Probability with Martingales

Cambridge University Press
978-0-521-40605-5 - Probability with Martingales
David Williams
Frontmatter
More information

Probability with Martingales

David Williams
Statistical Laboratory, DPMMS
Cambridge University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

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/9780521406055

First published 1991
18th printing 2015

Printed in the United Kingdom by Clays, St Ives plc.

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

ISBN 978-0-521-40605-5 Paperback

Cambridge University Press
978-0-521-40605-5 - Probability with Martingales
David Williams
Frontmatter
More information

Contents

PART C: CHARACTERISTIC FUNCTIONS

Preface – please read!

The most important chapter in this book is *Chapter E: Exercises.* I have left the interesting things for *you* to do. You can start *now* on the 'EG' exercises, but see 'More about exercises' later in this Preface.

The book, which is essentially the set of lecture notes for a third-year undergraduate course at Cambridge, is as lively an introduction as I can manage to the rigorous theory of probability. Since much of the book is devoted to martingales, it is bound to become very lively: look at those Exercises on Chapter 10! But, of course, there is that initial plod through the measure-theoretic foundations. It must be said however that measure theory, that most arid of subjects when done for its own sake, becomes amazingly more alive when used in probability, not only because it is then applied, but also because it is immensely enriched.

You cannot avoid measure theory: an *event* in probability is a measurable set, a *random variable* is a measurable function on the sample space, the *expectation* of a random variable is its integral with respect to the probability measure; and so on. To be sure, one can take some central results from measure theory as axiomatic in the main text, giving careful proofs in appendices; and indeed that is exactly what I have done.

Measure theory for its own sake is based on the fundamental addition rule for measures. Probability theory supplements that with the multiplication rule which describes independence; and things are already looking up. But what really enriches and enlivens things is that we deal with lots of σ-algebras, not just the one σ-algebra which is the concern of measure theory.

In planning this book, I decided for every topic what things I considered just a bit too advanced, and, often with sadness, I have ruthlessly omitted them.

For a more thorough training in many of the topics covered here, see Billingsley (1979), Chow and Teicher (1978), Chung (1968), Kingman and

Taylor (1966), Laha and Rohatgi (1979), and Neveu (1965). As regards measure theory, I learnt it from Dunford and Schwartz (1958) and Halmos (1959). After reading this book, you must read the still-magnificent Breiman (1968), and, for an excellent indication of what can be done with discrete martingales, Hall and Heyde (1980).

Of course, intuition is much more important than knowledge of measure theory, and you should take every opportunity to sharpen your intuition. There is no better whetstone for this than Aldous (1989), though it is a very demanding book. For appreciating the scope of probability and for learning how to think about it, Karlin and Taylor (1981), Grimmett and Stirzaker (1982), Hall (1988), and Grimmett's recent superb book, Grimmett (1989), on percolation are strongly recommended.

More about exercises. In compiling Chapter E, which consists exactly of the homework sheet I give to the Cambridge students, I have taken into account the fact that this book, like any other mathematics book, implicitly contains a vast number of other exercises, many of which are easier than those in Chapter E. I refer of course to the exercises *you* create by reading the statement of a result, and then trying to prove it for yourself, before you read the given proof. One other point about exercises: you will, for example, surely forgive my using expectation E in Exercises on Chapter 4 before E is treated with full rigour in Chapter 6.

Acknowledgements. My first thanks must go to the students who have endured the course on which the book is based and whose quality has made me try hard to make it worthy of them; and to those, especially David Kendall, who had developed the course before it became my privilege to teach it. My thanks to David Tranah and other staff of CUP for their help in converting the course into this book. Next, I must thank Ben Garling, James Norris and Chris Rogers without whom the book would have contained more errors and obscurities. (The many faults which surely remain in it are my responsibility.) Helen Rutherford and I typed part of the book, but the vast majority of it was typed by Sarah Shea-Simonds in a virtuoso performance worthy of Horowitz. My thanks to Helen and, most especially, to Sarah. Special thanks to my wife, Sheila, too, for all her help.

But my best thanks – and yours if you derive any benefit from the book – must go to three people whose names appear in capitals in the Index: J.L. Doob, A.N. Kolmogorov and P. Lévy: without them, there wouldn't have been much to write about, as Doob (1953) splendidly confirms.

Statistical Laboratory,
Cambridge

David Williams
October 1990

A Question of Terminology

Random variables: functions or equivalence classes?

At the level of this book, the theory would be more 'elegant' if we regarded a random variable as an *equivalence class* of measurable functions on the sample space, two functions belonging to the same equivalence class if and only if they are equal almost everywhere. Then the conditional-expectation map

$$X \mapsto \mathsf{E}(X|\mathcal{G})$$

would be a truly well-defined contraction map from $L^p(\Omega, \mathcal{F}, \mathsf{P})$ to $L^p(\Omega, \mathcal{G}, \mathsf{P})$ for $p \geq 1$; and we would not have to keep mentioning versions (representatives of equivalence classes) and would be able to avoid the endless 'almost surely' qualifications.

I have however chosen the 'inelegant' route: firstly, I prefer to work with *functions*, and confess to preferring

$$4 + 5 = 2 \bmod 7 \qquad \text{to} \qquad [4]_7 + [5]_7 = [2]_7.$$

But there is a substantive reason. I hope that this book will tempt you to progress to the much more interesting, and more important, theory where the parameter set of our process is uncountable (e.g. it may be the time-parameter set $[0, \infty)$). There, the equivalence-class formulation just will not work: the 'cleverness' of introducing quotient spaces loses the subtlety which is essential even for formulating the fundamental results on existence of continuous modifications, etc., unless one performs contortions which are hardly elegant. Even if these contortions allow one to *formulate* results, one would still have to use genuine functions to *prove* them; so where does the reality lie?!

A Guide to Notation

▶ signifies something important, ▶▶ something very important, and ▶▶▶ the Martingale Convergence Theorem.

I use ':=' to signify 'is defined to equal'. This Pascal notation is particularly convenient because it can also be used in the reversed sense.

I use analysts' (as opposed to category theorists') conventions:

▶ $$\mathbf{N} := \{1, 2, 3, \ldots\} \subseteq \{0, 1, 2, \ldots\} =: \mathbf{Z}^+.$$

Everyone is agreed that $\mathbf{R}^+ := [0, \infty)$.

For a set B contained in some universal set S, I_B denotes the indicator function of B: that is $I_B : S \to \{0, 1\}$ and

$$I_B(s) := \begin{cases} 1 & \text{if } s \in B, \\ 0 & \text{otherwise.} \end{cases}$$

For $a, b \in \mathbf{R}$,

$$a \wedge b := \min(a, b), \qquad a \vee b := \max(a, b).$$

CF:characteristic function; DF: distribution function; pdf: probability density function.

σ-algebra, $\sigma(\mathcal{C})$ (1.1); $\sigma(Y_\gamma : \gamma \in \mathcal{C})$ (3.8, 3.13). π-system (1.6); d-system (A1.2).

---

a.e.:	almost everywhere (1.5)
a.s.:	almost surely (2.4)
bΣ:	the space of bounded Σ-measurable functions (3.1)

$\mathcal{B}(S)$:	the Borel σ-algebra on S, $\mathcal{B} := \mathcal{B}(\mathbf{R})$ (1.2)
$C \bullet X$:	discrete stochastic integral (10.6)
$d\lambda/d\mu$:	Radon-Nikodým derivative (5.14)
$d\mathsf{Q}/d\mathsf{P}$:	Likelihood Ratio (14.13)
$\mathsf{E}(X)$:	expectation $\mathsf{E}(X) := \int_\Omega X(\omega)\mathsf{P}(d\omega)$ of X (6.3)
$\mathsf{E}(X; F)$:	$\int_F X d\mathsf{P}$ (6.3)
$\mathsf{E}(X\|\mathcal{G})$:	conditional expectation (9.3)
(E_n, ev):	$\liminf E_n$ (2.8)
$(E_n, \text{i.o.})$:	$\limsup E_n$ (2.6)
f_X:	probability density function (pdf) of X (6.12).
$f_{X,Y}$:	joint pdf (8.3)
$f_{X\|Y}$:	conditional pdf (9.6)
F_X:	distribution function of X (3.9)
lim inf:	for sets, (2.8)
lim sup:	for sets, (2.6)
$x = \uparrow \lim x_n$:	$x_n \uparrow x$ in that $x_n \le x_{n+1}$ $(\forall n)$ and $x_n \to x$.
log:	natural (base e) logarithm
$\mathcal{L}_X$, Λ_X:	law of X (3.9)
$\mathcal{L}^p$, L^p:	Lebesgue spaces (6.7, 6.13)
Leb:	Lebesgue measure (1.8)
$\mathrm{m}\Sigma$:	space of Σ-measurable functions (3.1)
M^T:	process M stopped at time T (10.9)
$\langle M \rangle$:	angle-brackets process (12.12)
$\mu(f)$:	integral of f with respect to μ (5.0, 5.2)
$\mu(f; A)$:	$\int_A f d\mu$ (5.0, 5.2)
φ_X:	CF of X (Chapter 16)
φ:	pdf of standard normal N(0,1) distribution
Φ:	DF of N(0,1) distribution
X^T:	X stopped at time T (10.9)