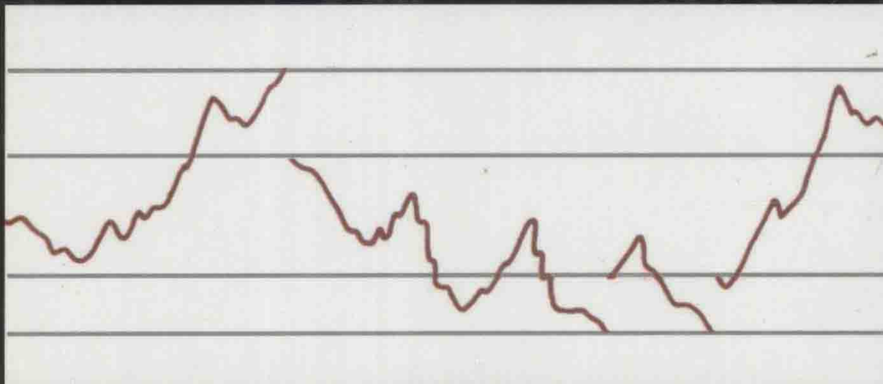
Brownian Models of Performance and Control



J. MICHAEL HARRISON

CAMBRIDGE

Brownian Models of Performance and Control

J. MICHAEL HARRISON
Stanford University, California

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107018396

First published 2013

Printed in the United States of America

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-01839-6 Hardback

Brownian Models of Performance and Control

Brownian Models of Performance and Control covers Brownian motion and stochastic calculus at the graduate level, and illustrates the use of that theory in various application domains, emphasizing business and economics. The mathematical development is narrowly focused and briskly paced, with many concrete calculations and a minimum of abstract notation. The applications discussed include: the role of reflected Brownian motion as a storage model, queueing model, or inventory model; optimal stopping problems for Brownian motion, including the influential McDonald–Siegel investment model; optimal control of Brownian motion via barrier policies, including optimal control of Brownian storage systems; and Brownian models of dynamic inference, also called Brownian learning models, or Brownian filtering models.

J. MICHAEL HARRISON has developed and analyzed stochastic models in several different domains related to business, including mathematical finance and processing network theory. His current research is focused on dynamic models of resource sharing, and on the application of stochastic control theory in economics and operations. Professor Harrison has been honored by the Institute for Operations Research and Management Science (INFORMS) with its Expository Writing Award (1998), the Lanchester Prize for best research publication (2001), and the John von Neumann Theory Prize (2004); he was elected to the National Academy of Engineering in 2008. He is a Fellow of INFORMS and of the Institute for Mathematical Statistics.

Preface

This is an expanded and updated version of a book that I published in 1985 with John Wiley and Sons, titled *Brownian Motion and Stochastic Flow Systems*. Like the original, it fits comfortably under the heading of "applied probability," its primary subjects being (i) stochastic system models based on Brownian motion, and (ii) associated methods of stochastic analysis.

Here the word "system" is used in the engineer's or economist's sense, referring to equipment, people, and operating procedures that work together with some economic purpose, broadly construed. Examples include telephone call centers, manufacturing networks, cash management operations, and data storage centers. This book emphasizes *dynamic stochastic models* of such man-made systems, that is, models in which system status evolves over time, subject to unpredictable factors like weather, demand shocks, or mechanical failures. Some of the models considered in the book are purely descriptive in nature, aimed at estimating performance characteristics like long-run average inventory or expected discounted cost, given a fixed set of system characteristics. Other models are explicitly aimed at optimizing some measure of system performance, especially through the exercise of a dynamic control capability.

Brownian models of system evolution and dynamic control are increasingly popular in economics and engineering, and in allied business fields like finance and operations. The reason for that popularity is mathematical tractability: In one instance after another, researchers working with Brownian models have been able to derive explicit solutions and clear-cut insights that were unobtainable using conventional models. In the ways that really matter, then, Brownian models provide the *simplest* possible representations of dynamic, stochastic phenomena.

Old and new content Chapters 1, 2, 3, 4, 6, and 7 of this book correspond to chapters in the 1985 original. The last of those six "old" chapters has been substantially revised and expanded; the other five have been revised in less substantial ways. In Chapters 1 and 3 the basic properties of Brownian motion are summarized and various standard formulas are derived. Chapter 4 is devoted to the Itô calculus for Brownian motion, emphasizing Itô's formula and its various generalizations.

Chapters 2 and 6 are concerned with descriptive models of stochastic storage systems, in which an input flow and an output flow are decoupled by an intermediate storage buffer: queuing models, inventory models, and cash balance models all describe systems of that kind. Chapter 2 develops some foundational theory, and Chapter 6 is specifically concerned with storage system models based on Brownian motion, called *Brownian storage models* or *Brownian storage systems* for brevity. Formulas are developed in Chapter 6 for various standard performance measures, taking the system parameters as given.

Four dynamic control problems associated with Brownian storage models are considered in Chapter 7: optimal policies are derived under different cost structures, using both discounted and average cost optimality criteria. The subject matter of Chapter 7, like that of Chapters 2 and 6, is traditionally associated with operations research, but the monographs by Dixit (1993) and Stokey (2009) show that it is also of interest in economics.

Chapters 5, 8, and 9 are new. The first one addresses optimal stopping problems for Brownian motion, including the influential investment model of McDonald and Siegel (1986). It illustrates the guess-and-verify approach that is ubiquitous in applications, and also summarizes some insightful general theory. Chapter 8 is concerned with Brownian models of dynamic inference, in which one observes a Brownian motion whose drift rate is initially unknown, or more generally, may depend on the state of an unobserved underlying process. The problem is to make inferences about the unknown parameter or unobserved process, given the Brownian path. Learning models of this kind have a long history in statistics, where the terms "sequential analysis" and "sequential detection" are used, and also in engineering, where the central problem is described as one of "filtering" information about a parameter of interest from the Brownian noise with which it is confounded. Brownian learning models also arise increasingly in economics, especially in dynamic investment theory. Finally, Chapter 9 treats a diverse collection of examples, some involving particular applications and others more methodological in character, that further develop themes introduced in earlier chapters.

The distinctive feature of my 1985 book was its combination of compactness and concreteness: a narrow focus and brisk pace; many concrete formulas and explicit calculations; and a minimum of abstract notation. I have made every effort to preserve that aspect of the original, and in particular, to include only such elements of general theory as are needed for the applications considered.

Intended audience This book is intended for researchers and advanced graduate students in economics, engineering, and operations research. As mathematical prerequisites, readers are assumed to have knowledge of elementary real analysis, including Riemann–Stieltjes integration, at the level of Bartle (1976), and of measure theoretic probability, including conditional expectation, at the level of Billingsley (1995). However, I have tried to make the book accessible to readers who may lack some of the prerequisite knowledge nominally assumed. Certain essential results from probability theory and real analysis are collected in the appendices, and many important definitions are reviewed in the text. As stated in the introduction of the 1985 original, "mathematically able readers who have at least a nodding acquaintance with σ-algebras will be able to get by... I hope this book will be immediately useful to readers with limited mathematical background, and may also serve to stimulate and guide further study." The book is aimed at non-mathematicians whose goal is to build and analyze stochastic models.

Reflected versus regulated Brownian motion A substantial portion of this book is devoted to a process that is called "reflected Brownian motion" in stochastic process theory. In my 1985 book I proposed the alternative name "regulated Brownian motion," arguing that the word "reflection" is confusing in this context. I used the newly coined name throughout the 1985 book, and a number of authors, especially in economics, have adopted my alternative terminology in their own work. However, having won very few converts outside economics, I am reluctantly reverting to the traditional terminology in this revision. The reason for that choice is nicely summarized in the following statement, excerpted from an anonymous review of the book proposal that I submitted several years ago to Cambridge University Press:

> One important purpose of a textbook is to prepare the reader for the research literature, and for better or for worse "reflected Brownian motion" is the universally accepted standard terminology. It would be a good idea in revising the book to conform to the standard terminology: the benefits for the reader would outweigh the aesthetic appeal of a more imaginative but nonstandard language.

Having decided in favor of "reflected Brownian motion" for this revision, I have also adopted related terminology like "reflection mapping" and "reflecting barrier." Despite that capitulation, the reason for my attempted revolt is worth repeating. For that purpose let $X = \{X_t,\ t \geq 0\}$ be a standard Brownian motion (zero drift and unit variance, starting at the origin) and then define

$$(0.1) \qquad Z_t := X_t - \inf_{0 \leq s \leq t} X_s, \qquad t \geq 0.$$

It has long been known that this new process Z has the same distribution as Y, where

$$(0.2) \qquad Y_t := |X_t|, \qquad t \geq 0.$$

Of course, "reflected Brownian motion" is a perfectly good name for Y, and mathematicians understandably felt that (0.2) was a more natural definition than (0.1), so Z came to be known as "an alternative representation of reflected Brownian motion." But the word "reflection" does not describe well the mapping embodied in (0.1), and it is this mapping with which one begins in applications (see Chapter 2). Moreover, we are generally interested in the situation where X is a Brownian motion with drift. Then Y and Z do *not* have the same distribution, but still Z is called "reflected Brownian motion." This terminology has even been extended to higher dimensions, where one encounters mysterious phrases like "Brownian motion with oblique reflection at the boundary." (Problem 6.13 describes a process that is usually characterized in this way.) Because the word "reflection" has the connotation of a symmetric "folding over," this terminology is potentially confusing, to say the least, but we seem to be stuck with it.

Emphasis on Itô calculus With respect to mathematical methods, this book emphasizes Itô stochastic calculus. If one has a probabilistic model based on Brownian motion, such as the process Z defined by (0.1), then all the interesting associated quantities will be solutions of certain differential equations. For example, in this book I wish to compute expected discounted costs for various processes as functions of the starting state. To calculate such a quantity, what differential equation must be solved, and what are the appropriate boundary conditions?

Using Itô's formula, such questions can be answered systematically, which allows one to recast the original problem in purely analytic terms. Many problems can be solved by direct probabilistic means, such as the martingale methods of Chapter 3, but to solve really hard problems it is necessary to have command of both probabilistic and analytic methods.

One of my primary objectives in writing the original 1985 book was to show exactly why and how Itô's formula is so useful for solving concrete problems. Chapters 4, 6, and 7, together with their problems, have been structured with this goal in mind. I hope that even readers who have no intrinsic interest in models of buffered stochastic flow will find that the applications discussed in Chapters 6 and 7 enrich their appreciation for the general theory.

A note on organization Readers are advised to begin with at least a quick look at the appendices. These serve not only to review prerequisite results but also to set notation and terminology. At the end of the book, just before the index, there is a list of all works cited in the text, including the pages on which they are cited. I have made no attempt to compile a comprehensive set of references on any of the subjects covered in the book, nor to suggest the relative contributions of different authors through frequency of citation.

Numbering conventions There is a single numbering system for enunciations (lemmas, theorems, etc.) within each chapter. Thus, for example, the first three enunciations of Chapter n could be Lemma $n.1$, Definition $n.2$, and Proposition $n.3$. Equations in Chapter n are numbered ($n.1$), ($n.2$), etc. Similar numbering is used for enunciations and equations in the two appendices.

Acknowledgments My initial exposure to much of the material in this book came in graduate courses from David Siegmund and Donald Iglehart, and in later interactions with David Kreps, Larry Shepp, and Rick Durrett. Erhan Çinlar read an initial draft of the 1985 original and made many helpful comments, as did Avi Mandelbaum and Ruth Williams on portions of later drafts. The original version of the book was also influenced by comments from students in courses I taught during the 1980s, including Peter Glynn, Bill Peterson, Richard Pitbladdo, Tom Sellke, and Ruth Williams.

The idea for this revised and expanded version came originally from Jan Van Mieghem, and in its preparation I have benefited from consultations with many colleagues at Stanford and elsewhere, including Bariş Ata, Jim Dai, Peter DeMarzo, Darrell Duffie, Brad Efron, Brett Green, Ioannis Karatzas, Andrzej Skrzypacz, Ilya Strebulaev, Ruth Williams, and Assaf Zeevi. Among those colleagues Bariş Ata deserves pride of place, having read the entire book and provided helpful suggestions on several different levels.

In the spring of 2011 I taught a graduate course at Stanford in which Joel Goh, Dmitry Orlov, Dmitry Smelov, Han-I Su, Nur Sunar, Felipe Varas Green, and Pavel Zryumov made presentations that significantly influenced the development of this book, in terms of both the new topics covered (or not covered) and the way in which those topics are treated. Finally, I am indebted to Cindy Kirby, who typed the entire book in LaTeX and expertly managed its format, and to the Stanford Graduate School of Business for financial support of this project.

Guide to Notation and Terminology

The expression $A := B$ means that A is equal to B as a matter of definition. In some sentences, the expression should be read "A, which is equal by definition to B, …." Conditional expectations are defined only up to an equivalence. Equations involving conditional expectation, or any other random variables, should be interpreted in the almost sure sense. The terms *positive* and *increasing* are used in the weak sense, as opposed to *strictly positive* and *strictly increasing*. The equation

$$P\{X \in dx\} = f(x)\,dx$$

means that f is a density function for the random variable X. That is,

$$P\{X \in A\} = \int_A f(x)\,dx$$

for any Borel set A. In the usual way, 1_A denotes the indicator function of a set A, which equals 1 on A and equals zero elsewhere. If $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$, then 1_A is described as an indicator random variable or as the indicator of event A. To specify the time at which a stochastic process X is observed, I may write either X_t or $X(t)$ depending on the situation. On esthetic grounds, I prefer the former notation, but the latter is superior when one must write expressions like $X(T_1 + T_2)$.

Let I be an interval subset of $\mathbb{R}$ (the real line). We say that a function $f : I \to \mathbb{R}$ is C^1 (or less commonly, that f belongs to C^1) if it is continuously differentiable on the interior of I, and moreover, $f'(\cdot)$ approaches a finite limit at each closed endpoint of I, if there are any. (This state of affairs is often expressed by saying that f is continuously differentiable *up to the boundary*.) Similarly, a C^2 function on I is twice continuously differentiable on the interior of I, and its first and second derivative approach finite limits at each closed endpoint, if there are any.

Continuing in this same vein, we say that $f : I \to \mathbb{R}$ is *piecewise* C^1 if it is continuously differentiable except at finitely many interior points, and moreover, $f'(\cdot)$ has finite left and right limits at each of the exceptional points, as well as a finite limit at each closed endpoint. A piecewise C^2 function is defined similarly, with both the first and second derivatives having finite left and right limits at each exceptional point, as well as finite limits at each closed endpoint. Thus, if f is piecewise C^2 on a compact interval subset of $\mathbb{R}$, both $f'(\cdot)$ and $f''(\cdot)$ are bounded.

Let f be an increasing continuous function on $[0, \infty)$. We say that f increases at a point $t > 0$ if $f(t + \epsilon) > f(t - \epsilon)$ for all $\epsilon > 0$. In this case, t is said to be a *point of increase* for f. Now let g be another continuous function on $[0, \infty)$ and consider the statement

$$f \text{ increases only when } g = 0.$$

This means that $g_t = 0$ at every point t where f increases. Many such statements appear in this book, and readers will find this terminology to be efficient if somewhat cryptic.

The last section of Appendix B discusses notational conventions for Riemann–Stieltjes integrals. As the reader will see, my general rule is to suppress the arguments of functions appearing in such integrals whenever possible. The same guiding principle is used in Chapters 4 to 6 with respect to stochastic integrals. For example, I write

$$\int_0^t X\,dW \quad \text{rather than} \quad \int_0^t X(s)\,dW(s)$$

to denote the stochastic integral of a process X with respect to a Brownian motion W. The former notation is certainly more economical, and it is also more correct mathematically, but my slavish adherence to the guiding principle may occasionally cause confusion. As an extreme example, consider the expression

$$\int_0^T e^{-\lambda t}(\Gamma - \lambda) f(Z)\,dg(X + L - U)$$

where λ is a constant, Γ is a differential operator, f and g are functions, and Z, X, L, and U are processes. This signifies the stochastic integral over $[0, T]$ of a process that has value $\exp(-\lambda t)[\Gamma f(Z_t) - \lambda f(Z_t)]$ at time t with respect to a process that has value $g(X_t + L_t - U_t)$ at time t.

The following is a list of symbols that are used with a single meaning, or at least with one dominant meaning, throughout the book. Section numbers, when given, locate either the definition of the symbol or the point of its first appearance (assuming that the appendices are read first).

$\square$	end of proof
$\wedge$ and $\vee$	maximum and minimum
$x^+ := x \vee 0$	positive part of x
$x^- := -(x \wedge 0)$	negative part of x
$\mathbb{R}$	the real line
C^1 and C^2	see text immediately above
$\mathbb{F} = \{\mathcal{F}_t,\ t \geq 0\}$	filtration (Section A.1)
$\mathcal{B}$	Borel σ-algebra on $\mathbb{R}$ (Section A.2)
$\mathcal{B}[0,\infty)$	Borel σ-algebra on $[0,\infty)$ (Section A.2)
$C := C[0,\infty)$	Section A.2
$\mathcal{C}$	Borel σ-algebra on C (Section A.2)
$\mathcal{N}(\mu,\sigma^2)$	normal distribution (Section 1.1)
$V_\beta(t)$	Wald martingale (Section 1.4)
$\Phi(x)$	$\mathcal{N}(0,1)$ distribution function (Section 1.7)
P_x and E_x	Section 3.1
$\Gamma f := \mu f' + \frac{1}{2}\sigma^2 f''$	Section 3.3
$\alpha_1(\lambda)$ and $\alpha_2(\lambda)$	Section 3.3
$\psi_1(x)$ and $\psi_2(x)$	Section 3.3
$\theta_1(x)$ and $\theta_2(x)$	Section 3.3
$\mathrm{E}(X;A)$	partial expectation (Section 3.3)
H	Section 4.1
$I_t(X)$	stochastic integral (Section 4.1)
H^2	Section 4.2
L^2	Section 4.2
S^2	Section 4.2
RCLL	right-continuous with left limits (Section 4.8)

Contents