APPLICATIONS OF MATHEMATICS

STOCHASTIC MODELLING AND APPLIED PROBABILITY

J. Michael Steele

Stochastic Calculus and Financial Applications

Springer

Stochastic Mechanics
Random Media
Signal Processing and Image Synthesis
Mathematical Economics
Stochastic Optimization and Finance
Stochastic Control

Applications of Mathematics

Stochastic Modelling and Applied Probability

45

Edited by
I. Karatzas
M. Yor

Advisory Board
P. Brémaud
E. Carlen
W. Fleming
D. Geman
G. Grimmett
G. Papanicolaou
J. Scheinkman

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Applications of Mathematics

1 Fleming/Rishel, **Deterministic and Stochastic Optimal Control** (1975)
2 Marchuk, **Methods of Numerical Mathematics,** Second Ed. (1982)
3 Balakrishnan, **Applied Functional Analysis,** Second Ed. (1981)
4 Borovkov, **Stochastic Processes in Queueing Theory** (1976)
5 Liptser/Shiryayev, **Statistics of Random Processes I: General Theory** (1977)
6 Liptser/Shiryayev, **Statistics of Random Processes II: Applications** (1978)
7 Vorob'ev, **Game Theory: Lectures for Economists and Systems Scientists** (1977)
8 Shiryayev, **Optimal Stopping Rules** (1978)
9 Ibragimov/Rozanov, **Gaussian Random Processes** (1978)
10 Wonham, **Linear Multivariable Control: A Geometric Approach,** Third Ed. (1985)
11 Hida, **Brownian Motion** (1980)
12 Hestenes, **Conjugate Direction Methods in Optimization** (1980)
13 Kallianpur, **Stochastic Filtering Theory** (1980)
14 Krylov, **Controlled Diffusion Processes** (1980)
15 Prabhu, **Stochastic Storage Processes: Queues, Insurance Risk, Dams, and Data Communication,** Second Ed. (1998)
16 Ibragimov/Has'minskii, **Statistical Estimation: Asymptotic Theory** (1981)
17 Cesari, **Optimization: Theory and Applications** (1982)
18 Elliott, **Stochastic Calculus and Applications** (1982)
19 Marchuk/Shaidourov, **Difference Methods and Their Extrapolations** (1983)
20 Hijab, **Stabilization of Control Systems** (1986)
21 Protter, **Stochastic Integration and Differential Equations** (1990)
22 Benveniste/Métivier/Priouret, **Adaptive Algorithms and Stochastic Approximations** (1990)
23 Kloeden/Platen, **Numerical Solution of Stochastic Differential Equations** (1992)
24 Kushner/Dupuis, **Numerical Methods for Stochastic Control Problems in Continuous Time** (1992)
25 Fleming/Soner, **Controlled Markov Processes and Viscosity Solutions** (1993)
26 Baccelli/Brémaud, **Elements of Queueing Theory** (1994)
27 Winkler, **Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: An Introduction to Mathematical Aspects** (1994)
28 Kalpazidou, **Cycle Representations of Markov Processes** (1995)
29 Elliott/Aggoun/Moore, **Hidden Markov Models: Estimation and Control** (1995)
30 Hernández-Lerma/Lasserre, **Discrete-Time Markov Control Processes: Basic Optimality Criteria** (1996)
31 Devroye/Györfi/Lugosi, **A Probabilistic Theory of Pattern Recognition** (1996)
32 Maitra/Sudderth, **Discrete Gambling and Stochastic Games** (1996)

(continued after index)

J. Michael Steele

Stochastic Calculus and Financial Applications

Springer

J. Michael Steele
The Wharton School
Department of Statistics
University of Pennsylvania
3000 Steinberg Hall–Dietrich Hall
Philadelphia, PA 19104-6302, USA

Managing Editors

I. Karatzas
Departments of Mathematics and Statistics
Columbia University
New York, NY 10027, USA

M. Yor
CNRS, Laboratoire de Probabilités
Université Pierre et Marie Curie
4, Place Jussieu, Tour 56
F-75252 Paris Cedex 05, France

With 3 figures.

Mathematics Subject Classification (2000): 60G44, 60H05, 91B28, 60G42

Library of Congress Cataloging-in-Publication Data
Steele, J. Michael.
Stochastic calculus and financial applications / J. Michael Steele.
p. cm. — (Applications of mathematics ; 45)
Includes bibliographical references and index.

1. Stochastic analysis. 2. Business mathematics. I. Title. II. Series.
QA274.2 .S74 2000
519.2—dc21 00-025890

Printed on acid-free paper.

© 2001 Springer-Verlag New York, Inc.
Softcover reoprint of the hardcover 1st edition 2001

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Timothy Taylor; manufacturing supervised by Jeffrey Taub.
Camera-ready copy prepared from the author's TeX files.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4419-2862-7 ISBN 978-1-4684-9305-4 (eBook)
DOI 10.1007/978-1-4684-9305-4

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Preface

This book is designed for students who want to develop professional skill in stochastic calculus and its application to problems in finance. The Wharton School course that forms the basis for this book is designed for energetic students who have had some experience with probability and statistics but have not had advanced courses in stochastic processes. Although the course assumes only a modest background, it moves quickly, and in the end, students can expect to have tools that are deep enough and rich enough to be relied on throughout their professional careers.

The course begins with simple random walk and the analysis of gambling games. This material is used to motivate the theory of martingales, and, after reaching a decent level of confidence with discrete processes, the course takes up the more demanding development of continuous-time stochastic processes, especially Brownian motion. The construction of Brownian motion is given in detail, and enough material on the subtle nature of Brownian paths is developed for the student to evolve a good sense of when intuition can be trusted and when it cannot. The course then takes up the Itô integral in earnest. The development of stochastic integration aims to be careful and complete without being pedantic.

With the Itô integral in hand, the course focuses more on models. Stochastic processes of importance in finance and economics are developed in concert with the tools of stochastic calculus that are needed to solve problems of practical importance. The financial notion of replication is developed, and the Black-Scholes PDE is derived by three different methods. The course then introduces enough of the theory of the diffusion equation to be able to solve the Black–Scholes partial differential equation and prove the uniqueness of the solution. The foundations for the martingale theory of arbitrage pricing are then prefaced by a well-motivated development of the martingale representation theorems and Girsanov theory. Arbitrage pricing is then revisited, and the notions of admissibility and completeness are developed in order to give a clear and professional view of the fundamental formula for the pricing of contingent claims.

This is a text with an attitude, and it is designed to reflect, wherever possible and appropriate, a prejudice for the concrete over the abstract. Given good general skill, many people can penetrate most deeply into a mathematical theory by focusing their energy on the mastery of well-chosen examples. This does not deny that good abstractions are at the heart of all mathematical subjects. Certainly, stochastic calculus has no shortage of important abstractions that have stood the test of time. These abstractions are to be cherished and nurtured. Still, as a matter of principle, each abstraction that entered the text had to clear a high hurdle.

Many people have had the experience of learning a subject in 'spirals.' After penetrating a topic to some depth, one makes a brief retreat and revisits earlier topics with the benefit of fresh insights. This text builds on the spiral model in several ways. For example, there is no shyness about exploring a special case before discussing a general result. There also are some problems that are solved in several different ways, each way illustrating the strength or weakness of a new technique.

Any text must be more formal than a lecture, but here the lecture style is followed as much as possible. There is also more concern with 'pedagogic' issues than is common in advanced texts, and the text aims for a coaching voice. In particular, readers are encouraged to use ideas such as George Pólya's "Looking Back" technique, numerical calculation to build intuition, and the art of guessing before proving. The main goal of the text is to provide a professional view of a body of knowledge, but along the way there are even more valuable skills one can learn, such as general problem-solving skills and general approaches to the invention of new problems.

This book is not designed for experts in probability theory, but there are a few spots where experts will find something new. Changes of substance are far fewer than the changes in style, but some points that might catch the expert eye are the explicit use of wavelets in the construction of Brownian motion, the use of linear algebra (and dyads) in the development of Skorohod's Embedding, the use of martingales to achieve the approximation steps needed to define the Itô integral, and a few more.

Many people have helped with the development of this text, and it certainly would have gone unwritten except for the interest and energy of more than eight years of Wharton Ph.D. students. My fear of omissions prevents me from trying to list all the students who have gone out of their way to help with this project. My appreciation for their years of involvement knows no bounds.

Of the colleagues who have helped personally in one way or another with my education in the matters of this text, I am pleased to thank Erhan Çinlar, Kai Lai Chung, Darrell Duffie, David Freedman, J. Michael Harrison, Michael Phelan, Yannis Karatzas, Wenbo Li, Andy Lo, Larry Shepp, Steve Shreve, and John Walsh. I especially thank Jim Pitman and Ruth Williams for their comments on an early draft of this text. They saved me from some grave errors, and they could save me from more if time permitted. Finally, I would like to thank Vladimir Pozdnyakov for hundreds of hours of conversation on this material. His suggestions were especially influential on the last five chapters.

J. Michael Steele
Philadelphia, PA

Contents

CHAPTER 1

Random Walk and First Step Analysis

The fountainhead of the theory of stochastic processes is simple random walk. Already rich in unexpected and elegant phenomena, random walk also leads one inexorably to the development of Brownian motion, the theory of diffusions, the Itô calculus, and myriad important applications in finance, economics, and physical science.

Simple random walk provides a model of the wealth process of a person who makes a living by flipping a fair coin and making fair bets. We will see it is a hard living, but first we need some notation. We let $\{X_i : 1 \le i < \infty\}$ denote a sequence of independent random variables with the probability distribution given by

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}.$$

Next, we let S_0 denote an arbitrary integer that we view as our gambler's initial wealth, and for $1 \le n < \infty$ we let S_n denote S_0 plus the partial sum of the X_i:

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n.$$

If we think of $S_n - S_0$ as the net winnings after n fair wagers of one dollar each, we almost have to inquire about the probability of the gambler winning A dollars before losing B dollars. To put this question into useful notation, we do well to consider the first time τ at which the partial sum S_n reaches level A or level $-B$:

$$\tau = \min\{n \ge 0 : S_n = A \text{ or } S_n = -B\}.$$

At the random time τ, we have $S_\tau = A$ or $S_\tau = -B$, so our basic problem is to determine $P(S_\tau = A \mid S_0 = 0)$. Here, of course, we permit the wealth of the idealized gambler to become negative — not an unrealistic situation.

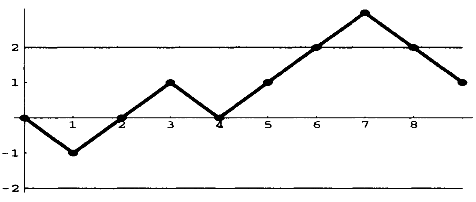

FIGURE 1.1. HITTING TIME OF LEVEL ± 2 IS 6

1.1. First Step Analysis

The solution of this problem can be obtained in several ways, but perhaps the most general method is *first step analysis*. One benefit of this method is that it is completely elementary in the sense that it does not require any advanced mathematics. Still, from our perspective, the main benefit of first step analysis is that it provides a benchmark by which to gauge more sophisticated methods.

For our immediate problem, first step analysis suggests that we consider the gambler's situation after one round of the game. We see that his wealth has either increased by one dollar or decreased by one dollar. We then face a problem that replicates our original problem except that the "initial" wealth has changed. This observation suggests that we look for a recursion relation for the function

$$f(k) = P(S_\tau = A \mid S_0 = k), \qquad \text{where } -B \leq k \leq A.$$

In this notation, $f(0)$ is precisely the desired probability of winning A dollars before losing B dollars.

If we look at what happens as a consequence of the first step, we immediately find the desired recursion for $f(k)$,

$$f(k) = \frac{1}{2} f(k-1) + \frac{1}{2} f(k+1) \text{ for } -B < k < A, \tag{1.1}$$

and this recursion will uniquely determine f when it is joined with the boundary conditions

$$f(A) = 1 \text{ and } f(-B) = 0.$$

The solution turns out to be a snap. For example, if we let $f(-B+1) = \alpha$ and substitute the values of $f(-B)$ and $f(-B+1)$ into equation (1.1), we find that $f(-B+2) = 2\alpha$. If we then substitute the values of $f(-B+1)$ and $f(-B+2)$ into equation (1.1) we find $f(-B+3) = 3\alpha$, whence it is no great leap to guess that we have $f(-B+k) = k\alpha$ for all $0 \leq k \leq A+B$.

Naturally, we verify the guess simply by substitution into equation (1.1). Finally, we determine that $\alpha = 1/(A+B)$ from the right boundary condition $f(A) = 1$ and the fact that for $k = A+B$ our conjectured formula for f requires $f(A) = (A+B)\alpha$. In the end, we arrive at a formula of remarkable simplicity and grace:

$$P(S_n \text{ reaches } A \text{ before } -B \mid S_0 = 0) = \frac{B}{A+B}. \tag{1.2}$$

Looking Back

When we look back at this formula, we find that it offers several reassuring checks. First, when $A = B$ we get $\frac{1}{2}$, as we would guess by symmetry. Also, if we replace A and B by $2A$ and $2B$ the value of the right-hand side of formula (1.2) does not change. This is also just as one would expect, say by considering the outcome of pairs of fair bets. Finally, if $A \to \infty$ we see the gambler's chance of reaching A before $-B$ goes to zero, exactly as common sense would tell us.

Simple checks such as these are always useful. In fact, George Pólya made "Looking Back" one of the key tenets of his lovely book *How to Solve It*, a volume that may teach as much about doing mathematics as any ever written. From time to time, we will take advantage of further advice that Pólya offered about looking back and other aspects of problem solving.

1.2. Time and Infinity

Our derivation of the hitting probability formula (1.2) would satisfy the building standards of all but the fussiest communities, but when we check the argument we find that there is a logical gap; we have tacitly assumed that τ is finite. How do we know for sure that the gambler's net winnings will eventually reach A or $-B$? This important fact requires proof, and we will call on a technique that exploits a general principle: if something is possible — and there are infinitely many "serious" attempts — then it will happen.

Consider the possibility that the gambler wins $A+B$ times in a row. If the gambler's fortune has not already hit $-B$, then a streak of $A+B$ wins is guaranteed to boost his fortune above A. Such a run of luck is unlikely, but it has positive probability—in fact, probability $p = 2^{-A-B}$. Now, if we let E_k denote the event that the gambler wins on each turn in the time interval $[k(A+B), (k+1)(A+B)-1]$, then the E_k are independent events, and $\tau > n(A+B)$ implies that all of the E_k with $0 \le k \le n$ fail to occur. Thus, we find

$$P(\tau > n(A+B) \mid S_0 = 0) \le P(\cap_{k=0}^{n-1} E_k^c) = (1-p)^n. \tag{1.3}$$

Since $P(\tau = \infty \mid S_0 = 0) \le P(\tau > n(A+B) \mid S_0 = 0)$ for all n, we see from equation (1.3) that $P(\tau = \infty \mid S_0 = 0) = 0$, just as we needed to show to justify our earlier assumption.

By a small variation on this technique, we can even deduce from equation (1.3) that τ has moments of all orders. As a warmup, first note that if $1(A)$ denotes the indicator function of the event A, then for any integer-valued nonnegative random variable Z we have the identity

$$Z = \sum_{k=1}^{\infty} 1(Z \ge k). \tag{1.4}$$

If we take expectations on both sides of the identity (1.4), we find a handy formula that textbooks sometimes prove by a tedious summation by parts:

$$E(Z) = \sum_{k=1}^{\infty} P(Z \ge k). \tag{1.5}$$

We will use equations (1.4) and (1.5) on many occasions, but much of the time we do not need an exact representation. In order to prove that $E(\tau^d) < \infty$ we can get along just as well with rough bounds. For example, if we sum the crude estimate

$$\tau^d 1[(k-1)(A+B) < \tau \le k(A+B)] \le k^d(A+B)^d 1[(k-1)(A+B) < \tau],$$

over k, then we have

$$\tau^d \le \sum_{k=1}^{\infty} k^d (A+B)^d 1[(A+B)(k-1) < \tau]. \tag{1.6}$$

We can then take expectations on both sides of the inequality (1.6) and apply the tail estimate (1.3). The ratio test finally provides the convergence of the bounding sum:

$$E(\tau^d) \le \sum_{k=1}^{\infty} k^d (A+B)^d (1-p)^{k-1} < \infty.$$

A Second First Step

Once we know that τ has a finite expectation, we are almost immediately drawn to the problem of determining the value of that expectation. Often, such ambitious questions yield only partial answers, but this time the answer could not be more complete or more beautiful.

Again, we use first step analysis, although now we are interested in the function defined by

$$g(k) = E(\tau \mid S_0 = k).$$

After one turn of the game, two things will have happened: the gambler's fortune will have changed, and a unit of time will have passed. The recurrence equation that we obtain differs from the one found earlier only in the appearance of an additional constant term:

$$g(k) = \frac{1}{2}g(k-1) + \frac{1}{2}g(k+1) + 1 \text{ for } -B < k < A. \tag{1.7}$$

Also, since the time to reach A or $-B$ is zero if S_0 already equals A or $-B$, we have new boundary conditions:

$$g(-B) = 0 \text{ and } g(A) = 0.$$

This time our equation is not so trivial that we can guess the answer just by calculating a couple of terms. Here, our guess is best aided by finding an appropriate analogy. To set up the analogy, we introduce the forward difference operator defined by

$$\Delta g(k-1) = g(k) - g(k-1),$$

and we note that applying the operator twice gives

$$\Delta^2 g(k-1) = g(k+1) - 2g(k) + g(k-1).$$

The recurrence equation (1.7) can now be written rather elegantly as a second order difference equation:

$$\frac{1}{2}\Delta^2 g(k-1) = -1 \text{ for } -B < k < A. \tag{1.8}$$

The best feature of this reformulation is that it suggests an immediate analogy. The integer function $g\colon \mathbb{N} \to \mathbb{R}$ has a constant second *difference*, and the real functions with a constant second *derivative* are just quadratic polynomials, so one is naturally led to look for a solution to equation (1.7) that is a quadratic over the integers. By the same analogy, equation (1.8) further suggests that the coefficient of k^2 in the quadratic should be -1. Finally, the two boundary conditions tell us that the quadratic must vanish at $-B$ and A, so we are left with only one reasonable guess,

$$g(k) = -(k-A)(k+B). \tag{1.9}$$

To verify that this guess is indeed an honest solution only requires substitution into equation (1.7). This time we are lucky. The solution does check, and our analogies have provided a reliable guide.

Finally, we note that when we specialize our formula to $k = 0$, we come to a result that could not be more striking:

$$E(\tau \mid S_0 = 0) = AB. \tag{1.10}$$

This formula is a marvel of simplicity — no better answer could even be imagined. Moreover, when we look back on equation (1.10), we find several interesting deductions.

For example, if we let $\tau' = \min\{n \geq 0 : S_n = -1\}$ and set

$$\tau'' = \min\{n \geq 0 : S_n = -1 \text{ or } S_n = A\},$$

then we see that $\tau'' \le \tau'$. But equation (1.10) tells us $E(\tau'') = A$ so we find that $E(\tau') \ge A$ for all A. The bottom line is that $E(\tau') = \infty$, or, in other words, the expected time until the gambler gets behind by even one dollar is infinite.

This remarkable fact might give the gambler some cause for celebration, except for the sad symmetrical fact that the expected time for the gambler to get ahead by one dollar is also infinite. Strangely, one of these two events must happen on the very first bet; thus we face one of the many paradoxical features of the fair coin game.

There are several further checks that we might apply to formula (1.10), but we will pursue just one more. If we consider the symmetric interval $[-A, A]$, is there some way that we might have guessed that the expected time until the first exit should be a quadratic function of A? One natural approach to this question is to consider the expected size of $|S_n|$. The central limit theorem and a bit of additional work will tell us that $E(|S_n|) \sim \sqrt{2n/\pi}$, so when both n and A are large we see that $E(|S_n|)$ will first leave the interval $[-A, A]$ when $n \sim \pi A^2/2$. This observation does not perfectly parallel our exit-time formula (1.10), but it does suggest that a quadratic growth rate is in the cards.

1.3. Tossing an Unfair Coin

It is often remarked that life is not fair, and, be that as it may, there is no doubt that many gambling games are not even-handed. Considerable insight into the difficulties that face a player of an unfair game can be found by analysis of the simplest model — the biased random walk defined by $S_n = S_0 + X_1 + X_2 + \cdots + X_n$, where

$$P(X_i = 1) = p \text{ and } P(X_i = -1) = 1 - p = q \text{ where } p \neq q.$$

To solve the ruin problem for biased random walk, we take $f(k)$ and τ as before and note that first step analysis leads us to

$$f(k) = pf(k+1) + qf(k-1).$$

This is another equation that is most easily understood if it is written in terms of the difference operator. First, we note that since $p + q = 1$ the equation can be rearranged to give

$$0 = p\{f(k+1) - f(k)\} - q\{f(k) - f(k-1)\},$$

from which we find a simple recursion for $\Delta f(k)$:

$$\Delta f(k) = (q/p)\Delta f(k-1). \tag{1.11}$$

Now, we simply iterate equation (1.11) to find

$$\Delta f(k+j) = (q/p)^j \Delta f(k),$$

so, if we set $\alpha = \Delta f(-B)$, we can exploit the fact that $f(-B) = 0$ and successive cancellations to find

$$f(k) = \sum_{j=0}^{k+B-1} \Delta f(j - B) = \alpha \sum_{j=0}^{k+B-1} (q/p)^j = \alpha \frac{(q/p)^{k+B} - 1}{(q/p) - 1}. \tag{1.12}$$

We can then eliminate α from equation (1.12) if we let $k = A$ and invoke our second boundary condition:

$$1 = f(A) = \alpha \frac{(q/p)^{A+B} - 1}{(q/p) - 1}.$$

After determining α, we return to equation (1.12) and take $k = 0$ to get to the bottom line; for biased random walk, we have a simple and explicit formula for the ruin probability:

$$P(S_n \text{ hits } A \text{ before } -B \mid S_0 = 0) = \frac{(q/p)^B - 1}{(q/p)^{A+B} - 1}. \tag{1.13}$$

This formula would transform the behavior of millions of the world's gamblers, if they could only take it to heart. Such a conversion is unlikely, though perhaps a few might be moved to change their ways if they would work out the implications of equation (1.13) for some typical casino games.

Time and Time Again

The expected time until the biased random walk hits either level A or $-B$ can also be found by first step analysis. If $g(k)$ denotes the expected time until the random walk hits A or $-B$ when we start at k, then the equation given by first step analysis is just

$$g(k) = pg(k+1) + qg(k-1) + 1.$$

As before, this equation is better viewed in difference form

$$\Delta g(k) = (q/p)\Delta g(k-1) - 1/p, \tag{1.14}$$

where the boundary conditions are the same as those we found for the unbiased walk

$$g(-B) = 0 \text{ and } g(A) = 0.$$

To solve equation (1.14), we first note that if we try a solution of the form ck then we find that $g_0(k) = k/(q-p)$ is one solution of the inhomogeneous equation (1.14). From our earlier work we also know that $\alpha + \beta(q/p)^k$ is a solution of the homogeneous equation (1.11), so to obtain a solution that handles the boundary conditions we consider solutions of the form

$$g(k) = \frac{k}{q-p} + \alpha + \beta(q/p)^k.$$

The two boundary conditions give us a pair of equations that we can solve to determine α and β in order to complete the determination of $g(k)$. Finally, when we specialize to $g(0)$, we find the desired formula for the expected hitting time of $-B$ or A for the biased random walk:

$$E(\tau \mid S_0 = 0) = \frac{B}{q-p} - \frac{A+B}{q-p}\,\frac{1-(q/p)^B}{1-(q/p)^{A+B}}. \tag{1.15}$$

The formulas for the hitting probabilities (1.13) and the expected hitting time (1.15) are more complicated than their cousins for unbiased walk, but they answer more complex questions. When we look back on these formulas, we naturally want to verify that they contain the results that were found earlier, but one cannot recapture the simpler formulas just by setting $p = q = \frac{1}{2}$. Nevertheless, formulas (1.13) and (1.15) are consistent with the results that were obtained for unbiased walks. If we let $p = \frac{1}{2} + \epsilon$ and $q = \frac{1}{2} - \epsilon$ in equations (1.13) and (1.15), we find that as $\epsilon \to 0$ equations (1.13) and (1.15) reduce to $B/(A+B)$ and AB, as one would expect.

1.4. Numerical Calculation and Intuition

The formulas for the ruin probabilities and expected hitting times are straightforward, but for someone interested in building serious streetwise intuition there is nothing that beats numerical computation.

- We now know that in a fair game of coin tosses and \$1 wagers the expected time until one of the players gets ahead by \$100 is 10,000 tosses, a much larger number than many people might expect.
- If our gambler takes up a game with probability $p = 0.49$ of winning on each round, he has less than a 2% chance of winning \$100 before losing \$200. This offers a stark contrast to the fair game, where the gambler would have a 2/3 probability of winning \$100 before losing \$200. The cost of even a small bias can be surprisingly high.

In the table that follows, we compute the probability of winning \$100 before losing \$100 in some games with odds that are typical of the world's casinos. The table assumes a constant bet size of \$1 on all rounds of the game.

TABLE 1.1. STREETWISE BENCHMARKS.

Chance on one round	0.500	0.495	0.490	0.480	0.470
Chance to win \$100	0.500	0.1191	0.0179	0.0003	6×10^{-6}
Duration of the game	10,000	7,616	4,820	2,498	1,667

One of the lessons we can extract from this table is that the traditional movie character who chooses to wager everything on a single round of roulette is not so foolish; there is wisdom to back up the bravado. In a game with a 0.47 chance to win on each bet, you are about 78,000 times more likely to win \$100 by betting \$100 on a single round than by playing just \$1 per round. Does this add something to your intuition that seems to go beyond the formula for the ruin probability?

1.5. First Steps with Generating Functions

We have obtained compelling results for the most natural problems of gambling in either fair or unfair games, and these results make a sincere contribution to our understanding of the real world. It would be perfectly reasonable to move to other problems before bothering to press any harder on these simple models. Nevertheless, the first step method is far from exhausted, and, if one has the time and interest, much more detailed information can be obtained with just a little more work.

For example, suppose we go back to simple random walk and consider the problem of determining the probability *distribution* of the first hitting time of level 1 given that the walk starts at zero. Our interest is no longer confined to a single number, so we need a tool that lets us put all of the information of a discrete distribution into a package that is simple enough to crack with first step analysis.

If we let τ denote this hitting time, then the appropriate package turns out to be the probability generating function:

$$\phi(z) = E(z^\tau \mid S_0 = 0) = \sum_{k=0}^{\infty} P(\tau = k \mid S_0 = 0) z^k. \tag{1.16}$$

If we can find a formula for $\phi(z)$ and can compute the Taylor expansion of $\phi(z)$ from that formula, then by identifying the corresponding coefficients we will have found $P(\tau = k \mid S_0 = 0)$ for all k. Here, one should also note that once we understand τ we also understand the distribution of the first time to go up k levels; the probability generating function in that case is given by $\phi(z)^k$ because the probability generating function of a sum of independent random variables is simply the product of the probability generating functions.

Now, although we want to determine a function, first step analysis proceeds much as before. When we take our first step, two things happen. First, there is the passage of one unit of time; and, second, we will have moved from zero to either -1 or 1. We therefore find on a moment's reflection that

$$\phi(z) = \frac{1}{2}E(z^{\tau+1} \mid S_0 = -1) + \frac{1}{2}E(z^{\tau+1} \mid S_0 = 1). \tag{1.17}$$

Now, $E(z^\tau \mid S_0 = -1)$ is the same as the probability generating function of the first time to reach level 2 starting at 0, and we noted earlier that this is exactly $\phi(z)^2$. We also have $E(z^\tau \mid S_0 = 1) = 1$, so equation (1.17) yields a quadratic equation for $\phi(z)$:

$$\phi(z) = \frac{1}{2}z\phi(z)^2 + \frac{1}{2}z. \tag{1.18}$$

In principle $\phi(z)$ is now determined, but we can get a thoroughly satisfying answer only if we exercise some discrete mathematics muscle. When we first apply the quadratic formula to solve equation (1.18) for $\phi(z)$ we find two candidate solutions. Since $\tau \geq 1$, the definition of $\phi(z)$ tells us that $\phi(0) = 0$, and only one of the solutions of equation (1.18) evaluates to zero when $z = 0$, so we can deduce that

$$\phi(z) = \frac{1 - \sqrt{1 - z^2}}{z}. \tag{1.19}$$

The issue now boils down to finding the coefficients in the Taylor expansion of $\phi(z)$. To get these coefficients by successive differentiation is terribly boring, but we can get them all rather easily if we recall Newton's generalization of the binomial theorem. This result tells us that for any exponent $\alpha \in \mathbb{R}$, we have

$$(1 + y)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} y^k, \tag{1.20}$$

where the binomial coefficient is defined to be 1 for $k = 0$ and is defined by

$$\binom{\alpha}{k} = \frac{\alpha(\alpha - 1)\cdots(\alpha - k + 1)}{k!} \tag{1.21}$$

for $k > 0$. Here, we should note that if α is equal to a nonnegative integer m, then the Newton coefficients (1.21) reduce to the usual binomial coefficients, and Newton's series reduces to the usual binomial formula.

When we apply Newton's formula to $(1 - z^2)^{\frac{1}{2}}$, we quickly find the Taylor expansion for ϕ:

$$\phi(z) = \frac{1 - \sqrt{1 - z^2}}{z} = \sum_{k=1}^{\infty} \binom{1/2}{k} (-1)^{k+1} z^{2k-1},$$

and when we compare this expansion with the definition of $\phi(z)$ given by equation (1.16), we can identify the corresponding coefficients to find

$$P(\tau = 2k-1 \mid S_0 = 0) = \binom{1/2}{k}(-1)^{k+1}. \tag{1.22}$$

The last expression is completely explicit, but it can be written a bit more comfortably. If we expand Newton's coefficient and rearrange terms, we quickly find a formula with only conventional binomials:

$$P(\tau = 2k-1 \mid S_0 = 0) = \frac{1}{2k-1}\binom{2k}{k}2^{-2k}. \tag{1.23}$$

This formula and a little arithmetic will answer any question one might have about the distribution of τ. For example, it not only tells us that the probability that our gambler's winnings go positive for the first time on the fifth round is $1/16$, but it also resolves more theoretical questions such as showing

$$E(\tau^\alpha) < \infty \text{ for all } \alpha < 1/2,$$

even though we have

$$E(\tau^\alpha) = \infty \text{ for all } \alpha \geq 1/2.$$

1.6. Exercises

The first exercise suggests how results on biased random walks can be worked into more realistic models. Exercise 1.2 then develops the fundamental recurrence property of simple random walk. Finally, Exercise 1.3 provides a mind-stretching result that may seem unbelievable at first.

EXERCISE 1.1 (Complex Models from Simple Ones). Consider a naive model for a stock that has a support level of \$20/share because of a corporate buy-back program. Suppose also that the stock price moves randomly with a downward bias when the price is above \$20 and randomly with an upward bias when the price is below \$20. To make the problem concrete, we let Y_n denote the stock price at time n, and we express our support hypothesis by the assumption that

$$P(Y_{n+1} = 21 \mid Y_n = 20) = 0.9, \text{ and } P(Y_{n+1} = 19 \mid Y_n = 20) = 0.1.$$

We then reflect the downward bias at price levels above \$20 by requiring for $k > 20$ that

$$P(Y_{n+1} = k+1 \mid Y_n = k) = 1/3 \text{ and } P(Y_{n+1} = k-1 \mid Y_n = k) = 2/3.$$

The upward bias at price levels below \$20 is expressed by assuming for $k < 20$ that

$$P(Y_{n+1} = k+1 \mid Y_n = k) = 2/3 \text{ and } P(Y_{n+1} = k-1 \mid Y_n = k) = 1/3.$$

Calculate the expected time for the stock price to fall from \$25 through the support level of \$20 all the way down to \$18.

EXERCISE 1.2 (Recurrence of SRW). If S_n denotes simple random walk with $S_0 = 0$, then the usual binomial theorem immediately gives us the probability that we are back at 0 at time $2k$:

$$P(S_{2k} = 0 \mid S_0 = 0) = \binom{2k}{k} 2^{-2k}. \tag{1.24}$$

(a) First use Stirling's formula $k! \sim \sqrt{2\pi k}\, k^k e^{-k}$ to justify the approximation

$$P(S_{2k} = 0) \sim (\pi k)^{-\frac{1}{2}},$$

and use this fact to show that if N_n denotes the number of visits made by S_k to 0 up to time n, then $E(N_n) \to \infty$ as $n \to \infty$.

(b) Finally, prove that we have

$$P(S_n = 0 \text{ for infinitely many } n) = 1.$$

This is called the *recurrence property* of random walk; with probability one simple random walk returns to the origin infinitely many times. Anyone who wants a hint might consider the plan of calculating the expected value of

$$N = \sum_{n=1}^{\infty} 1(S_n = 0)$$

in two different ways. The direct method using $P(S_n = 0)$ should then lead without difficulty to $E(N) = \infty$. The second method is to let

$$r = P(S_n = 0 \text{ for some } n \geq 1 | S_0 = 0)$$

and to argue that

$$E(N) = \frac{r}{1-r}.$$

To reconcile this expectation with the calculation that $E(N) = \infty$ then requires $r = 1$, as we wanted to show.

(c) Let $\tau_0 = \min\{n \geq 1 \colon S_n = 0\}$ and use first step analysis together with the first-passage time probability (1.23) to show that we also have

$$P(\tau_0 = 2k) = \frac{1}{2k-1} \binom{2k}{k} 2^{-2k}. \tag{1.25}$$

Use Stirling's formula $n! \sim n^n e^{-n} \sqrt{2\pi n}$ to show that $P(\tau_0 = 2k)$ is bounded above and below by a constant multiple of $k^{-3/2}$, and use these bounds to conclude that $E(\tau_0^{\alpha}) < \infty$ for all $\alpha < \frac{1}{2}$ yet $E(\tau_0^{\frac{1}{2}}) = \infty$.

EXERCISE 1.3. Consider simple random walk beginning at 0 and show that for any $k \neq 0$ the expected number of visits to level k before returning to 0 is exactly 1. Anyone who wants a hint might consider the number N_k of visits to level k before the first return to 0. We have $N_0 = 1$ and can use the results on hitting probabilities to show that for all $k \geq 1$ we have

$$P(N_k > 0) = \frac{1}{2}\frac{1}{k} \text{ and } P(N_k > j+1 \mid N_k > j) = \frac{1}{2} + \frac{1}{2}\frac{k-1}{k}.$$

CHAPTER 2

First Martingale Steps

The theory of martingales began life with the aim of providing insight into the apparent impossibility of making money by placing bets on fair games. The success of the theory has far outstripped its origins, and martingale theory is now one of the main tools in the study of random processes. The aim of this chapter is to introduce the most intuitive features of martingales while minimizing formalities and technical details. A few definitions given here will be refined later, but the redundancy is modest, and the future abstractions should go down more easily with the knowledge that they serve an honest purpose.

We say that a sequence of random variables $\{M_n : 0 \leq n < \infty\}$ is a *martingale* with respect to the sequence of random variables $\{X_n : 1 \leq n < \infty\}$, provided that the sequence $\{M_n\}$ has two basic properties. The first property is that for each $n \geq 1$ there is a function $f_n : \mathbb{R}^n \mapsto \mathbb{R}$ such that $M_n = f_n(X_1, X_2, \ldots, X_n)$, and the second property is that the sequence $\{M_n\}$ satisfies the fundamental *martingale identity*:

$$E(M_n \mid X_1, X_2, \ldots, X_{n-1}) = M_{n-1} \text{ for all } n \geq 1. \tag{2.1}$$

To round out this definition, we will also require that M_n have a finite expectation for each $n \geq 1$, and, for a while at least, we will require that M_0 simply be a constant.

The intuition behind this definition is easy to explain. We can think of the X_i as telling us the ith outcome of some gambling process, say the head or tail that one would observe on a coin flip. We can also think of M_n as the fortune of a gambler who places fair bets in varying amounts on the results of the coin tosses. Formula (2.1) tells us that the *expected* value of the gambler's fortune at time n given all the information in the first $n-1$ flips of the coin is simply M_{n-1}, the *actual* value of the gambler's fortune before the nth round of the coin flip game.

The martingale property (2.1) leads to a theory that brilliantly illuminates the fact that a gambler in a fair game cannot expect to make money, however cleverly he varies his bets. Nevertheless, the reason for studying martingales is not that they provide such wonderful models for gambling games. The compelling reason for studying martingales is that they pop up like mushrooms all over probability theory.

2.1. Classic Examples

To develop some intuition about martingales and their basic properties, we begin with three classic examples. We will rely on these examples throughout the text, and we will find that in each case there are interesting analogs for Brownian motion as well as many other processes.

Example 1

If the X_n are independent random variables with $E(X_n) = 0$ for all $n \geq 1$, then the partial sum process given by taking $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$ for $n \geq 1$ is a martingale with respect to the sequence $\{X_n \colon 1 \leq n < \infty\}$.

Example 2

If the X_n are independent random variables with $E(X_n) = 0$ and $\mathrm{Var}(X_n) = \sigma^2$ for all $n \geq 1$, then setting $M_0 = 0$ and $M_n = S_n^2 - n\sigma^2$ for $n \geq 1$ gives us a martingale with respect to the sequence $\{X_n \colon 1 \leq n < \infty\}$.

One can verify the martingale property in the first example almost without thought, so we focus on the second example. Often, the first step one takes in order to check the martingale property is to separate the conditioned and unconditioned parts of the process:

$$E(M_n \mid X_1, X_2, \ldots, X_{n-1}) = E(S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2 \mid X_1, X_2, \ldots, X_{n-1}).$$

Now, since S_{n-1}^2 is a function of $\{X_1, X_2, \ldots, X_{n-1}\}$, its conditional expectation given $\{X_1, X_2, \ldots, X_{n-1}\}$ is just S_{n-1}^2. When we consider the second summand, we note that when we calculate the conditional expectation given $\{X_1, X_2, \ldots, X_{n-1}\}$ the sum S_{n-1} can be brought outside of the expectation

$$E(S_{n-1}X_n \mid X_1, X_2, \ldots, X_{n-1}) = S_{n-1}E(X_n \mid X_1, X_2, \ldots, X_{n-1}).$$

Next, we note that $E(X_n \mid X_1, X_2, \ldots, X_{n-1}) = E(X_n) = 0$ since X_n is independent of $X_1, X_2, \ldots, X_{n-1}$; by parallel reasoning, we also find

$$E(X_n^2 \mid X_1, X_2, \ldots, X_{n-1}) = \sigma^2.$$

When we reassemble the pieces, the verification of the martingale property for $M_n = S_n^2 - n\sigma^2$ is complete.

Example 3

For the third example, we consider independent random variables X_n such that $X_n \geq 0$ and $E(X_n) = 1$ for all $n \geq 1$. We then let $M_0 = 1$ and set

$$M_n = X_1 \cdot X_2 \cdots X_n \text{ for } n \geq 1.$$

One can easily check that M_n is a martingale. To be sure, it is an obvious multiplicative analog to our first example. Nevertheless, this third martingale offers some useful twists that will help us solve some interesting problems.

For example, if the independent identically distributed random variables Y_n have a moment generating function

$$\phi(\lambda) = E(\exp(\lambda Y_n)) < \infty,$$

then the independent random variables $X_n = \exp(\lambda Y_n)/\phi(\lambda)$ have mean one so their product leads us to a whole *parametric family* of martingales indexed by λ:

$$M_n = \exp(\lambda \sum_{i=1}^{n} Y_i)/\phi(\lambda)^n.$$

Now, if there exists a $\lambda_0 \neq 0$ such that $\phi(\lambda_0) = 1$, then there is an especially useful member of this family. In this case, when we set $S_n = \sum_{i=1}^n Y_i$ we find that

$$M_n = e^{\lambda_0 S_n}$$

is a martingale. As we will see shortly, the fact that this martingale is an explicit function of S_n makes it a particularly handy tool for study of the partial sums S_n.

SHORTHAND NOTATION

Formulas that involve conditioning on $X_1, X_2, \ldots, X_n$ can be expressed more tidily if we introduce some shorthand. First, we will write $E(Z \mid \mathcal{F}_n)$ in place of $E(Z \mid X_1, X_2, \ldots, X_n)$, and when $\{M_n : 1 \leq n < \infty\}$ is a martingale with respect to $\{X_n : 1 \leq n < \infty\}$ we will just call $\{M_n\}$ a martingale with respect to $\{\mathcal{F}_n\}$. Finally, we use the notation $Y \in \mathcal{F}_n$ to mean that Y can be written as $Y = f(X_1, X_2, \ldots, X_n)$ for some function f, and in particular if A is an event in our probability space, we will write $A \in \mathcal{F}_n$ provided that the indicator function of A is a function of the variables $\{X_1, X_2, \ldots, X_n\}$. The idea that unifies this shorthand is that we think of $\mathcal{F}_n$ as a representation of the information in the set of observations $\{X_1, X_2, \ldots, X_n\}$. A little later, we will provide this shorthand with a richer interpretation and some technical polish.

2.2. New Martingales from Old

Our intuition about gambling tells us that a gambler cannot turn a fair game into an advantageous one by periodically deciding to double the bet or by cleverly choosing the time to quit playing. This intuition will lead us to a simple theorem that has many important implications. As a necessary first step, we need a definition that comes to grips with the fact that the gambler's life would be easy if future information could be used to guide present actions.

DEFINITION 2.1. *A sequence of random variables $\{A_n : 1 \leq n < \infty\}$ is called* nonanticipating *with respect to the sequence $\{\mathcal{F}_n\}$ if for all $1 \leq n < \infty$, we have*

$$A_n \in \mathcal{F}_{n-1}.$$

In the gambling context, a nonanticipating A_n is simply a function that depends only on the information $\mathcal{F}_{n-1}$ that is known before placing a bet on the nth round of the game. This restriction on A_n makes it feasible for the gambler to permit A_n to influence the size of the nth bet, say by doubling the bet that would have been made otherwise. In fact, if we think of A_n itself as the bet multiplier, then $A_n(M_n - M_{n-1})$ would be the change in the gambler's fortune that is caused by the nth round of play. The idea of a bet size multiplier leads us to a concept that is known in more scholarly circles as the martingale transform.

DEFINITION 2.2. *The process $\{\widetilde{M}_n : 0 \leq n < \infty\}$ defined for $n = 0$ by $\widetilde{M}_0 = 0$ and for $n \geq 1$ by*

$$\widetilde{M}_n = A_1(M_1 - M_0) + A_2(M_2 - M_1) + \cdots + A_n(M_n - M_{n-1})$$

is called the martingale transform *of $\{M_n\}$ by $\{A_n\}$.*

The martingale transform gives us a general method for building new martingales out of old ones. Under a variety of mild conditions, the transform of a martingale is again a martingale. The next theorem illustrates this principle in its most useful instance.

THEOREM 2.1 (Martingale Transform Theorem). *If $\{M_n\}$ is a martingale with respect to the sequence $\{\mathcal{F}_n\}$, and if $\{A_n : 1 \le n < \infty\}$ is a sequence of* bounded *random variables that are nonanticipating with respect to $\{\mathcal{F}_n\}$, then the sequence of martingale transforms $\{\widetilde{M}_n\}$ is itself a martingale with respect to $\{\mathcal{F}_n\}$.*

PROOF. We obviously have $\widetilde{M}_n \in \mathcal{F}_n$, and the boundedness of the A_k guarantees that $\widetilde{M}_n$ has a finite expectation for all n. Finally, the martingale property follows from a simple calculation. We simply note that

$$\begin{aligned} E\left(\widetilde{M}_n - \widetilde{M}_{n-1} \mid \mathcal{F}_{n-1}\right) &= E(A_n(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}) \\ &= A_n E(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) = 0, \end{aligned} \tag{2.2}$$

and the martingale identity

$$E(\widetilde{M}_n \,|\mathcal{F}_{n-1}) = \widetilde{M}_{n-1}$$

is equivalent to equation (2.2). □

STOPPING TIMES PROVIDE MARTINGALE TRANSFORMS

One of the notions that lies at the heart of martingale theory is that of a *stopping time.* Intuitively, a stopping time is a random variable that describes a rule that we could use to decide to stop playing a gambling game. Obviously, such a rule cannot depend on the outcome of a round of the game that has yet to be played. This intuition is captured in the following definition.

DEFINITION 2.3. *A random variable τ that takes values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ is called a* stopping time *for the sequence $\{\mathcal{F}_n\}$ if*

$$\{\tau \le n\} \in \mathcal{F}_n \quad \textit{for all} \quad 0 \le n < \infty.$$

In many circumstances, we are interested in the behavior of a random process, say Y_n, precisely at the stopping time τ. If $\tau < \infty$ with probability one, then we can define the *stopped process* Y_τ by setting

$$Y_\tau = \sum_{k=0}^{\infty} 1(\tau = k) Y_k.$$

The fact that we define Y_τ only when we have $P(\tau < \infty) = 1$ should underscore that our definition of a stopping time permits the possibility that $\tau = \infty$, and it should also highlight the benefit of finding stopping times that are finite with probability one. Nevertheless, we always have truncation at our disposal; if we let $n \wedge \tau = \min\{n, \tau\}$, then $n \wedge \tau$ is a bounded stopping time, and for any sequence of random variables Y_n the stopped process $Y_{n\wedge\tau}$ is well defined. Also, the truncated stopping times $n \wedge \tau$ lead to an important class of martingales that we will use on many future occasions.

THEOREM 2.2 (Stopping Time Theorem). *If $\{M_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$, then the stopped process $\{M_{n\wedge\tau}\}$ is also a martingale with respect to $\{\mathcal{F}_n\}$.*

PROOF. First, we note that there is no loss of generality in assuming $M_0 = 0$ since otherwise we can introduce the martingale $M_n' = M_n - M_0$. Next, we note that the bounded random variables A_k defined by

$$A_k = 1(\tau \geq k) = 1 - 1(\tau \leq k-1)$$

are nonanticipating since τ is a stopping time. Finally,

$$\sum_{k=1}^{n} A_k\{M_k - M_{k-1}\} = M_\tau 1(\tau \leq n-1) + M_n 1(\tau \geq n) = M_{n\wedge\tau},$$

so we that see $\{M_{n\wedge\tau}\}$ is the martingale transform of M_n by the process $\{A_n\}$ which is bounded and nonanticipating. Theorem 2.1 then confirms that $\{M_{n\wedge\tau}\}$ is a martingale. □

2.3. Revisiting the Old Ruins

The stopping time theorem provides a new perspective on our earlier calculation of the probability that a random walk S_n starting at 0 has a probability $B/(A+B)$ of hitting level A before hitting level $-B$. If we let

$$\tau = \min\{n : S_n = A \text{ or } S_n = -B\},$$

then the stopping time theorem and the fact that S_n is a martingale combine to tell us that $S_{n\wedge\tau}$ is also a martingale, so we have

$$E[S_{n\wedge\tau}] = E[S_{0\wedge\tau}] = 0 \quad \text{for all} \quad n \geq 0. \tag{2.3}$$

Now, we checked earlier that τ is finite with probability one, so we also have

$$\lim_{n\to\infty} S_{n\wedge\tau} = S_\tau \text{ with probability one.}$$

The random variables $|S_{n\wedge\tau}|$ are bounded by $\max(A, B)$ so the dominated convergence theorem[1] tells us

$$\lim_{n\to\infty} E[S_{n\wedge\tau}] = E[S_\tau],$$

so equation (2.3) tells us

$$0 = E[S_\tau]. \tag{2.4}$$

Remarkably, we have a second way to calculate $E[S_\tau]$. We have the random variable representation

$$S_\tau = A1(S_\tau = A) - B1(S_\tau = -B),$$

so if we take expectations, we find

$$E[S_\tau] = P(S_\tau = A)\cdot A - (1 - P(S_\tau = A))\cdot B. \tag{2.5}$$

From equations (2.5) and (2.4), we therefore find that

$$0 = E[S_\tau] = P(S_\tau = A)\cdot A - (1 - P(S_\tau = A))\cdot B,$$

and we can solve this equation to find the classical formula:

$$P(S_\tau = A) = \frac{B}{A+B}.$$

[1]This is the first time we have used one of the three great tools of integration theory: the dominated convergence theorem, the monotone convergence theorem, and Fatou's lemma. A discussion of these results and a quick review of the Lebesgue integral can be found in the Appendix on Mathematical Tools.

Once More Quickly

The martingale method for calculating the ruin probability may seem long-winded compared to first step analysis, but the impression is due to the detail with which the calculations were given. With some experience behind us, we can pick up the pace considerably. For example, if we now calculate $E[\tau]$, the expected hitting time of A or $-B$ for an unbiased random walk, we can see that the martingale method is actually very quick.

For unbiased simple random walk $S_n = X_1 + X_2 + \cdots + X_n$, where the X_i are independent symmetric Bernoulli random variables, we have $\mathrm{Var}(X_i) = 1$, so we know from the first calculation in this section that $M_n = S_n^2 - n$ is a martingale. Next, we note that

$$|M_{n\wedge\tau}| \leq \max(A^2, B^2) + \tau,$$

and, since we showed earlier that $E[\tau] < \infty$, we see that for all $n \geq 1$ the random variables $M_{n\wedge\tau}$ are dominated by an integrable random variable. The martingale property of $M_{n\wedge\tau}$ gives us $E[M_{n\wedge\tau}] = 0$ for all $n \geq 0$, and $M_{n\wedge\tau}$ converges to M_τ with probability one by the finiteness of τ, so the dominated convergence theorem finally gives us

$$E[M_\tau] = \lim_{n\to\infty} E[M_{n\wedge\tau}] = 0.$$

What makes this fact useful is that again we have a second way to calculate $E[M_{n\wedge\tau}]$. If we let $\alpha = P(S_\tau = A)$ then we have $P(S_\tau = -B) = 1 - \alpha$, so we can calculate $E[M_\tau]$ directly from the elementary definition of expectation and our earlier discovery that $\alpha = B/(A+B)$ to find

$$E[M_\tau] = E[S_\tau^2] - E[\tau] = \alpha A^2 + (1-\alpha)B^2 - E[\tau] = AB - E\tau.$$

Because our martingale argument already established that $E[M_\tau] = 0$, we again find the lovely formula $E[\tau] = AB$.

Now With Bias

How about the ruin probabilities for biased random walk? This case is more interesting since we will make use of a new martingale. To be precise, if X_i are independent random variables with $P(X_i = 1) = p$ and $P(X_i = -1) = q$ where $q = 1 - p$ and $p \neq \frac{1}{2}$, then we can define a new martingale by setting $M_0 = 1$ and setting

$$M_n = (q/p)^{S_n} \quad \text{for all} \quad n \geq 1.$$

One easily verifies that M_n is indeed a martingale, so we can go directly to the calculation of $P(S_\tau = A)$. By our usual argument, $P(\tau < \infty) = 1$, so as $n \to \infty$ we see that $M_{n\wedge\tau}$ converges with probability one to M_τ. Because $M_{n\wedge\tau}$ is a martingale, we have $E[M_{n\wedge\tau}] = 1$ for all $n \geq 0$, and the random variables $M_{n\wedge\tau}$ are bounded, so the dominated convergence theorem tells us that

$$E[M_\tau] = \lim_{n\to\infty} E[M_{n\wedge\tau}] = 1.$$

We also have a bare-handed calculation of $E[M_\tau]$,

$$E[M_\tau] = P(S_\tau = A) \cdot (q/p)^A + (1 - P(S_\tau = A)) \cdot (q/p)^{-B},$$

so from the fact that $E[M_\tau] = 1$ we find an equation for $P(S_\tau = A)$. When we solve this equation, we find

$$P(S_\tau = A) = \frac{(q/p)^B - 1}{(q/p)^{A+B} - 1},$$

exactly as we found by first step analysis.

SOME PERSPECTIVE

The martingale $M_n = (q/p)^{S_n}$ may seem to have popped out of thin air, but it is actually an old friend. The more principled (and less magical) way of coming to this martingale is to use the parametric family of martingales that we built using the moment generating function. For each step X_i of a biased random walk, the moment generating function is given by

$$\phi(\lambda) = E(\exp(\lambda X_i)) = pe^\lambda + qe^{-\lambda}, \tag{2.6}$$

so by our earlier calculations we know that the process defined by

$$M_n = e^{\lambda S_n}/(pe^\lambda + qe^{-\lambda})^n$$

is a martingale for all λ.

Now, if we can find λ_0 such that

$$\phi(\lambda_0) = pe^{\lambda_0} + qe^{-\lambda_0} = 1, \tag{2.7}$$

then we see that the simple process $M_n = e^{\lambda_0 S_n}$ is a martingale. To make the last martingale completely explicit, we only need to find e^{λ_0}. To do this, we multiply equation (2.7) by $x = e^{\lambda_0}$ to get a quadratic equation in x, and then we solve that equation to find two solutions: $x = 1$ and $x = q/p$. The solution $x = 1$ gives us the trivial martingale $M_n \equiv 1$, but when we take the second choice we find the martingale $M_n = (q/p)^{S_n}$ that we found to be so handy in the solution of the ruin problem for biased random walk.

2.4. Submartingales

The applicability of martingale theory can be extended greatly if we relax the martingale identity to an analogous inequality. This wider class of processes retains many of the good features of martingales, yet it is far more flexible and robust.

DEFINITION 2.4. *If the integrable random variables $M_n \in \mathcal{F}_n$ satisfy the inequality*

$$M_{n-1} \le E(M_n \mid \mathcal{F}_{n-1}) \quad \textit{for all} \quad n \ge 1,$$

we say $\{M_n\}$ is a submartingale *adapted to $\{\mathcal{F}_n\}$.*

Submartingales are handy because many natural operations on submartingales (or martingales) lead us directly to another submartingale. For example, if $\{M_n\}$ is a martingale then $\{|M_n|\}$ is a submartingale, and if $p \ge 1$, then $\{|M_n|^p\}$ is also a submartingale, provided that $E(|M_n|^p) < \infty$ for all $n \ge 0$. As we will see shortly, these results are best understood as corollaries of a general inequality for the conditional expectation of a convex function of a random variable.