

Cambridge University Press
978-0-521-40605-5 - Probability with Martingales
David Williams
Frontmatter
[More information](#)

Probability with Martingales

Cambridge University Press
978-0-521-40605-5 - Probability with Martingales
David Williams
Frontmatter
[More information](#)

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
978-0-521-40605-5 - Probability with Martingales
David Williams
Frontmatter
[More information](#)

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

© Cambridge University Press 1991

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

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 has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface – please read!	xi
A Question of Terminology	xiii
A Guide to Notation	xiv
Chapter 0: A Branching-Process Example	1
0.0. Introductory remarks. 0.1. Typical number of children, X . 0.2. Size of n^{th} generation, Z_n . 0.3. Use of conditional expectations. 0.4. Extinction probability, π . 0.5. Pause for thought: measure. 0.6. Our first martingale. 0.7. Convergence (or not) of expectations. 0.8. Finding the distribution of M_∞ . 0.9. Concrete example.	
PART A: FOUNDATIONS	
Chapter 1: Measure Spaces	14
1.0. Introductory remarks. 1.1. Definitions of algebra, σ -algebra. 1.2. Examples. Borel σ -algebras, $\mathcal{B}(S)$, $\mathcal{B} = \mathcal{B}(\mathbf{R})$. 1.3. Definitions concerning set functions. 1.4. Definition of measure space. 1.5. Definitions concerning measures. 1.6. Lemma. Uniqueness of extension, π -systems. 1.7. Theorem. Carathéodory's extension theorem. 1.8. Lebesgue measure Leb on $((0, 1], \mathcal{B}(0, 1])$. 1.9. Lemma. Elementary inequalities. 1.10. Lemma. Monotone-convergence properties of measures. 1.11. Example/Warning.	
Chapter 2: Events	23
2.1. Model for experiment: $(\Omega, \mathcal{F}, \mathbf{P})$. 2.2. The intuitive meaning. 2.3. Examples of (Ω, \mathcal{F}) pairs. 2.4. Almost surely (a.s.) 2.5. Reminder: $\limsup, \liminf, \downarrow \lim$, etc. 2.6. Definitions. $\limsup E_n, (E_n, \text{i.o.})$. 2.7.	

First Borel-Cantelli Lemma (BC1). 2.8. Definitions. $\liminf E_n, (E_n, ev)$.
 2.9. Exercise.

Chapter 3: Random Variables 29

3.1. Definitions. Σ -measurable function, $m\Sigma, (m\Sigma)^+, b\Sigma$. 3.2. Elementary Propositions on measurability. 3.3. Lemma. Sums and products of measurable functions are measurable. 3.4. Composition Lemma. 3.5. Lemma on measurability of infs, liminfs of functions. 3.6. Definition. Random variable. 3.7. Example. Coin tossing. 3.8. Definition. σ -algebra generated by a collection of functions on Ω . 3.9. Definitions. Law, Distribution Function. 3.10. Properties of distribution functions. 3.11. Existence of random variable with given distribution function. 3.12. Skorokod representation of a random variable with prescribed distribution function. 3.13. Generated σ -algebras – a discussion. 3.14. The Monotone-Class Theorem.

Chapter 4: Independence 38

4.1. Definitions of independence. 4.2. The π -system Lemma; and the more familiar definitions. 4.3. Second Borel-Cantelli Lemma (BC2). 4.4. Example. 4.5. A fundamental question for modelling. 4.6. A coin-tossing model with applications. 4.7. Notation: IID RVs. 4.8. Stochastic processes; Markov chains. 4.9. Monkey typing Shakespeare. 4.10. Definition. Tail σ -algebras. 4.11. Theorem. Kolmogorov's 0-1 law. 4.12. Exercise/Warning.

Chapter 5: Integration 49

5.0. Notation, etc. $\mu(f) := \int f d\mu, \mu(f; A)$. 5.1. Integrals of non-negative simple functions, SF^+ . 5.2. Definition of $\mu(f), f \in (m\Sigma)^+$. 5.3. Monotone-Convergence Theorem (MON). 5.4. The Fatou Lemmas for functions (FATOU). 5.5. 'Linearity'. 5.6. Positive and negative parts of f . 5.7. Integrable function, $\mathcal{L}^1(S, \Sigma, \mu)$. 5.8. Linearity. 5.9. Dominated Convergence Theorem (DOM). 5.10. Scheffé's Lemma (SCHEFFÉ). 5.11. Remark on uniform integrability. 5.12. The standard machine. 5.13. Integrals over subsets. 5.14. The measure $f\mu, f \in (m\Sigma)^+$.

Chapter 6: Expectation 58

Introductory remarks. 6.1. Definition of expectation. 6.2. Convergence theorems. 6.3. The notation $E(X; F)$. 6.4. Markov's inequality. 6.5. Sums of non-negative RVs. 6.6. Jensen's inequality for convex functions. 6.7. Monotonicity of \mathcal{L}^p norms. 6.8. The Schwarz inequality. 6.9. \mathcal{L}^2 : Pythagoras, covariance, etc. 6.10. Completeness of \mathcal{L}^p ($1 \leq p < \infty$). 6.11. Orthogonal projection. 6.12. The 'elementary formula' for expectation. 6.13. Hölder from Jensen.

Contents

vii

Chapter 7: An Easy Strong Law 71

7.1. ‘Independence means multiply’ – again! 7.2. Strong Law – first version.
 7.3. Chebyshev’s inequality. 7.4. Weierstrass approximation theorem.

Chapter 8: Product Measure 75

8.0. Introduction and advice. 8.1. Product measurable structure, $\Sigma_1 \times \Sigma_2$.
 8.2. Product measure, Fubini’s Theorem. 8.3. Joint laws, joint pdfs. 8.4.
 Independence and product measure. 8.5. $\mathcal{B}(\mathbf{R})^n = \mathcal{B}(\mathbf{R}^n)$. 8.6. The n -fold
 extension. 8.7. Infinite products of probability triples. 8.8. Technical note
 on the existence of joint laws.

PART B: MARTINGALE THEORY

Chapter 9: Conditional Expectation 83

9.1. A motivating example. 9.2. Fundamental Theorem and Definition
 (Kolmogorov, 1933). 9.3. The intuitive meaning. 9.4. Conditional ex-
 pectation as least-squares-best predictor. 9.5. Proof of Theorem 9.2. 9.6.
 Agreement with traditional expression. 9.7. Properties of conditional ex-
 pectation: a list. 9.8. Proofs of the properties in Section 9.7. 9.9. Regular
 conditional probabilities and pdfs. 9.10. Conditioning under independence
 assumptions. 9.11. Use of symmetry: an example.

Chapter 10: Martingales 93

10.1. Filtered spaces. 10.2. Adapted processes. 10.3. Martingale, super-
 martingale, submartingale. 10.4. Some examples of martingales. 10.5. Fair
 and unfair games. 10.6. Previsible process, gambling strategy. 10.7. A fun-
 damental principle: you can’t beat the system! 10.8. Stopping time. 10.9.
 Stopped supermartingales are supermartingales. 10.10. Doob’s Optional-
 Stopping Theorem. 10.11. Awaiting the almost inevitable. 10.12. Hitting
 times for simple random walk. 10.13. Non-negative superharmonic func-
 tions for Markov chains.

Chapter 11: The Convergence Theorem 106

11.1. The picture that says it all. 11.2. Upcrossings. 11.3. Doob’s Upcross-
 ing Lemma. 11.4. Corollary. 11.5. Doob’s ‘Forward’ Convergence Theorem.
 11.6. Warning. 11.7. Corollary.

Chapter 12: Martingales bounded in \mathcal{L}^2	110
12.0. Introduction. 12.1. Martingales in \mathcal{L}^2 : orthogonality of increments. 12.2. Sums of zero-mean independent random variables in \mathcal{L}^2 . 12.3. Random signs. 12.4. A symmetrization technique: expanding the sample space. 12.5. Kolmogorov's Three-Series Theorem. 12.6. Cesàro's Lemma. 12.7. Kronecker's Lemma. 12.8. A Strong Law under variance constraints. 12.9. Kolmogorov's Truncation Lemma. 12.10. Kolmogorov's Strong Law of Large Numbers (SLLN). 12.11. Doob decomposition. 12.12. The angle-brackets process $\langle M \rangle$. 12.13. Relating convergence of M to finiteness of $\langle M \rangle_\infty$. 12.14. A trivial 'Strong Law' for martingales in \mathcal{L}^2 . 12.15. Lévy's extension of the Borel-Cantelli Lemmas. 12.16. Comments.	
Chapter 13: Uniform Integrability	126
13.1. An 'absolute continuity' property. 13.2. Definition. UI family. 13.3. Two simple sufficient conditions for the UI property. 13.4. UI property of conditional expectations. 13.5. Convergence in probability. 13.6. Elementary proof of (BDD). 13.7. A necessary and sufficient condition for \mathcal{L}^1 convergence.	
Chapter 14: UI Martingales	133
14.0. Introduction. 14.1. UI martingales. 14.2. Lévy's 'Upward' Theorem. 14.3. Martingale proof of Kolmogorov's 0-1 law. 14.4. Lévy's 'Downward' Theorem. 14.5. Martingale proof of the Strong Law. 14.6. Doob's Submartingale Inequality. 14.7. Law of the Iterated Logarithm: special case. 14.8. A standard estimate on the normal distribution. 14.9. Remarks on exponential bounds; large deviation theory. 14.10. A consequence of Hölder's inequality. 14.11. Doob's \mathcal{L}^p inequality. 14.12. Kakutani's Theorem on 'product' martingales. 14.13. The Radon-Nikodým theorem. 14.14. The Radon-Nikodým theorem and conditional expectation. 14.15. Likelihood ratio; equivalent measures. 14.16. Likelihood ratio and conditional expectation. 14.17. Kakutani's Theorem revisited; consistency of LR test. 14.18. Note on Hardy spaces, etc.	
Chapter 15: Applications	153
15.0. Introduction – please read! 15.1. A trivial martingale-representation result. 15.2. Option pricing; discrete Black-Scholes formula. 15.3. The Mabinogion sheep problem. 15.4. Proof of Lemma 15.3(c). 15.5. Proof of result 15.3(d). 15.6. Recursive nature of conditional probabilities. 15.7. Bayes' formula for bivariate normal distributions. 15.8. Noisy observation of a single random variable. 15.9. The Kalman-Bucy filter. 15.10. Harnesses entangled. 15.11. Harnesses unravelled, 1. 15.12. Harnesses unravelled, 2.	

PART C: CHARACTERISTIC FUNCTIONS

Chapter 16: Basic Properties of CFs 172

16.1. Definition. 16.2. Elementary properties. 16.3. Some uses of characteristic functions. 16.4. Three key results. 16.5. Atoms. 16.6. Lévy's Inversion Formula. 16.7. A table.

Chapter 17: Weak Convergence 179

17.1. The 'elegant' definition. 17.2. A 'practical' formulation. 17.3. Skorokhod representation. 17.4. Sequential compactness for $\text{Prob}(\bar{\mathbf{R}})$. 17.5. Tightness.

Chapter 18: The Central Limit Theorem 185

18.1. Lévy's Convergence Theorem. 18.2. o and O notation. 18.3. Some important estimates. 18.4. The Central Limit Theorem. 18.5. Example. 18.6. CF proof of Lemma 12.4.

APPENDICES

Chapter A1: Appendix to Chapter 1 192

A1.1. A non-measurable subset A of S^1 . A1.2. d -systems. A1.3. Dynkin's Lemma. A1.4. Proof of Uniqueness Lemma 1.6. A1.5. λ -sets: 'algebra' case. A1.6. Outer measures. A1.7. Carathéodory's Lemma. A1.8. Proof of Carathéodory's Theorem. A1.9. Proof of the existence of Lebesgue measure on $((0, 1], \mathcal{B}(0, 1])$. A1.10. Example of non-uniqueness of extension. A1.11. Completion of a measure space. A1.12. The Baire category theorem.

Chapter A3: Appendix to Chapter 3 205

A3.1. Proof of the Monotone-Class Theorem 3.14. A3.2. Discussion of generated σ -algebras.

Chapter A4: Appendix to Chapter 4 208

A4.1. Kolmogorov's Law of the Iterated Logarithm. A4.2. Strassen's Law of the Iterated Logarithm. A4.3. A model for a Markov chain.

Chapter A5: Appendix to Chapter 5 211

A5.1. Doubly monotone arrays. A5.2. The key use of Lemma 1.10(a). A5.3. 'Uniqueness of integral'. A5.4. Proof of the Monotone-Convergence Theorem.

Chapter A9: Appendix to Chapter 9	214
A9.1. Infinite products: setting things up. A9.2. Proof of A9.1(e).	
Chapter A13: Appendix to Chapter 13	217
A13.1. Modes of convergence: definitions. A13.2. Modes of convergence: relationships.	
Chapter A14: Appendix to Chapter 14	219
A14.1. The σ -algebra \mathcal{F}_T , T a stopping time. A14.2. A special case of OST. A14.3. Doob's Optional-Sampling Theorem for UI martingales. A14.4. The result for UI submartingales.	
Chapter A16: Appendix to Chapter 16	222
A16.1. Differentiation under the integral sign.	
Chapter E: Exercises	224
References	243
Index	246

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

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

would be a truly well-defined contraction map from $L^p(\Omega, \mathcal{F}, \mathbf{P})$ to $L^p(\Omega, \mathcal{G}, \mathbf{P})$ 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

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

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:

$$\blacktriangleright \quad \mathbf{N} := \{1, 2, 3, \dots\} \subseteq \{0, 1, 2, \dots\} =: \mathbf{Z}^+.$$

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 $I_B : S \rightarrow \{0, 1\}$ and

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

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

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

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)

$b\Sigma$: the space of bounded Σ -measurable functions (3.1)

A Guide to Notation

xv

$\mathcal{B}(S)$:	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)
dQ/dP :	Likelihood Ratio (14.13)
$E(X)$:	expectation $E(X) := \int_{\Omega} X(\omega)P(d\omega)$ of X (6.3)
$E(X; F)$:	$\int_F X dP$ (6.3)
$E(X \mathcal{G})$:	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)
\liminf :	for sets, (2.8)
\limsup :	for sets, (2.6)
$x = \uparrow \lim x_n$:	$x_n \uparrow x$ in that $x_n \leq x_{n+1}$ ($\forall n$) and $x_n \rightarrow x$.
\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)
$m\Sigma$:	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)