

THE  
THEORY  
OF  
PROBABILITY

EXPLORATIONS AND APPLICATIONS



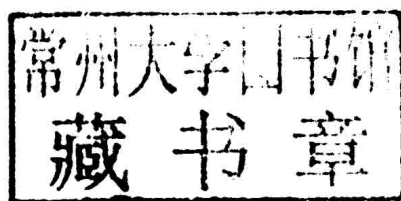
Santosh S. Venkatesh

CAMBRIDGE

# THE THEORY OF PROBABILITY

SANTOSH S. VENKATESH

*University of Pennsylvania*



CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town,  
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press  
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

[www.cambridge.org](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9781107024472](http://www.cambridge.org/9781107024472)

© Cambridge University Press 2013

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 2013

Printed and bound in the United Kingdom by the MPG Books Group

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

ISBN 978-1-107-02447-2 Hardback

Additional resources for this publication at [www.cambridge.org/venkatesh](http://www.cambridge.org/venkatesh)

## THE THEORY OF PROBABILITY

From classical foundations to advanced modern theory, this self-contained and comprehensive guide to probability weaves together mathematical proofs, historical context, and richly detailed illustrative applications.

A theorem discovery approach is used throughout, setting each proof within its historical setting, and is accompanied by a consistent emphasis on elementary methods of proof. Each topic is presented in a modular framework, combining fundamental concepts with worked examples, problems and digressions which, although mathematically rigorous, require no specialised or advanced mathematical background.

Augmenting this core material are over 80 richly embellished practical applications of probability theory, drawn from a broad spectrum of areas both classical and modern, each tailor-made to illustrate the magnificent scope of the formal results. Over 500 homework problems and more than 250 worked examples are included.

Providing a solid grounding in practical probability, without sacrificing mathematical rigour or historical richness, this insightful book is a fascinating reference, and essential resource, for all engineers, computer scientists and mathematicians.

SANTOSH S. VENKATESH is an Associate Professor of Electrical and Systems Engineering at the University of Pennsylvania, whose research interests include probability, information, communication and learning theory, pattern recognition, computational neuroscience, epidemiology and computer security. He is a member of the David Mahoney Institute for Neurological Sciences, and has been awarded the Lindback Award for Distinguished Teaching.

‘This is a gentle and rich book that is a delight to read. Gentleness comes from the attention to detail; few readers will ever find themselves “stuck” on any steps of the derivations or proofs. Richness comes from the many examples and historical anecdotes that support the central themes. The text will support courses of many styles and it is especially attractive for self-guided study.’

*J. J. Michael Steele, University of Pennsylvania*

‘This book does an excellent job of covering the basic material for a first course in the theory of probability. It is notable for the entertaining coverage of many interesting examples, several of which give a taste of significant fields where the subject is applied.’

*Venkat Anantharam, University of California, Berkeley*

‘This book presents one of the most refreshing treatments of the theory of probability. By providing excellent coverage with both intuition and rigor, together with engaging examples and applications, the book presents a wonderfully readable and thorough introduction to this important subject.’

*Sanjeev Kulkarni, Princeton University*

**for**

ces, bes, bessie,  
mummyyumyum, mozer,  
muti, **cecily stewart venkatesh**, anil,  
boneil, banana peel, studebaker, suffering  
succotash, anna, banana, baby dinosaur, bumpus,  
pompous, boom boom music, twelve ways of biscuit,  
booboo, binjammin, rumpus, gumpus, grumpus,  
goofy, goofus, **anil benjamin venkatesh**, chachu, achu,  
chach, ach, uch, uchie, sweetie, honey, bachu, joy, achoo  
tree, muggle artefacts, anna did it, bump, lump, sump,  
chachubachubach, stormcloud, jammie, sara, baby  
tyrannosaurus, simmy, akka, **saraswathi joy venkatesh**,  
vid, id, did, the kid, vidya, viddy, viddydid, viddydid-  
dydid, little viddles, viddlesticks, fiddlesticks, wun-  
kles, winkie, wunks, winks, pooh, pooh sticks,  
imposerous, imposterous, ginormous, broom  
closet, hobbit, gump, rump, peewee,  
googoogaga, **vidya margaret**  
**venkatesh**

who put up with a frequently distracted and unlovable husband and father.



# Preface

GENTLE READER: Henry Fielding begins his great comic novel *Tom Jones* with these words.

An author ought to consider himself, not as a gentleman who gives a private or eleemosynary treat, but rather as one who keeps a public ordinary, at which all persons are welcome for their money. [...] Men who pay for what they eat, will insist on gratifying their palates, however nice and even whimsical these may prove; and if every thing is not agreeable to their taste, will challenge a right to censure, to abuse, and to d—n their dinner without controul.

To prevent therefore giving offence to their customers by any such disappointment, it hath been usual, with the honest and well-meaning host, to provide a bill of fare, which all persons may peruse at their first entrance into the house; and, having thence acquainted themselves with the entertainment which they may expect, may either stay and regale with what is provided for them, or may depart to some other ordinary better accommodated to their taste.

To take a hint from these honest victuallers, as Fielding did, it strikes me therefore that I should at once and without delay explain my motivations for writing this book and what the reader may reasonably hope to find in it. To the expert reader who finds a discursive prolegomenon irritating, I apologise. There have been so many worthy and beautiful books published on the subject of probability that any new entry must needs perhaps make a case for what is being added to the canon.

THE PAST IS PROLOGUE: The subject of chance is rich in tradition and history. The study of games of chance paved the way for a theory of probability, the nascent science of which begot divers applications, which in turn led to more theory, and yet more applications. This fecund interplay of theory and application is one of the distinguishing features of the subject. It is too much to hope to cover all of the facets of this interaction within the covers of one volume—or indeed many such volumes—and I shall look to the history for guidance.

A central thread running through the theory right from its inceptions in antiquity is the concept peculiar to chance of “statistical independence”. This is the notion that rescues probability from being merely a fragrant by-water of the general theory of measure. To be sure one can articulate the abstract idea of “independent functions” but it appears to have little traction in measure outside of the realm of probability where it not only has a profound impact on the theory but has a peculiarly powerful appeal to intuition.

Historically, the concept of statistical independence was identified first with independent trials in games of chance. Formulations of this principle led to most of the classical results in the theory of probability from the seventeenth century onwards to the early portion of the twentieth century. But the theme is far from exhausted: the last quarter of the twentieth century has seen the serendipitous emergence of new, hugely profitable directions of inquiry on deep and unsuspected aspects of independence at the very heart of probability.

The chronological summary of these new directions that I have included below naturally cannot in its brevity do full justice to all the actors who have helped expand the field. Without pretending to completeness it is intended for the expert reader who may appreciate a quick overview of the general tendency of these results and their connections to the earlier history.

- ✦ Vapnik and Chervonenkis’s beautiful investigation of uniform convergence in 1968 expanded hugely on Glivenko and Cantelli’s classical results dating to 1933. This work spurred the development of empirical process theory and served as an impetus for the burgeoning science of machine learning.
- ✦ Stein unveiled his method of approximate computation of expectations in 1970. The method sketched a subtle and fundamentally different approach to the ubiquitous central limit theorem which dates back to de Moivre in 1733. While it was only slowly that the novelty and genuine power of the idea came to be appreciated, Stein’s method has not only placed the classical canon in an inviting new light, but has opened new doors in the investigation of central tendency.
- ✦ The application of Stein’s method to Poisson approximation was fleshed out by Chen in 1976 and breathed new life into the theory sparked by Poisson’s approximation to the binomial in 1837. The theory that has emerged has provided flexible and powerful new tools in the analysis of rare events, extrema, and exceedances.
- ✦ The Lovász local lemma appeared in 1975 and provided a subtle view of the classical probability sieves used by de Montmort in 1708 and whose provenance goes as far back as the number-theoretic sieves known to the Greeks. It is hard to overstate the abiding impact the local lemma has had;

it and the related sieve arguments that it engendered are now a staple of combinatorial models.

- ✦ The idea that the phenomenon of concentration of measure is very pervasive began to gain traction in the 1980s through the efforts of Gromov and Milman. Talagrand's stunning paper of 1995 placed an exclamation point on the idea. It is most satisfying that this powerful idea constitutes a vast extension of scope of perhaps the most intuitive and oldest idea in probability—the law of large numbers.

These newer developments all in one way or the other expand on the central idea of independence and, to the great pleasure of this author, connect back, as I have indicated, to some of the oldest themes in probability. In close to twenty five years of teaching and stuttering attempts at writing at the University of Pennsylvania and visiting stints at the California Institute of Technology and the Helsinki University of Technology I have attempted to connect these themes from different perspectives, the story being modulated by the developments that were occurring as I was teaching and writing. This book is the result: I could perhaps have titled it, more whimsically, *A Tale of Independence*, and I would not have been far wrong.

**PHILOSOPHY AND THE CUSTOM:** The reader who is new to the subject will find a comprehensive exploration of the theory and its rich applications within these covers. But there is something here for the connoisseur as well; any such will find scattered vignettes through the book that will charm, instruct, and illuminate. (This paragraph was written on a day when the sun was shining, the lark was on the wing, and the author was feeling good about the material.)

One of my goals in teaching the subject, and ultimately in writing down what I taught, was to illustrate the intimate connections between abstract theory and vibrant application—there is perhaps no other mathematical science where art, application, and theory coexist so beautifully. This is a serious book withal, an honest book; the proofs of the theorems meet the exacting standards of a professional mathematician. But my pedagogical inclination, shaped by years of teaching, has always been to attempt to discover theorems, perhaps as they might first have been unearthed, rather than to present them one after the other, like so many slices of dry bread, as if they were a mere litany of facts to be noted. And, at the same time, attempt to place the unfolding development of the formal theory in context by both preparing the ground and promptly illustrating its scope with meaty and colourful applications. The novelty is not in new results—although, to be sure, there are new proofs and problems here and there—but in arrangement, presentation, and perspective.

I have endeavoured to make the proofs self-contained and complete, preferably building on repeated elementary themes rather than on a stable of new tricks or sophisticated theorems from other disciplines. It has never struck me that it is pedagogically useful to attempt to prove a theorem by appeal to

another result that the reader has as little chance of knowing; she is being asked to take something on faith in any case and if that is so one may as well just ask her to believe the theorem in question.

Of course there is always the question of what background may be reasonably assumed. My audiences have ranged from undergraduate upperclassmen to beginning graduate students to advanced graduate students and specialists; and they have come from an eclectic welter of disciplines ranging across engineering, computer science, mathematics, statistics, and pure and applied science. A common element in their backgrounds has been a solid foundation in undergraduate mathematics, say, as taught in a standard three or four-course calculus sequence that is a staple in engineering, science, or mathematics curricula. A reader with this as background and an interest in mathematical probability will, with sufficient good will and patience, be able to make her way through most of this book; more advanced tools and techniques are developed where they are needed and a short Appendix fills in lacunae that may have crept into a calculus sequence.

And then there is the question of measure. Probability is, with the possible exception of geometry, the most intuitive of the mathematical sciences and students in a first course on the subject tend to have a strong intuition for it. But, as a graduate student once told me, a subsequent measure-theoretic course on the subject felt as though it were dealing with another subject altogether; a too early focus on measure-theoretic foundations has the unfortunate effect of viewing the subject at a vast remove from its rich intuitive base and the huge application domain, as though measure is from Mars, probability from Venus. And I have found this sentiment echoed repeatedly among students. Something valuable is lost if the price of rigour is a loss of intuition.

I have attempted to satisfy the demands of intuition and rigour in the narrative by beginning with the elementary theory (though a reader should not confuse the word elementary to mean easy or lacking subtlety) and blending in the theory of measure half way through the book. While measure provides the foundation of the modern theory of probability, much of its import, especially in the basic theory, is to provide a guarantee that limiting arguments work seamlessly. The reader willing to take this on faith can plunge into the rich theory and applications in the later chapters in this book, returning to shore up the measure-theoretic details as time and inclination allow. I have found to my pleasant surprise over the years that novices have boldly plunged into passages where students with a little more experience are sadly hampered by the fear of a misstep and tread with caution. Perhaps it is the case that the passage from the novitiate to the cloister is through confusion.

A serious student of a subject is not an idle spectator to a variety show but learns best by active involvement. This is particularly true of mathematical subjects. I have accordingly included a large collection of problems for solution, scattered quite evenly throughout the text. While there are new problems here

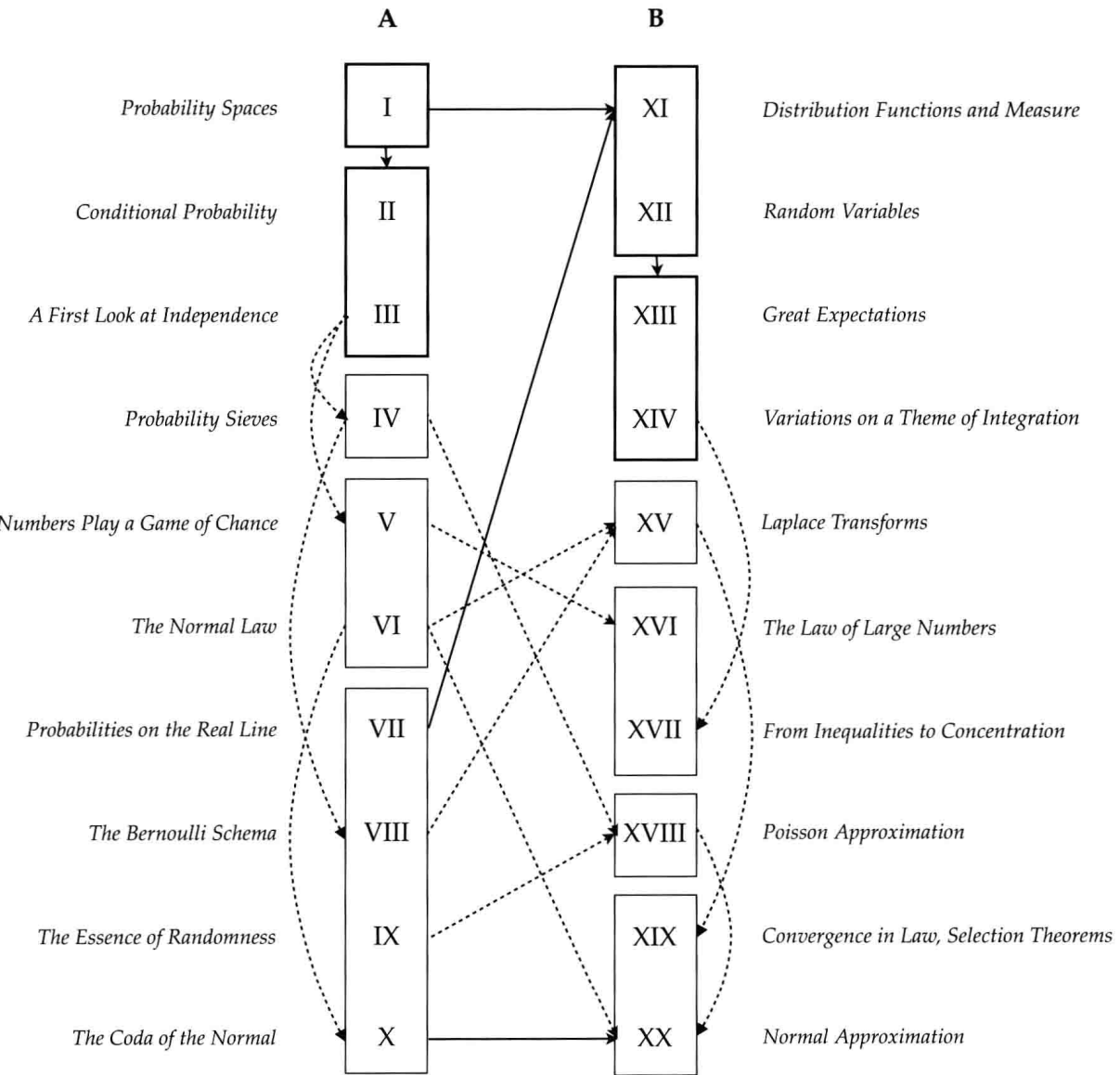
and there to be sure, I have not, by and large, attempted to reinvent the wheel and have taken liberally from the large corpus of problems which are part of the folklore of this venerable subject; providing attribution here is complicated by the confused history and the serial reuse of good problems but where I am aware of a primary source I have provided a name or a reference. Very few of these problems are of a cookbook nature; in some cases they build on the developments in the main body of the text and in others explore significant new areas. Assessing the probable difficulty of a problem for a reader is a tricky task but I have flagged some problems as containing difficult or dangerous digressions so that the reader has some visual guidance.

It has been said with some justification that mathematicians are indifferent historians and, while I cannot claim to have set new standards of accuracy in this regard, I have attempted to provide representative sources; either the original when the antecedents are clear, or a scholarly work which has summarised or clarified the work of many predecessors. Bearing in mind the broad canvas of exploration and the eclectic backgrounds of my students, I have kept the citations generally targeted and specific, not encyclopaedic. While I have faithfully adhered to the original spelling of names in the Latin alphabet, there is no commonly accepted convention for the transliteration of names from other alphabets such as the Cyrillic and variant spellings are to be expected. Bearing in mind Philip J. Davis's admonition in his charming book *The Thread: a Mathematical Yarn* that only admirers of Čaykovskiy's music may write Čebysev in a reference to the great nineteenth-century Russian mathematician, I have kept transliterations phonetic, simple, and common.

Inevitably, the price to be paid for an honest account of the foundational theory and its applications is in coverage. One cannot be all things to all people. I suppose I could plead personal taste in the shape of the narrative but as Kai Lai Chung has remarked in the preface of his classical book on probability, in mathematics, as in music, literature, or cuisine, there is good taste and bad taste; and any author who pleads personal taste must be willing to be judged thereby. And so the discerning reader must decide for herself whether I have been wise in my choices. The reader who wishes to learn more about the theory of Markov chains, renewal theory, information theory, stochastic processes, martingale limit theory, ergodic theory and dynamical systems, or Itô integration must needs look elsewhere. But she will be well prepared with a sound and ample base from which she can sally forth.

The occasional reference to a female reader is idiosyncratic but is not intended to imply that either gender has a monopoly on mathematical thinking; this father was influenced not only by his daughters who over the years peered over his shoulders as he typed and giggled at the titles, but also by the fact that, in life, as in this book, the goddess chance rules.

THE BILL OF FARE; OR, HOW TO READ THIS BOOK: The layout of the text is shown in Figure 1 on the following page. It is arranged in two parts of roughly



**Figure 1:** The layout of the book. Bold arrows indicate precursor themes which should be absorbed first; dashed arrows indicate connections across themes.

equal size divided into ten chapters apiece. The first part contains the elements of the subject but the more experienced reader will already find previews of deep results obtained by elementary methods scattered through the material: sieve methods and the local lemma (IV); connections with number theory and the laws of large numbers (V); the central limit theorem and large deviations (VI); fluctuation theory (VIII); covering problems and queuing (IX); and mixing and Brownian motion (X). The second part (XI–XX) contains the more abstract foundations where these and other themes are echoed and amplified, and their modern incarnations alluded to earlier fleshed out.

The material was originally written as a sequence of “essays” and while the demands of classroom instruction have meant that the original free-flowing narrative now has some order superimposed upon it in the sequencing of the chapters, the arrangement of the material within and across chapters still has a strong flavour of a menu from which items can be sampled. A reader should study at least the core sections of the introductory chapters of each part (shown enclosed in heavier weight boxes in the figure) before embarking on the succeeding material. With a little good will on the part of the reader each connected block of chapters may then be read independently of the others with, perhaps, a glance at notational conventions of antecedent material. Any such can be readily tracked down via the detailed Index.

To provide further guidance to the reader, the margin of the first page of each chapter contains an annotation cataloguing the nature of the “essays” to follow:  $\mathcal{C}$  connotes core sections where key concepts are developed and the basic theorems proved;  $\mathcal{A}$  connotes sections containing applications; and, following Bourbaki and Knuth, the “dangerous bend” sigil  $\curvearrowright$  connotes sections containing material that is more technical, difficult, or digressionary in nature. *A beginning reader should at a minimum read through the core sections; these contain the “essentials”.* Interleaved with these she will find a smorgasbord of applications and digressionary material scattered liberally through the text; *these may generally be sampled in any order;* succeeding theory does not, as a general rule, build upon these. To visually guide the reader, section headings for applications and dangerous digressions appear italicised both in page headers and in the Table of Contents. Digressionary material is flagged additionally by appearing in small print; *this material can be safely skipped on a first reading without loss of continuity.* Where an entire section is devoted to a tangential or technical tributary, I have also flagged the section with the “dangerous bend” sigil  $\curvearrowright$ . Table 1 on the following page shows the breakdown of core, application, and dangerous bend sections.

It is my hope that the more experienced reader will be tempted to flit and explore; for the novice reader there is enough material here for an organised course of study spread over a year if the text is followed linearly. There are also various possibilities for a single-semester course of instruction depending on background and sophistication. I have road tested the following variants.

- ♣ Chapters I–III, VII–X (skipping the bits in small print) can form the core of an

| Chapter Titles |       |                                        | $\mathcal{C}$  | $\mathcal{A}$  |  |
|----------------|-------|----------------------------------------|----------------|----------------|-------------------------------------------------------------------------------------|
| A              | I     | Probability Spaces                     | 1–7            |                | 8, 9                                                                                |
|                | II    | Conditional Probabilities              | 1, 3, 7        | 2, 4–6, 8–10   |                                                                                     |
|                | III   | A First Look at Independence           | 1, 4           | 2, 3           | 5                                                                                   |
|                | IV    | Probability Sieves                     | 1, 4, 6, 9     | 2, 3, 5, 7, 10 | 8, 11                                                                               |
|                | V     | Numbers Play a Game of Chance          | 1–7            |                |                                                                                     |
|                | VI    | The Normal Law                         | 1–5            |                | 6–9                                                                                 |
|                | VII   | Probabilities on the Real Line         | 1–6, 9, 10     | 7, 8           |                                                                                     |
|                | VIII  | The Bernoulli Schema                   | 1, 2, 6, 7, 10 | 3–5, 9         | 8                                                                                   |
|                | IX    | The Essence of Randomness              | 1, 5, 8–10     | 2–4, 6, 7      |                                                                                     |
|                | X     | The Coda of the Normal                 | 1–4            | 5, 6, 8, 9     | 7, 10–12                                                                            |
| B              | XI    | Distribution Functions and Measure     | 1–3            |                | 4, 5                                                                                |
|                | XII   | Random Variables                       | 1–8            |                | 9, 10                                                                               |
|                | XIII  | Great Expectations                     | 1–5, 8         | 6, 7           |                                                                                     |
|                | XIV   | Variations on a Theme of Integration   | 1–5, 10        | 6, 8, 9, 11    | 7                                                                                   |
|                | XV    | Laplace Transforms                     | 1–3            | 4–8            | 9                                                                                   |
|                | XVI   | The Law of Large Numbers               | 1, 2, 9        | 3–7            | 8, 10–12                                                                            |
|                | XVII  | From Inequalities to Concentration     | 1, 3–5         | 2, 7–10        | 6                                                                                   |
|                | XVIII | Poisson Approximation                  | 1–5, 7         | 6, 8, 9        |                                                                                     |
|                | XIX   | Convergence in Law, Selection Theorems | 1–3, 5, 6      |                | 4, 7–9                                                                              |
|                | XX    | Normal Approximation                   | 1, 5, 8        | 6, 7, 9–12     | 2–4                                                                                 |
| C              | XXI   | Sequences, Functions, Spaces           | 1–3            |                |                                                                                     |

**Table 1:** Distribution of sections in the chapter layout.  $\mathcal{C}$ : the core sections of each chapter contain the key concepts, definitions, and the basic theorems; these should be read in sequence.  $\mathcal{A}$ : the application sections are optional and may be sampled in any order. The dangerous bend sections contain subtleties, intriguing, but perilous, examples, technical details not critical to the main flow of the narrative, or simply fun digressions; they should be skipped on a first reading.

honours course for experienced undergraduates, supplemented, at the instructor's discretion, by theory from V and VI or applications from IV, XVI, or XX.

- ✦ Chapters I, XI, XII, V, VI, XIII, XIV, and XVI, complemented by selections from X, XV, XIX, or XX, constitute a more abstract, foundational course aimed at graduate students with some prior exposure to probability.
- ✦ A seminar course for advanced graduate students can be cobbled together, based on interest, from the following thematic groups of chapters: (XV), (V, XVI, XVII), (IV, XVIII), and (VI, X, XIX, XX).

*A word on cross-references and terminology.* On pedagogical grounds it seemed to me to be worthwhile to minimise cross-references at the cost of a little repetition. I have accordingly adopted a parsimonious referencing convention and numbered items only where needed for a later reference. Numbered objects like theorems, lemmas, and examples are numbered sequentially *by section* to keep the numbering spare and unencumbered; where needed to unambiguously identify the object under discussion I have amplified the reference to include details of the section or the chapter in which it may be found. To illustrate, Theorem 2 in Section 4 of Chapter IV (this is the first Borel–Cantelli lemma) is referred to in increasing levels of specificity as Theorem 2, Theorem 4.2, or Theorem IV.4.2 depending on whether the reference to it occurs in the same section, another section of the same chapter, or another chapter. Other numbered objects like lemmas, slogans, definitions, examples, and equations are treated in the same way. While I have included a large collection of figures, tables, and problems, these are rarely cross-referenced and I have numbered them sequentially within each chapter; where needed these are identified by chapter.

It is as well to settle a point of terminology here. In keeping with a somewhat cavalier customary usage I use the terms *positive*, *negative*, *increasing*, and *decreasing* rather *elastically* to mean *non-negative*, *non-positive*, *non-decreasing*, and *non-increasing*, respectively; in these cases I reserve the use of the qualifier “strictly” to eschew the possibility of equality. Thus, I say that the sequence  $\{x_n\}$  is increasing to mean  $x_n \leq x_{n+1}$  for each  $n$  and modify the statement to  $\{x_n\}$  is strictly increasing if I mean that  $x_n < x_{n+1}$ . Likewise, I say  $x$  is positive to mean  $x \geq 0$  and say that  $x$  is strictly positive when I mean  $x > 0$ .

ACKNOWLEDGEMENTS: Joel Franklin taught me probability and showed me its beauty. My thinking on the subject was shaped by his gentle guidance, sagacity, and unquenchable enthusiasm. I owe him a deep debt of gratitude.

My students over the years have proved willing foils as I have tested this material on them and my teaching assistants over this period have been pearls beyond price. Shao Fang, Gaurav Kasbekar, Jonathan Nukpezah, Alireza Tahbaz Salehi, Shahin Shahrampour, Evangelos Vergetis, and Zhengwei Wu worked out problems, proofread various portions of the material, and provided suggestions and critiques. Between them they have helped me hone the material and collar ambiguities and errata. Those that remain must, of course, be laid at my door; I can only hope that they are of the venial kind.

My colleagues at Penn and elsewhere have been uniformly encouraging, commenting fulsomely on earlier versions, and I am very grateful for their support and kindness. As has been mentioned by other authors, a long list of acknowledgements excites more attention by its omissions than by its inclusions, and I will not attempt to list everyone to whom I am indebted here. I will only mention that I owe Herb Wilf, Tom Cover, and Mike Steele particular debts of gratitude: Herb, at the inception, for encouraging me to write the book when its outline was barely discernible in a haphazard collection of notes, Tom, at the halfway point when the book was spiralling, apparently ineluctably, into ever-deepening complexity, for reminding me of the virtues of simplicity, and Mike, at the conclusion, for his kind words of encouragement in the dark hour just before dawn when an author looks at his completed work and despairs of it.

I should not forget to mention the good folk at Cambridge University Press who have been so supportive: Phil Meyler waited patiently for this book for a very long time—I can only hope it has been worth the wait—Elizabeth Horne helped select the cover art and provided careful and detailed suggestions on how to improve the look and feel of the manuscript as a whole, Abigail Jones shepherded the book expertly through the production process, and Sarah Lewis whisked the copy-editor's broom over the document with a light and sure touch.