

Preface

Markov first studied the stochastic processes that came to be named after him in 1906. Approximately a century later, there is an active and diverse interdisciplinary community of researchers using Markov chains in computer science, physics, statistics, bioinformatics, engineering, and many other areas.

The classical theory of Markov chains studied *fixed* chains, and the goal was to estimate the rate of convergence to stationarity of the distribution at time t , as $t \rightarrow \infty$. In the past two decades, as interest in chains with large state spaces has increased, a different asymptotic analysis has emerged. Some target distance to the stationary distribution is prescribed; the number of steps required to reach this target is called the *mixing time* of the chain. Now, the goal is to understand how the mixing time grows as the size of the state space increases.

The modern theory of Markov chain mixing is the result of the convergence, in the 1980's and 1990's, of several threads. (We mention only a few names here; see the chapter Notes for references.)

For statistical physicists Markov chains become useful in Monte Carlo simulation, especially for models on finite grids. The mixing time can determine the running time for simulation. However, Markov chains are used not only for simulation and sampling purposes, but also as models of dynamical processes. Deep connections were found between rapid mixing and spatial properties of spin systems, e.g., by Dobrushin, Shlosman, Stroock, Zegarliniski, Martinelli, and Olivieri.

In theoretical computer science, Markov chains play a key role in sampling and approximate counting algorithms. Often the goal was to prove that the mixing time is polynomial in the logarithm of the state space size. (In this book, we are generally interested in more precise asymptotics.)

At the same time, mathematicians including Aldous and Diaconis were intensively studying card shuffling and other random walks on groups. Both spectral methods and probabilistic techniques, such as coupling, played important roles. Alon and Milman, Jerrum and Sinclair, and Lawler and Sokal elucidated the connection between eigenvalues and expansion properties. Ingenious constructions of “expander” graphs (on which random walks mix especially fast) were found using probability, representation theory, and number theory.

In the 1990's there was substantial interaction between these communities, as computer scientists studied spin systems and as ideas from physics were used for sampling combinatorial structures. Using the geometry of the underlying graph to find (or exclude) bottlenecks played a key role in many results.

There are many methods for determining the asymptotics of convergence to stationarity as a function of the state space size and geometry. We hope to present these exciting developments in an accessible way.

We will only give a taste of the applications to computer science and statistical physics; our focus will be on the common underlying mathematics. The prerequisites are all at the undergraduate level. We will draw primarily on probability and linear algebra, but we will also use the theory of groups and tools from analysis when appropriate.

Why should mathematicians study Markov chain convergence? First of all, it is a lively and central part of modern probability theory. But there are ties to several other mathematical areas as well. The behavior of the random walk on a graph reveals features of the graph's geometry. Many phenomena that can be observed in the setting of finite graphs also occur in differential geometry. Indeed, the two fields enjoy active cross-fertilization, with ideas in each playing useful roles in the other. Reversible finite Markov chains can be viewed as resistor networks; the resulting discrete potential theory has strong connections with classical potential theory. It is amusing to interpret random walks on the symmetric group as card shuffles—and real shuffles have inspired some extremely serious mathematics—but these chains are closely tied to core areas in algebraic combinatorics and representation theory.

In the spring of 2005, mixing times of finite Markov chains were a major theme of the multidisciplinary research program *Probability, Algorithms, and Statistical Physics*, held at the Mathematical Sciences Research Institute. We began work on this book there.

Overview

We have divided the book into two parts.

In **Part I**, the focus is on techniques, and the examples are illustrative and accessible. Chapter 1 defines Markov chains and develops the conditions necessary for the existence of a unique stationary distribution. Chapters 2 and 3 both cover examples. In Chapter 2, they are either classical or useful—and generally both; we include accounts of several chains, such as the gambler's ruin and the coupon collector, that come up throughout probability. In Chapter 3, we discuss Glauber dynamics and the Metropolis algorithm in the context of “spin systems.” These chains are important in statistical mechanics and theoretical computer science.

Chapter 4 proves that, under mild conditions, Markov chains do, in fact, converge to their stationary distributions and defines *total variation distance* and *mixing time*, the key tools for quantifying that convergence. The techniques of Chapters 5, 6, and 7, on coupling, strong stationary times, and methods for lower bounding distance from stationarity, respectively, are central to the area.

In Chapter 8, we pause to examine card shuffling chains. Random walks on the symmetric group are an important mathematical area in their own right, but we hope that readers will appreciate a rich class of examples appearing at this stage in the exposition.

Chapter 9 describes the relationship between random walks on graphs and electrical networks, while Chapters 10 and 11 discuss hitting times and cover times.

Chapter 12 introduces eigenvalue techniques and discusses the role of the relaxation time (the reciprocal of the spectral gap) in the mixing of the chain.

In **Part II**, we cover more sophisticated techniques and present several detailed case studies of particular families of chains. Much of this material appears here for the first time in textbook form.

Chapter 13 covers advanced spectral techniques, including comparison of Dirichlet forms and Wilson’s method for lower bounding mixing.

Chapters 14 and 15 cover some of the most important families of “large” chains studied in computer science and statistical mechanics and some of the most important methods used in their analysis. Chapter 14 introduces the path coupling method, which is useful in both sampling and approximate counting. Chapter 15 looks at the Ising model on several different graphs, both above and below the critical temperature.

Chapter 16 revisits shuffling, looking at two examples—one with an application to genomics—whose analysis requires the spectral techniques of Chapter 13.

Chapter 17 begins with a brief introduction to martingales and then presents some applications of the evolving sets process.

Chapter 18 considers the cutoff phenomenon. For many families of chains where we can prove sharp upper and lower bounds on mixing time, the distance from stationarity drops from near 1 to near 0 over an interval asymptotically smaller than the mixing time. Understanding why cutoff is so common for families of interest is a central question.

Chapter 19, on lamplighter chains, brings together methods presented throughout the book. There are many bounds relating parameters of lamplighter chains to parameters of the original chain: for example, the mixing time of a lamplighter chain is of the same order as the cover time of the base chain.

Chapters 20 and 21 introduce two well-studied variants on finite discrete time Markov chains: continuous time chains and chains with countable state spaces. In both cases we draw connections with aspects of the mixing behavior of finite discrete-time Markov chains.

Chapter 22, written by Propp and Wilson, describes the remarkable construction of coupling from the past, which can provide exact samples from the stationary distribution.

Chapter 23 closes the book with a list of open problems connected to material covered in the book.

For the Reader

Starred sections contain material that either digresses from the main subject matter of the book or is more sophisticated than what precedes them and may be omitted.

Exercises are found at the ends of chapters. Some (especially those whose results are applied in the text) have solutions at the back of the book. We of course encourage you to try them yourself first!

The Notes at the ends of chapters include references to original papers, suggestions for further reading, and occasionally “complements.” These generally contain related material not required elsewhere in the book—sharper versions of lemmas or results that require somewhat greater prerequisites.

The Notation Index at the end of the book lists many recurring symbols.

Much of the book is organized by method, rather than by example. The reader may notice that, in the course of illustrating techniques, we return again and again to certain families of chains—random walks on tori and hypercubes, simple card shuffles, proper colorings of graphs. In our defense we offer an anecdote.

In 1991 one of us (Y. Peres) arrived as a postdoc at Yale and visited Shizuo Kakutani, whose rather large office was full of books and papers, with bookcases and boxes from floor to ceiling. A narrow path led from the door to Kakutani's desk, which was also overflowing with papers. Kakutani admitted that he sometimes had difficulty locating particular papers, but he proudly explained that he had found a way to solve the problem. He would make four or five copies of any really interesting paper and put them in different corners of the office. When searching, he would be sure to find at least one of the copies. . . .

Cross-references in the text and the Index should help you track earlier occurrences of an example. You may also find the chapter dependency diagrams below useful.

We have included brief accounts of some background material in Appendix A. These are intended primarily to set terminology and notation, and we hope you will consult suitable textbooks for unfamiliar material.

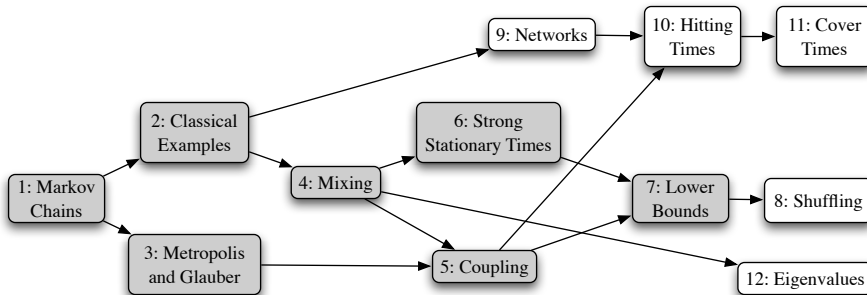
Be aware that we occasionally write symbols representing a real number when an integer is required (see, e.g., the $\binom{n}{\delta k}$'s in the proof of Proposition 13.31). We hope the reader will realize that this omission of floor or ceiling brackets (and the details of analyzing the resulting perturbations) is in her or his best interest as much as it is in ours.

For the Instructor

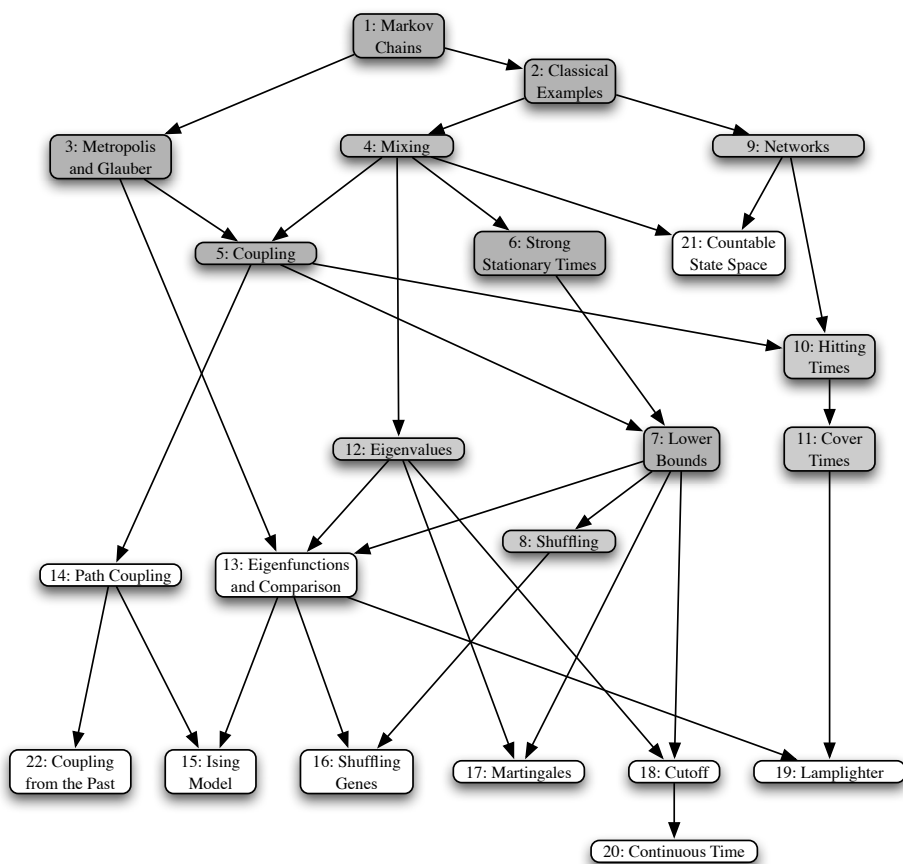
The prerequisites this book demands are a first course in probability, linear algebra, and, inevitably, a certain degree of mathematical maturity. When introducing material which is standard in other undergraduate courses—e.g., groups—we provide definitions, but often hope the reader has some prior experience with the concepts.

In Part I, we have worked hard to keep the material accessible and engaging for students. (Starred sections are more sophisticated and are not required for what follows immediately; they can be omitted.)

Here are the dependencies among the chapters of Part I:



Chapters 1 through 7, shown in gray, form the core material, but there are several ways to proceed afterwards. Chapter 8 on shuffling gives an early rich application but is not required for the rest of Part I. A course with a probabilistic focus might cover Chapters 9, 10, and 11. To emphasize spectral methods and combinatorics, cover Chapters 8 and 12 and perhaps continue on to Chapters 13 and 17.



The logical dependencies of chapters. The core Chapters 1 through 7 are in dark gray, the rest of Part I is in light gray, and Part II is in white.

While our primary focus is on chains with finite state spaces run in discrete time, continuous-time and countable-state-space chains are both discussed—in Chapters 20 and 21, respectively.

We have also included Appendix B, an introduction to simulation methods, to help motivate the study of Markov chains for students with more applied interests. A course leaning towards theoretical computer science and/or statistical mechanics might start with Appendix B, cover the core material, and then move on to Chapters 14, 15, and 22.

Of course, depending on the interests of the instructor and the ambitions and abilities of the students, any of the material can be taught! Above we include a full diagram of dependencies of chapters. Its tangled nature results from the interconnectedness of the area: a given technique can be applied in many situations, while a particular problem may require several techniques for full analysis.

For the Expert

Several other recent books treat Markov chain mixing. Our account is more comprehensive than those of Häggström (2002), Jerrum (2003), or Montenegro and Tetali (2006), yet not as exhaustive as Aldous and Fill (1999). Norris (1998) gives an introduction to Markov chains and their applications, but does not focus on mixing. Since this is a textbook, we have aimed for accessibility and comprehensibility, particularly in Part I.

What is different or novel in our approach to this material?

- Our approach is probabilistic whenever possible. We introduce the random mapping representation of chains early and use it in formalizing randomized stopping times and in discussing grand coupling and evolving sets. We also integrate “classical” material on networks, hitting times, and cover times and demonstrate its usefulness for bounding mixing times.
- We provide an introduction to several major statistical mechanics models, most notably the Ising model, and collect results on them in one place.
- We give expository accounts of several modern techniques and examples, including evolving sets, the cutoff phenomenon, lamplighter chains, and the L -reversal chain.
- We systematically treat lower bounding techniques, including several applications of Wilson’s method.
- We use the transportation metric to unify our account of path coupling and draw connections with earlier history.
- We present an exposition of coupling from the past by Propp and Wilson, the originators of the method.