Contents

Preface

This book began life as a set of handwritten notes, distributed to students in my one-semester graduate course on probability theory, a course that had humble aims: to help the students understand results such as the strong law of large numbers, the central limit theorem, conditioning, and some martingale theory. Along the way they could expect to learn a little measure theory and maybe even a smattering of functional analysis, but not as much as they would learn from a course on Measure Theory or Functional Analysis.

In recent years the audience has consisted mainly of graduate students in statistics and economics, most of whom have not studied measure theory. Most of them have no intention of studying measure theory systematically, or of becoming professional probabilists, but they do want to learn some rigorous probability theory—in one semester.

Faced with the reality of an audience that might have neither the time nor the inclination to devote itself completely to my favorite subject, I sought to compress the essentials into a course as self-contained as I could make it. I tried to pack into the first few weeks of the semester a crash course in measure theory, with supplementary exercises and a whirlwind exposition (Appendix A) for the enthusiasts. I tried to eliminate duplication of mathematical effort if it served no useful role. After many years of chopping and compressing, the material that I most wanted to cover all fitted into a one semester course, divided into 25 lectures each lasting from 60 to 75 minutes. My handwritten notes filled fewer than a hundred pages.

I had every intention of making my little stack of notes into a little book. But I couldn't resist expanding a bit here and a bit there, adding in useful reference material, spelling out ideas that I had struggled with on first acquaintance, slipping in extra topics that my students have seemed to need when writing dissertations, pulling in material from other courses that I have taught and neat tricks that I have learned from my friends. And soon it wasn't so little any more.

Many of the additions ended up in starred Sections, which contain harder material or topics that can be skipped over without loss of continuity.

My treatment includes a few eccentricities that might upset some of my professional colleagues. My most obvious departure from tradition is in the use of linear functional notation for expectations, an approach that I first encountered in books by de Finetti. I attempt to explain the virtues of this notation in the first two Chapters. Another slight novelty—at least for anyone already exposed to the Kolmogorov interpretation of conditional expectations—appears in my treatment of conditioning, in Chapter 5. For many years I have worried about the wide gap between the free-wheeling conditioning calculations of an elementary probability course and the formal manipulations demanded by rigor. I claim that a treatment starting from the idea of conditional distributions offers one way of bridging the gap, at least for many of the statistical applications of conditioning that have troubled me the most.

The twelve Chapters and six Appendixes contain general explanations, remarks, opinions, and blocks of more formal material. Theorems and Lemmas contain the most important mathematical details. Examples contain gentler, or less formal, explanations and illustrations. Supporting theoretical material is presented either in the form of Exercises, with terse solutions, or as Problems (at the ends of the Chapters) that work step-by-step through material that missed the cutoff as Exercises, Lemmas, or Theorems. Some Problems are routine, to give students an opportunity to digest the ideas in the text without great mental effort; some Problems are hard.

A possible one-semester course

Here is a list of the material that I usually try to cover in the one-semester graduate course.

Chapter 1: Spend one lecture on why measure theory is worth the effort, using a few of the Examples as illustrations. Introduce de Finetti notation, identifying sets with their indicator functions, and writing $\mathbb{P}$ for both probabilities of sets and expectations of random variables. Mention, very briefly, the fair price Section as an alternative to the frequency interpretation.

Chapter 2: Cover the unstarred Sections carefully, but omitting many details from the Examples. Postpone Section 7 until Chapter 3. Postpone Section 8 until Chapter 6. Describe briefly the generating class theorem for functions, from Section 11, without proofs.

Chapter 3: Cover Section 1, explaining the connection with the elementary notion of a density. Take a short excursion into Hilbert space (explaining the projection theorem as an extension of the result for Euclidean spaces) before presenting the simple version of Radon-Nikodym. Mention briefly the classical concept of absolute continuity, but give no details. Maybe say something about total variation.

Chapter 4: Cover Sections 1 and 2, leaving details of some arguments to the students. Give a reminder about generating classes of functions. Describe construction of $\mu \otimes \Lambda$, only for a finite kernel Λ, via the iterated integral. Cover product measures, using some of the Examples from Section 4. Explain the need for the blocking idea from Section 6, using the Maximal Inequality to preview the idea of a stopping time. Mention the truncation idea behind the version of the SLLN for independent, identically distributed random variables with finite first moments, but skip most of the proof.

Chapter 5: Discuss Section 1 carefully. Cover the high points of Sections 2 through 4. (They could be skipped without too much loss of continuity, but I prefer not to move straight into Kolmogorov conditioning). Cover Section 6.

Chapter 6: Cover Sections 1 through 4, but skipping over some Examples. Characterize uniformly integrable martingales, using Section 6 and some of material postponed from Section 8 of Chapter 2, unless short of time.

Chapter 7: Cover the first four Sections, skipping some of the examples of central limit theorems near the end of Section 2. Downplay multivariate results.

Chapter 8: Cover Sections 1, 2, 4, and 6.

If time is left over, cover a topic from the remaining Chapters.

Acknowledgments

I am particularly grateful to Richard Gill, who is a model of the constructive critic. His comments repeatedly exposed weaknesses and errors in the manuscript. My colleagues Joe Chang and Marten Wegkamp asked helpful questions while using earlier drafts to teach graduate probability courses. Andries Lenstra provided some important historical references.

Many cohorts of students worked through the notes, revealing points of obscurity and confusion. In particular, Jeankyung Kim, Gheorghe Doros, Daniela Cojocaru, and Peter Radchenko read carefully through several chapters and worked through numerous Problems. Their comments led to a lot of rewriting.

Finally, I thank Lauren Cowles for years of good advice, and for her inexhaustible patience with an author who could never stop tinkering.

David Pollard
New Haven

February 2001