory of Markov processes, (English trans 1961 Russian edition).
Slutsky EE (1937) Qualche proposizione relativa alla teoria delle funzioni aleatorie. Giorn Ist Ital Attuari 8:183–199.
Spitzer F (1956) A combinatorial lemma and its application to probability theory. Trans Amer Math Soc 82(2):323–339.
Spitzer F (1976) Principles of random walk, 2nd edn. Springer, New York.
Strassen V (1964) An invariance principle for the law of the iterated logarithm. Z. Wahrscheinlichkeits. Ver Geb 3:211–226.
Stroock DW, Varadhan SRS (1980) Multidimensional diffusion processes. Springer, Berlin.
Sun Y, Stein ML (2015) A stochastic space-time model for intermittent precipitation occurrences. Ann Appl Statist 9(4):2110–2132.
Watanabe S (1964) On discontinuous additive functionals and Levy measures of a Markov process. Japanese J Math 34:53–70.
Waymire E, Williams S (1994) A general decomposition theory for random cascades. Bull Amer Math Soc 31:216–222.
Waymire E, Williams S (1995) Multiplicative cascades: dimension spectra and dependence. In: The journal of fourier analysis and applications Kahane, Special Issue. CRC Press, Boca Raton, pp 589–609.
Waymire E, Williams S (1996) A cascade decomposition theory with applications to Markov and exchangeable cascades. Trans Amer Math Soc 348:585–632.
Yaglom AM (1947) Certain limit theorems of the theory of branching random processes, (Russian). Doklady Akad Nauk SSSR (NS) 56:795–798.
Zhao W, Su W, Wang G (2021) Interactions of velocity structures between large and small scales in micro-electrokinetic turbulence. Phys Rev Fluids 6:074610

Related Textbooks and Monographs

Athreya KB, Ney PE (1972) Branching processes. Springer, New York.
Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge.
Bhattacharya R, Ranga Rao R (2010) Normal approximation and asymptotic expansions. SIAM classics in applied mathematics series. Originally published by Wiley, Philadelphia, (1976).
Bhattacharya R, Waymire E (2009) Stochastic processes with applications. SIAm classics in applied mathematics series. Originally published by Wiley, Philadelphia (1990).
Bhattacharya R, Waymire E (2007, 2016) A basic course in probability theory. Springer universitext series, 1st, 2nd edn. Springer International Publishing, Berlin. (ERRATA: http://sites.science.oregonstate.edu/~waymire/)
Billingsley P (1968) Convergence of probability measures. Wiley, New York.
Billingsley P (1986) Probability and measure, 2nd edn. Wiley, New York.
Bradley RC (2007) Introduction to strong mixing conditions, vols 1–3, Kendrick Press, Heber City.
Breiman L (1968, 1992) Probability, Reprinted in SIAM classics in applied mathematics series, Philadelphia, PA, from Addison Wesley, Reading (1968).
Brémaud P (1981) Point processes and queues: martingale dynamics. Springer, Berlin.
Chung KL (1974) A course in probability theory, 2nd edn. Academic, New York.

Davis M, Etheridge A (2006) Louis Bachelier's theory of specultation: the origins of modern finance. Princeton University Press, Princeton.

Doob JL (1953) Stochastic processes. Wiley, New York.

Durrett R (1984) Brownian motion and martingales in analysis. Wadsworth, Belmont.

Durrett R (1995) Probability theory and examples, 2nd edn. Wadsworth, Brooks & Cole, Pacific, Grove.

Ethier SN, Kurtz TG (1985) Markov processes: characterization and convergence. Wiley, New York.

Feller W (1968, , vol 1, 3rd edn/1971, vol 2, 2nd edn) An introduction to probability theory and its applications. Wiley, New York.

Foelmer H, Schied A (2002) Stochastic finance: an introduction in discrete time. Studies in mathematics, vol 27. Walter de Gruyter, Berlin.

Folland G (1984) Real analysis. Wiley, New York.

Fürth R (1954) Investigations on the theory of the brownian movement. Dover, Downers Grove.

Grimmett G (1999) Percolation, 2nd edn. Springer, Berlin.

Gikhman II, Skorokhod AV (1969) An introduction to the theory of random processes. W.B. Saunders, Philadelphia.

Hardy GH, Wright EM (1938) An introduction to the theory of numbers, 1st edn. Clarendon Press, Oxford.

Harris TE (1963) The theory of branching processes. Springer, Berlin.

Itô K, McKean HP (1965) Diffusion processes and their sample paths. Reprint Springer, Berlin.

Itô K, McKean HP (1974) Diffusion processes and their sample paths, 2nd printing. Die Grundlehren der mathematischen Wissenschaften, 125.

Jacod J, Protter P (2003) Probability essentials. Springer universitext series, 2nd edn. Springer, New York.

Kac M (1979) In: Baclawski K, Donsker MD (eds) Probability, number theory and statistical physics: selected papers. MIT Press, Cambridge.

Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York.

Kallenberg O (2001) Foundations of modern probability, 2nd edn. Springer, New York.

Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic, New York.

Karlin S, Taylor HM (1981) A second course in stochastic processes. Academic, New York.

Kesten H (1982) Percolation theory for mathematicians. Birkhauser, Basel.

Kolmogorov AN (1950) Foundations of the theory of probability. Chelsea, New York (English trans 1933 German original).

Koshnevisan D (2007) Probability. American Mathematical Society, Providence.

Lamperti J (1996) Probability: a survey of the mathematical theory. Wiley, New York.

Landau LD, Lifshitz EM (1987) Fluid mechanics, course of theoretical physics, vol 6, 2nd revised edn. Pergamon Press, Oxford.

Laplace PS (1812) Théorie analytique des probabilités. Reprinted in Oeuvres Complète de Laplace, vol 7. Gauthier-Villars, Paris.

Lawler G, Limic V (2010) Random walk: a modern introduction. Cambridge studies in advanced mathematics, no 123. Cambridge University Press, Cambridge.

Lemarié-Rieusset PG (2002) Recent developments in the Navier-Stokes problem, Chapman & Hall/CRC Research Notes in Mathematics, vol 431. Chapman & Hall/CRC, Boca Raton, FL. xiv+395p

Lévy P (1925) Calcul des Probabilités. Gauthier–Villars, Paris.

Lévy P (1937) Théorie de l'addition des variables aléatores. Gauthier-Villars, Paris.

Lévy P (1965) Processus stochastiques et mouvement brownien, 2nd edn. Gauthier-Villars, Paris (1st edn. 1948).

Liggett TM (1985) Interacting particle systems. Springer, New York.

Liggett TM (2009) Continuous time markov processes: an introduction. American Mathematical Society, Providence.

Lindvall T (1992) Lectures on the coupling method. Wiley, New York.

Lyons R, Peres Y (2016) Probability on trees and networks. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge.
Mitzenmacher M, Upfal E (2005) Probability and computing (randomized routing probabilistic analysis). Cambridge University Press, Cambridge.
Nelson E (1959) Regular probability measures on function space. Ann Math 630–643.
Nelson E (1967) Dynamical theories of brownian motion. Princeton University Press, Princeton.
Neveu J (1971) Mathematical foundations of the calculus of probability. Holden-Day, San Francisco.
Neveu J (1975) Discrete parameter martingales. North-Holland, Amsterdam.
Parthasarathy KR (1967) Probability measures on metric spaces. Academic, New York.
Pitman J (1993) Probability. Springer, New York.
Perrin J (1929) Atoms. Ox Bow Press, Woodbridge (English translation), Reprinted 1990, 295 pages.
Protter P (1990) Stochastic integration and differential equations. Springer, Berlin.
Ramasubramanian S (2009) Lectures on insurance models. American Mathematical Society, Providence.
Reed MC, Simon B (1972) Functional analysis, vol 1. Methods of mathematical physics. Academic, New York.
Resnick S (1987) Extreme values, regular variation, and point processes. Springer, New York.
Revuz D, Yor M (1999) Continuous martingales and brownian motion, 3rd edn. Springer, Berlin.
Rogers LCG, Williams D (2000) Diffusions, Markov processes and martingales, vol 1, 2nd edn., vol 2. Cambridge University Press, Cambridge.
Rolski T, Schmidli H, Schmidt V, Teugels JL (1999) Stochastic processes for insurance and finance. Wiley, Chester.
Royden HL (1988) Real analysis, 3rd edn. MacMillan, New York.
Rudin W (1967) Fourier analysis on groups, 2nd edn. Wiley, New York.
Samorodnitsky G, Taqqu M (1994) Stable non-Gaussian Random processes: Stochastic models with infinite variance. Chapman and Hall, New York.
Samorodnitsky G, Taqqu M (2016) Stochastic processes and long range dependence. Springer series in operations research and financial engineering. Springer, New York.
Shi Z (2012) Branching random walks: Ecole D'Ete de probabilites de saint-flour XLII. Springer, New York.
Shiryaev AN (1990) Probability. Springer, New York.
Shreve S (2004) Stochastic calculus for finance. Springer, New York.
Skorokhod's AV (1965) Studies in the theory of Markov processes (English trans 1961 Russian edition).
Spitzer F (1976) Principles of random walk, 2nd edn. Springer, New York.
Stroock DW (1993) Probability theory: an analytic view. Cambridge University Press, Cambridge.
Stroock DW, Varadhan SRS (1980) Multidimensional diffusion processes. Springer, Berlin.
Tao T (2012) Topics in random matrix theory. Graduate studies in mathematics, vol 132. American Mathematical Society, Providence.
Walsh JB (2012) Knowing the odds: an introduction to probability. American Mathematical Society, Providence.
Wax N (1954) Selected papers on noise and stochastic processes. Dover, Downers Grove.
Williams D (1991) Probability with martingales. Cambridge University Press, Cambridge.

Author Index

A
Agresti, 16
Aidékon, 256
Alexandrov, 203, 240
Athreya, 373

B
Barral, 251, 252
Baudet, 245
Benzi, 245
Bhattacharya, 3, 8, 66, 75, 89, 94, 110, 186, 202, 243, 310, 329, 350, 361, 365, 367, 370
Biferale, 245
Biggins, 252, 261
Billingsley, 75
Black, 280
Blackwell, 311, 341
Blumenthal, 73
Boldyrev, 245
Bradley, 358
Bremaud, 185
Burd, 246

C
Chauvin, 377
Chavarria, 245
Chen, 245, 253, 367, 370
Chentsov, 61
Christensen, 111
Chung, 251
Ciliberto, 245
Cox, 280

D
Dascaliuc, 372, 373, 377, 378
Dey, 252
Dobson, 367, 370
Doeblin, 99
Doob, 14
Dubrulle, 244
Durrett, 253
Dynkin, 73

F
Fefferman, 363
Feller, 58, 188, 198, 216, 222, 312, 322, 342, 347
Ferguson, 294
Fields, 58
Fisher, 253
Foelmer, 279
Fuchs, 251

G
Gantert, 262
Gikhman, 8
Guenther, 367, 370
Guivarc'h, 251
Gupta, 244, 350

H
Hardy, 312
Hartman, 210
Höfelsauer, 262

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8

Hofstad van der, 37
Holley, 253
Hopf, 364, 365
Hostad van der, 35
Hurst, 347
Hutchinson, 325

J
Jin, 251
Johnson, 252, 256

K
Kahane, 244, 248
Keane, 35, 37
Kemperman, 36
Kendall, 377
Kesten, 119
Khinchine, 89, 210
Kolmogorov, 75, 117, 210, 244, 253
Konstantopoulos, 37
Kupiainen, 252
Kyprianou, 252

L
Lévy, 97, 359
Ladyzhenskaya, 363–365
Lalley, 325, 326
Landau, 244, 363
Lawler, 45
LeGall, 268
Leray, 364, 365
Levesque, 244
Lévy, 218
Le Cam, 47, 185
Le Gall, 268, 274
Le Jan, 363, 365
Lifshitz, 363
Liggett, 253
Limic, 45
Lindvall, 110
Lyons, 174, 244

M
Mandelbrot, 244, 347, 356
Mann, 203
Mesa, 244
Michalowski, 372, 377
Miles, 111
Molchin, 245
Moran, 350

N
Nelson, 3, 222
Neveu, 4, 377
Newburgh, 67
Ney, 119
Nikula, 252
Nisida, 96

O
Obukhov, 244
Orum, 367, 370
Ossiander, 245, 367, 370

P
Peidle, 67
Pemantle, 174, 244
Peres, 174, 244
Perrin, 66
Petrovskii, 253
Peyriére, 244, 246
Pham, 378
Piskunov, 253
Politano, 245
Pouquet, 245
Poveda, 244

R
Ramasubramanian, 322, 330, 331, 335, 336
Ramirez, 372
Reed, 369
Richardson, 244
Rogers, 347
Rolski, 345
Ross, 280
Rubenstein, 280
Rueckner, 67

S
Salas, 244
Samorodnitsky, 358
Scheffé, 192, 213
Schied, 279
Schmidli, 345
Schmidt, 345
Schoenberg, 359
Scholes, 280
Seneta, 183
She, Z.S., 244
Shi, 256, 259
Simon, 369

Sirao, 96
Skorokhod, 8, 75, 203
Slutsky, 61
Spitzer, 38, 119
Stein, 185
Strassen, 90
Stroock, 78
Struglia, 245
Su, 245
Sun, 185
Sznitman, 363, 365

T
Taqqu, 358
Thomann, 367, 370, 372, 373, 377, 378

V
Varadhan, 78

W
Wald, 203
Wallis, 347, 356
Wang, 245
Watanabe, 185
Waymire, 3, 8, 66, 75, 89, 94, 110, 186, 202, 243–246, 248, 252, 256, 310, 329, 350, 361, 365, 372, 373, 377, 378
Williams, 244, 246, 248
Wintner, 210
Wright, 312

Y
Yaglom, 117
Yushekivic, 73

Z
Zhao, 245

Subject Index

Symbols
L^p-maximal inequality, 130
ε–optimal stopping time, 299
$\{\mathcal{F}_t : t \in T\}$-submartingale, 124

A
Accessible state, 101
Adjustment coefficient, 145
Admissible strategy, 284
After-s process, 76
After-τ process, 78
Alpha-Ricatti equation, 372
American call option, 288, 303
American option, 297
American put option, 298, 303
Arbitrage free, 284, 286
Arcsine law for argmax on [0, 1] for Brownian motion, 224
Arcsine law for Brownian motion last zero, 219
Arcsine law for last zero for Brownian motion, 92
Arcsine law for occupation time for Brownian motion, 222
Arcsine laws for Brownian motion, 82
Ascending ladder heights, 341
Asymmetric simple random walk boundary distribution, 141
Attainable contingent claim, 284

B
Backward recursion, 291, 292
Bertrand's ballot theorem, 37
Bessel majorizing kernel, 368
Bienaymé–Galton–Watson Simple branching Process, 113
Bienaymé–Galton–Watson branching process, 12, 154
Binomial cascade, 245
Binomial tree model, 280
Blumenthal zero–one law, 93
Borel–Tanner distribution, 274
Boundary case for branching random walk, 253
Boundary value distribution
  Brownian motion, 85
Branching process, 12
  and simple random walk contour, 266
  family tree model, 175
Branching with geometric offspring, 118
Breakthrough curve, 191
Brownian bridge, 212, 235
Brownian bridge in non-parametric statistics, 212, 236
Brownian excursion, 269
Brownian meander, 269, 274
Brownian motion, 7, 8, 18
  boundary value distribution, 85, 86
  construction, 63
  extremes (max/min), 212
  first passage time process, 83
  hitting a two-point boundary, 167
  hitting time of a line, 196
  law of the iterated logarithm, 89
  nowhere differentiable, 67
  with one-point absorption, 227
  properties, 66

R. Bhattacharya, E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Graduate Texts in Mathematics 292, https://doi.org/10.1007/978-3-030-78939-8

Brownian motion (*cont.*)
with reflection at the origin, 229
transience under non-zero drift, 88
transition probabilities, 12
zero set, 91
Brownian sheet, 63, 64

C
Cadlag, 75
Cantor set, 325, 326
Change of measure, 156, 244
Closeable, 128
Closed martingale, 128
Coefficient of variation, 112, 328
Compensator, 190
Complete market, 286, 288
Complex martingale, 126
Compound Poisson process, 55
Conditional Poisson arrival times, 50
Contingent claim, 280
Continuous convolution property, 55
Continuous parameter Gaussian process, 7
Continuous parameter Markov process, 12
Continuous time random walk, 55
Continuous time simple random walk, 23
Contour path, 263
Contour path of branching process, 265
Contour walk, 267
Convergence of reverse martingales, 158
Counting process, 185
Counting process non-explosion criteria, 58
Coupled process, 102
Coupling, 47, 99, 103
inequality, 48
lemma, 103
Cox process, 190
Cox-Ross-Rubenstein model; mathematical finance, 280
Cramér–Lundberg model, 143, 337
Critical branching, 114
Critical branching process, 13
Critical strong disorder, 253

D
DeFinetti's theorem, 160
Delayed renewal process, 105
Derivative martingale, 253
Differential equations
Poisson process, 57
Dilogarithm, 376
Dilogarithmic majorizing kernel, 368
Directly Riemann integrable, 322
Discrete parameter Markov process, 9
Distinguished path, 178
Distinguished path analysis, 244
Distribution, 2
Distribution of area under Brownian motion, 212
Distribution of Brownian motion, 67
Distribution of maximum of Brownian bridge, 212
Divided differences, 58
Donsker's invariance principle, 205, 206
Doob maximal inequality, 128
Doob–Meyer decomposition, 127
Double or nothing strategy, 148
Doubly-stochastic Poisson process, 190
Downcrossing, 151
Dwass' formula, 268, 274

E
Elementary renewal theorem, 324
Empirical process, 237
Equivalent martingale measure (EMM), 285, 286
Error function, 96
Escape time distribution for Brownian motion, 84
Euler equations, 364
European call option, 280, 303
European put option, 289
Eventually decreasing, 335
Excess of loss policy, 145
Exchangeable martingale differences, 133
Exchangeable sequences, 159
Excursion interval, 269
Excursions of Brownian motion and branching process, 268
Expiration time, 280
Explosion, 375
Explosion in finite time, 58
Explosive, 52
Exponential martingale, 126, 132
Extinction, 155
Extinction of subcritical Bienaymé–Galton–Watson branching process, 14
Extinction probability as fixed point, 114
Extreme value distribution, 54

F
Fair game, 124
Feasible reinsurance policy, 145
Feller-Erdös-Pollard renewal theorem, 107

Feller property, 78
Filtration, 75, 124
  continuous, 170
Finite dimensional distribution, 2
Finite dimensional distributions
  Poisson process, 52
First departure bound, 250
First passage decomposition for Brownian motion, 84
First passage time distribution for simple symmetric random walk, 29
First passage time distribution for standard Brownian motion, 193
First passage time for Brownian motion, 194
First passage time Laplace transform, 83, 95
First passage time process, 81, 83
First return time, 22, 73
Fisher–KPP equation, 253, 378
Fourier transform of tempered distribution, 370
Fourier transform, 365
Fractals, 325
Fractional Brownian random field, 359
Frechet distribution, 54, 59
Functional central limit theorem, 202, 203, 205, 206
  invariance principle, 202
Functions of rapid decay, 370

G
Gambler's ruin, 24, 142
Gamma density, 49
Gamma distribution, 57
Gaussian process, 6
Gaussian random field, 6
Genealogical evolutionary process, 373
Genealogy of branching, 175
Generalized Fibonacci sequence, 111
General point process, 185
General random walk, 4
General renewal model, 336
Generalized ballot theorem, 38
Generation height, 175
Geometric random walk, 25, 282
Ghost edge, 263
Glivenko–Cantelli lemma, 239, 241
Gnedenko–Korolyuk formula, 241
Green's function of Markov chain, 101
Gumbel distribution, 54
Gumbel extreme value distribution, 54

H
Hölder continuity of sample paths for Brownian motion, 63
Harris-Ulam labeling, 175
Hazard function, 335
Heat equation, 65
Heavy-tailed claims, 344
Heavy-tailed distribution, 330
Hedging principle, 281
Hedging strategy, 282, 284
Historic probability, 280, 284
Holding time, 49
Homogeneous increments, 50
Homogeneous transition probabilities, 12
Hyperexplosion, 376

I
I.i.d random variables, 4
Incompressible fluid, 363
Independent coupling, 110
Independent increments, 50, 125
Index set, 123
Infinitely divisible distribution, 55
Infinitesimal probabilities
  Poisson process, 57
Inherited property of branching process, 155
Inhomogeneous Bienaymé–Galton–Watson and extinction, 15
Inhomogeneous Poisson process, 53
Inspection paradox, 112, 328
Integrated tail distribution, 335
Intensity parameter, 52
Inter-arrival time, 49
Interval recurrence, 23
Invariance principle, 202, 203, 206
Invariant probability, 102
Inverse Gaussian distribution, 191
Irreducible aperiodic Markov chain convergence, 104
Irreducible Markov chain, 99

J
Jean Perrin on Brownian motion, 66
Joint distribution of maximum and minimum for Brownian motion, 85

K
Karamata representation of slowly varying function, 330
Kesten–Stigum theorem, 174
Key renewal theorem, 321

non-lattice case, 322
Khinchine law of the iterated logarithm, 89
Kolmogorov backward and forward equations, 66
Kolmogorov–Chentsov theorem, 61
Kolmogorov consistency conditions, 3
Kolmogorov existence theorem, 3
Kolmogorov maximal inequality, 133
Kolmogorov–Smirnov statistic, 239
Kolmogorov's probability decay rate for branching processes, 117
Kolmogorov's upper class function test at the origin, 96
Kolmogorov–Yaglom–Kesten–Ney–Spitzer Theorem, 117

L

Ladder heights, 317
Ladder times, 317
Last visit to zero for simple symmetric random walk, 216
Lattice distribution, 307
Law of rare events, 47
Law of the iterated logarithm (LIL) for Brownian motion, 89
Law of the iterated logarithm for sums of i.i.d. random variables, 209
Lazy random walk, 22
Le Jan–Sznitman cascade, 365
Le Jan–Sznitman cascade without coin tossing, 365
Lebesgue decomposition, 156, 180
Left continuous filtration, 187
Left-skip free property, 37
Leray projection, 366
Lévy–Khinchine formula: special case, 56, 59
Lévy representation for reflecting Brownian motion, 230
Lévy-Skorohod formula for reflected Brownian motion, 82
Lifetimes, 105
Light-tailed claim size distribution, 144
Light-tailed distribution, 330, 336
Likelihood ratios, 126, 132
Line of descent, 179
Local clt for simple symmetric random walk, 197
Local limit theorem, 192
Lukasiewicz path, 268
Lundberg bound, 337
Lundberg constant, 337
Lundberg inequality, 144

M

Majorizing kernel, 367
Mann-Wald theorem, 203
Many-to-one formula, 261
Market probability, 280
Markov chain, 18
Markov chain return times, 100
Markov process, 9
Markov property (homogeneous), 71
Markov property for Brownian motion, 77
Markov property of Brownian motion, 77
Markov time, 72
Martingale, 14
characterization of Poisson process, 185
characterization of the Poisson process, 188
closed on the right, 128
difference sequence, 124, 125
gambling strategy, 148
measure, 285
quadratic variation, 127
Mathematical finance, 280
Maximal coupling, 110
Maximal moment inequality, 130
Method of images for simple random walk, 37
Method of images for simple symmetric random walk, 30
Modification of a stochastic process, 167
Monotone coupling, 110
Multidimensional Brownian motion, 65
Multi-dimensional simple symmetric random walk, 5
Multiparameter Brownian motion, 359
Multiplicative cascades, 244

N

Navier–Stokes equations, 363
Nonanticipative locally bounded step function, 132
Net profit condition (NPC), 144, 336
Non-explosive, 52
Non-lattice distribution, 307
Nowhere differentiability of Brownian motion, 67

O

Optimal gambling strategy, 148
Optimal stopping time, 291
Option; mathematical finance, 280
Optional sampling, 143
Optional sampling theorems, 142
Optional stopping theorem, 136, 165
Optional time, 164
Order statistic property (o.s.p.), 50

Otter-Dwass formulae, 267
  Otter formula, 268

P
Packing function, 326
Parameter set, 123
Partition function, 252
Path duality, 223
Path space, 2
Periodic Markov chain, 99
Permutation–invariant, 159
Pointwise standard recurrent Brownian motion, 87
Poisson
  renewal process, 308
  sprocess, 5
Poisson approximation, 56
  to binomial, 47, 48
  error bounds, 47
Poisson arrival time, 50
Poisson distribution, 47
Poisson process, 50, 53, 186
  age of last occurrence prior to t, 58
  distribution, 52
  finite dimensional distributions, 52
  homogenization, 55
  infinitesimal probabilities, 56, 57
  inhomogeneous, 53
  splitting, 57
  thinning, 57
  time change, 55
Poisson random counting measure, 53
Poisson random field, 53
Poisson random field construction, 53
Poisson renewal counting process, 49
Poisson residual lifetime, 58
Pollaczek–Khinchine formula, 337
Polya recurrence/transience criteria for simple symmetric random walk, 43
Predictable, 186
Pre-τ sigmafield, 164
Predictable strategy, 282
Pre-τ σ-field, 73, 135
Pricing measure, 287
Product probability, 4
Progeny, 263
Progressive measurability, 165
Progressively measurable, 165, 186
Proportionate loss policy, 145
Put-call parity, 289

Q
Queueing process, 14

R
Radon–Nikodym theorem, 162
Random field, 2
Random walks, 125
  boundary distribution, 19
  boundary distribution under zero drift, 21
  bridge, 242
  extreme values, 20
  extremes (max/min), 38
  Markov property, 72
Range of random walk, 24
Recurrence of one-dimensional Brownian motion with zero drift, 87
Recurrent state, 22, 100
Recurrent stochastic process, 22
Reflecting Brownian motion, 82, 230
  Lévy representation, 230
Reflection principle for Brownian motion, 81
Reflection principle for simple random walk, 28
Reflection principle for simple symmetric random walk, 38
Regime switching model, 288
Regularization of submartingales theorem, 170
Regularly varying, 330
Reinsurance, 145
  policy, 145
  relative security load, 145
Renewal age, 111
Renewal counting process, 308
Renewal epochs, 106
Renewal equation, 109
Renewal lifespan, 111
Renewal measure, 109, 306, 307
Renewal process
  delayed, 308
  ordinary, 308
Renewal times, 106, 308
Reservoir storage model, 143
Residual life process, 106
Reverse martingale, 131
Right-closed martingale, 128
Risk free measure, 285
Risk reserve process, 145
Risk, ruin, reinsurance, 143
Ruin probability, 143
Running maximum for Brownian motion, 81

S
Sample path, 2
Sample path regularity of sub-martingales, 170
Schwartz space, 370

Self-financing, 284
  hedging strategy existence, 282
  strategy, 282
Self-similar fractals, 325
Seneta theorem, 183
Shifted Brownian motion, 228
Sierpinksi gasket, 328
Similarity contraction map, 325
Similarity dimension, 326
Simple asymmetric random walk, 141
Simple point process, 185
Simple random walk, 5, 17
  construction, 18
  Markov property, 72
  range, 24
Simple symmetric random walk, 5
  boundary distribution, 140
  recurrence, 74
Size-bias change of measure, 180
Size-bias offspring distribution, 180
Skorokhod embedding theorem, 203, 204
Skorokhod space, 75
Slowly varying, 330
Sparre-Andersen model, 143, 336
Spence's function, 376
Spine decomposition, 181, 244
Spitzer comparison lemma, 119
Stable distribution, 82
Standard Brownian motion, 65
Standard space, 2
State space, 2
Stationary increments, 50
Stationary process, 7, 102
Stationary transition probabilities, 9, 12
Stirling formula, 34
Stochastic integral, 132
Stochastic Lebesgue-Stieltjes lemma, 187
Stochastic process, 2
Stochastic recursion, 182
Stock price model, 280
Stopped process, 135
Stopping time, 72, 102
Strassen law of the iterated logarithm, 90
Strike price, 280
Strong disorder, 252
Strong law of large numbers, 159
Strong Markov property, 73
  for Brownian motion, 80
  for Poisson process, 80
  for random walk, 74
  for right-continuous Markov process, 78
Subcritical branching, 114
Subexponential distribution, 330
Submartingale, 124
Submartingale convergence theorem, 153
Successful coupling, 103
Supercritical branching, 114
Supermartingale, 124
Symmetrically dependent sequence, 159

T
Tail σ-field, 159
Tempered distribution, 370
Tied-down Brownian motion, 235
Time-homogeneous Markov process, 76
Time-homogeneous transition probability, 9
Total progeny, 263
Trace $\sigma-$field, 165
Transience of one-dimensional Brownian motion with drift, 87, 88
Transient state, 21, 100
Transient stochastic process, 21
Transition probability, 9, 12
Translation-invariant, 7
Tree graph, 178

U
Upcrossing, 151
Upcrossing inequality, 153
Upper class at infinity, 93

V
Value at risk, 147
Version of a stochastic process, 61, 167

W
Waiting time paradox, 112, 328
Wald's First Identity, 140
Wald's identities, 140
Wald's Second Identity, 140
Weak disorder, 252
Wiener measure, 67

Y
Yaglom's exponential law for branching processes, 117
Yule branching process, 59, 377
Yule non-explosion, 377
Yule process, 377

Z
Zero-one law for inherited properties, 155
Zero set of Brownian motion, 222