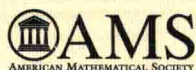


美国数学会经典影印系列



# Markov Chains and Mixing Times

Markov 链与混合时间

David A. Levin

Yuval Peres

Elizabeth L. Wilmer



高等教育出版社

美国数学会经典影印系列



# Markov Chains and Mixing Times

Markov 链与混合时间

David A. Levin

Yuval Peres

Elizabeth L. Wilmer

*With a chapter on "Coupling from the Past" by  
James G. Propp and David B. Wilson*

高等教育出版社·北京

图字: 01-2016-2517 号

*Markov Chains and Mixing Times*, by David A. Levin, Yuval Peres and Elizabeth L. Wilmer,  
first published by the American Mathematical Society.

Copyright © 2009 by the authors. All rights reserved.

This present reprint edition is published by Higher Education Press Limited Company under authority  
of the American Mathematical Society and is published under license.

Special Edition for People's Republic of China Distribution Only. This edition has been authorized by  
the American Mathematical Society for sale in People's Republic of China only, and is not for export therefrom.

本书原版最初由美国数学会于 2009 年出版, 原书名为 *Markov Chains and Mixing Times*,  
作者为 David A. Levin, Yuval Peres and Elizabeth L. Wilmer。原书作者保留原书所有版权。

原书版权声明: Copyright © 2009 by the authors。

本影印版由高等教育出版社有限公司经美国数学会独家授权出版。

本版只限于中华人民共和国境内发行。本版经由美国数学会授权仅在中华人民共和国境内销售, 不得出口。

## Markov 链与混合时间

Markov Lian yu Hunhe Shijian

## 图书在版编目 (CIP) 数据

Markov 链与混合时间 = Markov Chains and Mixing Times :  
英文 / (美) 大卫·A·莱文 (David A. Levin), (以) 尤瓦尔·  
佩雷斯 (Yuval Peres), (美) 伊丽莎白·L·威尔默 (Elizabeth  
L. Wilmer) 著. — 影印本. — 北京: 高等教育出版社, 2017.4  
ISBN 978-7-04-046994-3

I. ①M… II. ①大… ②尤… ③伊… III. ①马尔柯夫链  
—英文IV. ①O211.62

中国版本图书馆 CIP 数据核字 (2016) 第 326796 号

策划编辑 李 鹏

责任编辑 李 鹏

封面设计 张申申

责任印制 赵义民

出版发行 高等教育出版社

社址 北京市西城区德外大街 4 号

邮政编码 100120

购书热线 010-58581118

咨询电话 400-810-0598

网址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

网上订购 <http://www.hepmall.com.cn>

<http://www.hepmall.com>

<http://www.hepmall.cn>

印刷 北京中科印刷有限公司

开本 787mm×1092mm 1/16

印张 24.75

字数 620 千字

版次 2017 年 4 月第 1 版

印次 2017 年 4 月第 1 次印刷

定价 169.00 元

本书如有缺页、倒页、脱页等质量问题,

请到所购图书销售部门联系调换

版权所有 侵权必究

[物料号 46994-00]



## 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, Zegarlinski, 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.

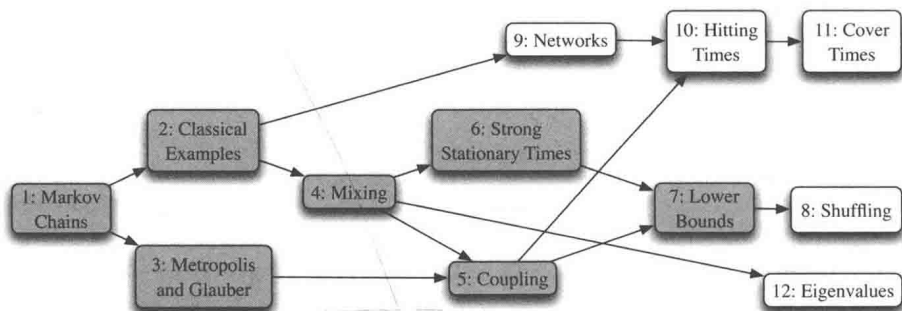
Be aware that we occasionally write symbols representing a real number when an integer is required (see, e.g., the  $\binom{n}{\delta k}$ 's in the proof of Proposition 13.31). We hope the reader will realize that this omission of floor or ceiling brackets (and the details of analyzing the resulting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

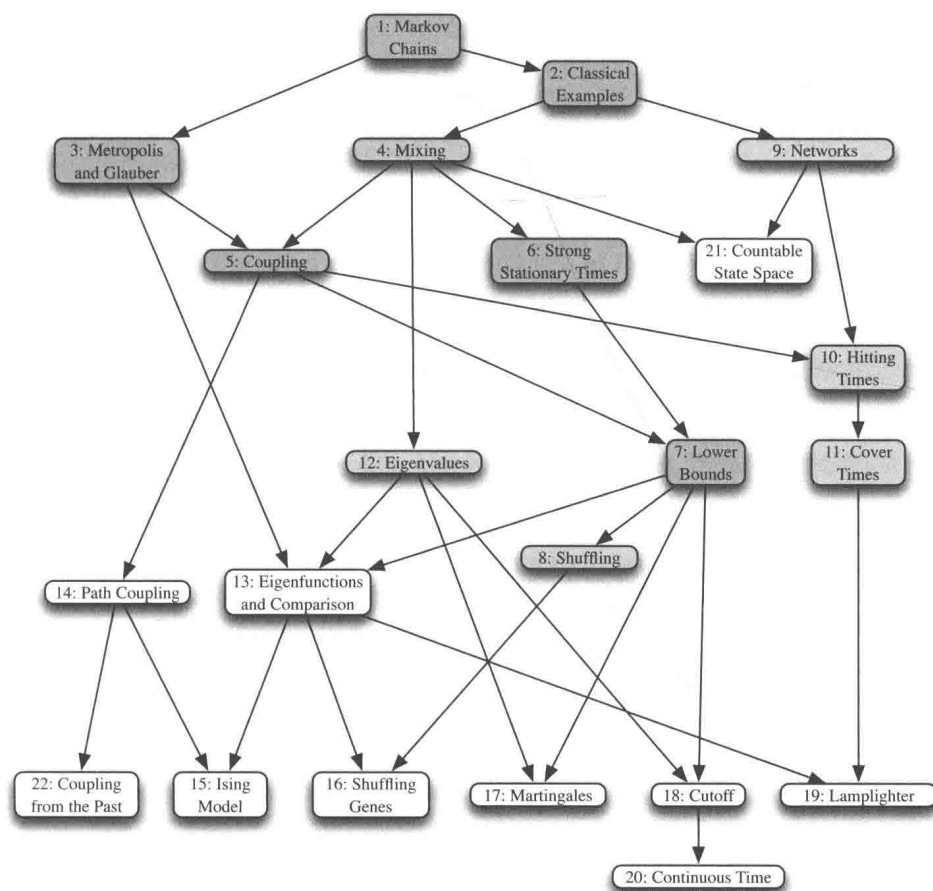
In Part I, we have worked hard to keep the material accessible and engaging for students. (Starred sections are more sophisticated and are not required for what follows immediately; they can be omitted.)

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.

## Acknowledgements

The authors thank the Mathematical Sciences Research Institute, the National Science Foundation VIGRE grant to the Department of Statistics at the University of California, Berkeley, and National Science Foundation grants DMS-0244479 and DMS-0104073 for support. We also thank Hugo Rossi for suggesting we embark on this project. Thanks to Blair Ahlquist, Tonci Antunovic, Elisa Celis, Paul Cuff, Jian Ding, Ori Gurel-Gurevich, Tom Hayes, Itamar Landau, Yun Long, Karola Mészáros, Shobhana Murali, Weiyang Ning, Tomoyuki Shirai, Walter Sun, Sithparran Vanniasagaram, and Ariel Yadin for corrections to an earlier version and making valuable suggestions. Yelena Shvets made the illustration in Section 6.5.4. The simulations of the Ising model in Chapter 15 are due to Raissa D'Souza. We thank László Lovász for useful discussions. We are indebted to Alistair Sinclair for his work co-organizing the M.S.R.I. program *Probability, Algorithms, and Statistical Physics* in 2005, where work on this book began. We thank Robert Calhoun for technical assistance.

Finally, we are greatly indebted to David Aldous and Persi Diaconis, who initiated the modern point of view on finite Markov chains and taught us much of what we know about the subject.

# Contents

Preface	xi
Overview	xii
For the Reader	xiii
For the Instructor	xiv
For the Expert	xvi
Acknowledgements	xvii
Part I: Basic Methods and Examples	1
Chapter 1. Introduction to Finite Markov Chains	3
1.1. Finite Markov Chains	3
1.2. Random Mapping Representation	6
1.3. Irreducibility and Aperiodicity	8
1.4. Random Walks on Graphs	9
1.5. Stationary Distributions	10
1.6. Reversibility and Time Reversals	14
1.7. Classifying the States of a Markov Chain*	16
Exercises	18
Notes	20
Chapter 2. Classical (and Useful) Markov Chains	21
2.1. Gambler's Ruin	21
2.2. Coupon Collecting	22
2.3. The Hypercube and the Ehrenfest Urn Model	23
2.4. The Pólya Urn Model	25
2.5. Birth-and-Death Chains	26
2.6. Random Walks on Groups	27
2.7. Random Walks on $\mathbb{Z}$ and Reflection Principles	30
Exercises	34
Notes	35
Chapter 3. Markov Chain Monte Carlo: Metropolis and Glauber Chains	37
3.1. Introduction	37
3.2. Metropolis Chains	37
3.3. Glauber Dynamics	40
Exercises	44
Notes	44
Chapter 4. Introduction to Markov Chain Mixing	47
4.1. Total Variation Distance	47

4.2. Coupling and Total Variation Distance	49
4.3. The Convergence Theorem	52
4.4. Standardizing Distance from Stationarity	53
4.5. Mixing Time	55
4.6. Mixing and Time Reversal	55
4.7. Ergodic Theorem*	58
Exercises	59
Notes	60
Chapter 5. Coupling	63
5.1. Definition	63
5.2. Bounding Total Variation Distance	64
5.3. Examples	65
5.4. Grand Couplings	70
Exercises	73
Notes	74
Chapter 6. Strong Stationary Times	75
6.1. Top-to-Random Shuffle	75
6.2. Definitions	76
6.3. Achieving Equilibrium	77
6.4. Strong Stationary Times and Bounding Distance	78
6.5. Examples	80
6.6. Stationary Times and Cesaro Mixing Time*	83
Exercises	84
Notes	85
Chapter 7. Lower Bounds on Mixing Times	87
7.1. Counting and Diameter Bounds	87
7.2. Bottleneck Ratio	88
7.3. Distinguishing Statistics	92
7.4. Examples	96
Exercises	98
Notes	98
Chapter 8. The Symmetric Group and Shuffling Cards	99
8.1. The Symmetric Group	99
8.2. Random Transpositions	101
8.3. Riffle Shuffles	106
Exercises	109
Notes	111
Chapter 9. Random Walks on Networks	115
9.1. Networks and Reversible Markov Chains	115
9.2. Harmonic Functions	116
9.3. Voltages and Current Flows	117
9.4. Effective Resistance	118
9.5. Escape Probabilities on a Square	123
Exercises	124
Notes	125

Chapter 10. Hitting Times	127
10.1. Definition	127
10.2. Random Target Times	128
10.3. Commute Time	130
10.4. Hitting Times for the Torus	133
10.5. Bounding Mixing Times via Hitting Times	134
10.6. Mixing for the Walk on Two Glued Graphs	138
Exercises	139
Notes	141
Chapter 11. Cover Times	143
11.1. Cover Times	143
11.2. The Matthews Method	143
11.3. Applications of the Matthews Method	147
Exercises	151
Notes	152
Chapter 12. Eigenvalues	153
12.1. The Spectral Representation of a Reversible Transition Matrix	153
12.2. The Relaxation Time	154
12.3. Eigenvalues and Eigenfunctions of Some Simple Random Walks	156
12.4. Product Chains	160
12.5. An $\ell^2$ Bound	163
12.6. Time Averages	165
Exercises	167
Notes	168
Part II: The Plot Thickens	169
Chapter 13. Eigenfunctions and Comparison of Chains	171
13.1. Bounds on Spectral Gap via Contractions	171
13.2. Wilson's Method for Lower Bounds	172
13.3. The Dirichlet Form and the Bottleneck Ratio	175
13.4. Simple Comparison of Markov Chains	179
13.5. The Path Method	182
13.6. Expander Graphs*	185
Exercises	187
Notes	187
Chapter 14. The Transportation Metric and Path Coupling	189
14.1. The Transportation Metric	189
14.2. Path Coupling	191
14.3. Fast Mixing for Colorings	193
14.4. Approximate Counting	195
Exercises	198
Notes	199
Chapter 15. The Ising Model	201
15.1. Fast Mixing at High Temperature	201
15.2. The Complete Graph	203

15.3. The Cycle	204
15.4. The Tree	206
15.5. Block Dynamics	208
15.6. Lower Bound for Ising on Square*	211
Exercises	213
Notes	214
Chapter 16. From Shuffling Cards to Shuffling Genes	217
16.1. Random Adjacent Transpositions	217
16.2. Shuffling Genes	221
Exercise	226
Notes	227
Chapter 17. Martingales and Evolving Sets	229
17.1. Definition and Examples	229
17.2. Optional Stopping Theorem	231
17.3. Applications	233
17.4. Evolving Sets	235
17.5. A General Bound on Return Probabilities	239
17.6. Harmonic Functions and the Doob $h$ -Transform	241
17.7. Strong Stationary Times from Evolving Sets	243
Exercises	245
Notes	245
Chapter 18. The Cutoff Phenomenon	247
18.1. Definition	247
18.2. Examples of Cutoff	248
18.3. A Necessary Condition for Cutoff	252
18.4. Separation Cutoff	254
Exercise	255
Notes	255
Chapter 19. Lamplighter Walks	257
19.1. Introduction	257
19.2. Relaxation Time Bounds	258
19.3. Mixing Time Bounds	260
19.4. Examples	262
Notes	263
Chapter 20. Continuous-Time Chains*	265
20.1. Definitions	265
20.2. Continuous-Time Mixing	266
20.3. Spectral Gap	268
20.4. Product Chains	269
Exercises	273
Notes	273
Chapter 21. Countable State Space Chains*	275
21.1. Recurrence and Transience	275
21.2. Infinite Networks	277

21.3. Positive Recurrence and Convergence	279
21.4. Null Recurrence and Convergence	283
21.5. Bounds on Return Probabilities	284
Exercises	285
Notes	286
Chapter 22. Coupling from the Past	287
22.1. Introduction	287
22.2. Monotone CFTP	288
22.3. Perfect Sampling via Coupling from the Past	293
22.4. The Hardcore Model	294
22.5. Random State of an Unknown Markov Chain	296
Exercise	297
Notes	297
Chapter 23. Open Problems	299
23.1. The Ising Model	299
23.2. Cutoff	300
23.3. Other Problems	301
Appendix A. Background Material	303
A.1. Probability Spaces and Random Variables	303
A.2. Metric Spaces	308
A.3. Linear Algebra	308
A.4. Miscellaneous	309
Appendix B. Introduction to Simulation	311
B.1. What Is Simulation?	311
B.2. Von Neumann Unbiasing*	312
B.3. Simulating Discrete Distributions and Sampling	313
B.4. Inverse Distribution Function Method	314
B.5. Acceptance-Rejection Sampling	314
B.6. Simulating Normal Random Variables	317
B.7. Sampling from the Simplex	318
B.8. About Random Numbers	318
B.9. Sampling from Large Sets*	319
Exercises	322
Notes	325
Appendix C. Solutions to Selected Exercises	327
Bibliography	353
Notation Index	363
Index	365



## Part I: Basic Methods and Examples

*Everything should be made as simple as possible, but not simpler.*

–Paraphrase of a quotation from Einstein (1934).

