Preface

To some extent, it would be accurate to summarize the contents of this book as an intolerably protracted description of what happens when either one raises a transition probability matrix $\mathbf{P}$ (i.e., all entries $(\mathbf{P})_{ij}$ are non-negative and each row of $\mathbf{P}$ sums to 1) to higher and higher powers or one exponentiates $\mathbf{R}(\mathbf{P}-\mathbf{I})$, where $\mathbf{R}$ is a diagonal matrix with non-negative entries. Indeed, when it comes right down to it, that is all that is done in this book. However, I, and others of my ilk, would take offense at such a dismissive characterization of the theory of Markov chains and processes with values in a countable state space, and a primary goal of mine in writing this book was to convince its readers that our offense would be warranted.

The reason why I, and others of my persuasion, refuse to consider the theory here as no more than a subset of matrix theory is that to do so is to ignore the pervasive role that probability plays throughout. Namely, probability theory provides a model which both motivates and provides a context for what we are doing with these matrices. To wit, even the term "transition probability matrix" lends meaning to an otherwise rather peculiar set of hypotheses to make about a matrix. Specifically, it suggests that we think of the matrix entry $(\mathbf{P})_{ij}$ as giving the probability that, in one step, a system in state i will make a transition to state j. Moreover, if we adopt this interpretation for $(\mathbf{P})_{ij}$, then we must interpret the entry $(\mathbf{P}^n)_{ij}$ of $\mathbf{P}^n$ as the probability of the same transition in n steps. Thus, as $n \to \infty$, $\mathbf{P}^n$ is encoding the long time behavior of a randomly evolving system for which $\mathbf{P}$ encodes the one-step behavior, and, as we will see, this interpretation will guide us to an understanding of $\lim_{n\to\infty}(\mathbf{P}^n)_{ij}$. In addition, and perhaps even more important, is the role that probability plays in bridging the chasm between mathematics and the rest of the world. Indeed, it is the probabilistic metaphor which allows one to formulate mathematical models of various phenomena observed in both the natural and social sciences. Without the language of probability, it is hard to imagine how one would go about connecting such phenomena to $\mathbf{P}^n$.

In spite of the propaganda at the end of the preceding paragraph, this book is written from a mathematician's perspective. Thus, for the most part, the probabilistic metaphor will be used to elucidate mathematical concepts rather than to provide mathematical explanations for non-mathematical phenomena. There are two reasons for my having chosen this perspective. First, and foremost, is my own background. Although I have occasionally tried to help people who are engaged in various sorts of applications, I have not accumulated a large store of examples which are easily translated into terms which are appropriate for a book at this level. In fact, my experience has taught me that people engaged in applications are more than competent to handle the routine problems that they encounter, and that they come to someone like me only as a last resort. As a consequence, the questions which they ask me tend to be quite difficult and the answers to those few which I can solve usually involve material which is well beyond the scope of the present book. The second reason for my writing this book in the way that I have is that I think the material itself is of sufficient interest to stand on its own. In spite of what funding agencies would have us believe, mathematics *qua* mathematics is a worthy intellectual endeavor, and I think there is a place for a modern introduction to stochastic processes which is unabashed about making mathematics its top priority.

I came to this opinion after several semesters during which I taught the introduction to stochastic processes course offered by the M.I.T. department of mathematics. The clientele for that course has been an interesting mix of undergraduate and graduate students, less than half of whom concentrate in mathematics. Nonetheless, most of the students who stay with the course have considerable talent and appreciation for mathematics, even though they lack the formal mathematical training which is requisite for a modern course in stochastic processes, at least as such courses are now taught in mathematics departments to their own graduate students. As a result, I found no ready-made choice of text for the course. On the one hand, the most obvious choice is the classic text *A First Course in Stochastic Processes*, either the original one by S. Karlin or the updated version [4] by S. Karlin and H. Taylor. Their book gives a no nonsense introduction to stochastic processes, especially Markov processes, on a countable state space, and its consistently honest, if not always easily assimilated, presentation of proofs is complemented by a daunting number of examples and exercises. On the other hand, when I began, I feared that adopting Karlin and Taylor for my course would be a mistake of the same sort as adopting Feller's book for an undergraduate introduction to probability, and this fear prevailed the first two times I taught the course. However, after using, and finding wanting, two derivatives of Karlin's classic, I took the plunge and assigned Karlin and Taylor's book. The result was very much the one which I predicted: I was far more enthusiastic about the text than were my students.

In an attempt to make Karlin and Taylor's book more palatable for the students, I started supplementing their text with notes in which I tried to couch the proofs in terms which I hoped they would find more accessible, and my efforts were rewarded with a quite positive response. In fact, as my notes became more and more extensive and began to diminish the importance of the book, I decided to convert them into what is now this book, although I realize that my decision to do so may have been stupid. For one thing, the market is already close to glutted with books that purport to cover this material. Moreover, some of these books are quite popular, although my experience with them leads me to believe that their popularity is not always correlated with the quality of the mathematics they contain. Having made that pejorative comment, I will not make public which are the books that led me to this conclusion. Instead, I will only mention the books on this topic, besides Karlin and Taylor's, which I very much like. J. Norris's book [5] is an excellent introduction to Markov processes which, at the same time, provides its readers with a good place to exercise their measure-theoretic skills. Of course, Norris's book is only appropriate for students who have measure-theoretic skills to exercise. On the other hand, for students who possess those skills, his book is a place where they can see measure theory put to work in an attractive way. In addition, Norris has included many interesting examples and exercises which illustrate how the subject can be applied. For more advanced students, an excellent treatment of Markov chains on a general state space can be found in the book [6] by D. Revuz.

The present book includes most of the mathematical material contained in [5], but the proofs here demand much less measure theory than his do. In fact, although I have systematically employed measure theoretic terminology (Lebesgue's dominated convergence theorem, the monotone convergence theorem, etc.), which is explained in Chap. 7, I have done so only to familiarize my readers with the jargon that they will encounter if they delve more deeply into the subject. In fact, because the state spaces in this book are countable, the applications which I have made of Lebesgue's theory are, with one notable exception, entirely trivial. The one exception is that I need to know the existence of countably infinite families of mutually independent random variables. In Sect. 7.2 I discuss how one goes about proving their existence, but, as distinguished from the first edition of this text, I do not go into details and instead refer to the treatment in [8]. Be that as it may, the reader who is ready to accept that such families exist has little need to consult Chap. 7 except for terminology and the derivation of a few essentially obvious facts about series.

The organization of this book should be more or less self-evident from the table of contents. In Chap. 1, I give a bare hands treatment of the basic facts, with particular emphasis on recurrence and transience, about nearest neighbor random walks on the square, d-dimensional lattice $\mathbb{Z}^d$. Chapter 2 introduces the study of ergodic properties, and this becomes the central theme which ties together Chaps. 2 through 6. In Chap. 2, the stochastic processes under consideration are Markov chains (i.e., the time parameter is discrete), and the driving force behind the development there is an idea which was introduced by Doeblin. Restricted as the applicability of Doeblin's idea may be, it has the enormous advantage over the material in Chaps. 4 and 5 that it provides an estimate on the rate at which the chain is converging to its equilibrium distribution. Chapter 3 begins with the classification of states in terms of recurrence and transience and then introduces some computational techniques for computing stationary probabilities. As an application, the final section of Chap. 3 gives a proof that Wilson's algorithm works and that it can be used to derive Kirchhoff's matrix tree theorem. The contents of this section are based on ideas that I learned from S. Sternberg, who learned them from M. Kozdron, who in turn learned them from G. Lawler. It was Kozdron who had the idea of using Wilson's algorithm to derive Kirchhoff's theorem.

In Chap. 4, I study the ergodic properties of Markov chains that do not necessarily satisfy Doeblin's condition. The main result here is the one summarized in equation

(4.1.15). Even though it is completely elementary, the derivation of (4.1.15), is, without doubt, the most demanding piece of analysis in the entire book. So far as I know, every proof of (4.1.15) requires work at some stage. In supposedly "simpler" proofs, the work is hidden elsewhere (either measure theory, as in [5] and [6], or in operator theory, as in [2]). The treatment given here, which is a re-working of the one in [4] based on Feller's renewal theorem, demands nothing more of the reader than a thorough understanding of arguments involving limits superior, limits inferior, and their role in proving that limits exist. In Chap. 5, Markov chains are replaced by continuous-time Markov processes (still on a countable state space). I do this first in the case when the rates are bounded and therefore problems of possible explosion do not arise. Afterwards, I allow for unbounded rates and develop criteria, besides boundedness, which guarantee non-explosion. The remainder of the chapter is devoted to transferring the results obtained for Markov chains in Chaps. 2 and 4 to the continuous-time setting.

Aside from Chap. 7, which is more like an appendix than an integral part of the book, the book ends with Chap. 6. The goal in Chap. 6 is to obtain quantitative results, reminiscent of, if not as strong as, those in Chap. 2, when Doeblin's theory either fails entirely or yields rather poor estimates. The new ingredient in Chap. 6 is the assumption that the chain or process is reversible (i.e., the transition probability is self-adjoint in the L^2-space of its stationary distribution), and the engine which makes everything go is the associated Dirichlet form. In the final section, the power of the Dirichlet form methodology is tested in an analysis of the Metropolis (a.k.a. as simulated annealing) algorithm. Finally, as I said before, Chap. 7 is an appendix in which the ideas and terminology of Lebesgue's theory of measure and integration are reviewed. Sect. 7.2.1.

I have finally reached the traditional place reserved for thanking those individuals who, either directly or indirectly, contributed to this book. The principal direct contributors are the many students who suffered with various and spontaneously changing versions of this book. I am particularly grateful to Adela Popescu whose careful reading of the first edition brought to light many minor and a few major errors that have been removed and, undoubtedly, replaced by new ones. In addition, I am grateful to Sternberg and Kozdron for introducing me to the ideas in Sect. 3.3.

Thanking, or even identifying, the indirect contributors is trickier. Indeed, they include all the individuals, both dead and alive, from whom I received my education, and I am not about to bore you with even a partial list of who they were or are. Nonetheless, there is one person who, over a period of more than ten years, patiently taught me to appreciate the sort of material treated here. Namely, Richard A. Holley, to whom I have dedicated this book, is a *true probabilist*. To wit, for Dick, intuitive understanding usually precedes his mathematically rigorous comprehension of a probabilistic phenomenon. This statement should lead no one to to doubt Dick's powers as a rigorous mathematician. On the contrary, his intuitive grasp of probability theory not only enhances his own formidable mathematical powers, it has saved me and others from blindly pursuing flawed lines of reasoning. As all who have worked with him know, reconsider what you are saying if ever, during

some diatribe into which you have launched, Dick quietly says “I don’t follow that.” In addition to his mathematical prowess, every one of Dick’s many students will attest to his wonderful generosity. I was not his student, but I was his colleague, and I can assure you that his generosity is not limited to his students.

Cambridge, MA, USA Daniel W. Stroock
August 2013

Springer

http://www.springer.com/978-3-642-40522-8

An Introduction to Markov Processes
Stroock, D.W.
2014, XVII, 203 p., Hardcover
ISBN: 978-3-642-40522-8