David Freedman

MARKOV CHAINS

David Freedman

MARKOV CHAINS

With 40 Figures

Springer-Verlag
New York Heidelberg Berlin

David Freedman
Department of Statistics
University of California
Berkeley, CA 94720
U.S.A.

AMS Subject Classifications: 60J10, 60J27

Library of Congress Cataloging in Publication Data
Freedman, David, 1938–
Markov chains.
Originally published: San Francisco: Holden-Day, 1971 (Holden-Day series in probability and statistics)
Bibliography: p.
Includes index.
1. Markov processes. I. Title. II. Series: Holden-Day series in probability and statistics.
QA274.7.F74 1983 519.2′33 82-19577.

The original version of this book was published by Holden-Day, Inc. in 1971.

Printed and bound by R. R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90808-0 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90808-0 Springer-Verlag Berlin Heidelberg New York

PREFACE

A long time ago I started writing a book about Markov chains, Brownian motion, and diffusion. I soon had two hundred pages of manuscript and my publisher was enthusiastic. Some years and several drafts later, I had a thousand pages of manuscript, and my publisher was less enthusiastic. So we made it a trilogy:

Markov Chains
Brownian Motion and Diffusion
Approximating Countable Markov Chains

familiarly — *MC*, *B & D*, and *ACM*.

I wrote the first two books for beginning graduate students with some knowledge of probability; if you can follow Sections 10.4 to 10.9 of *Markov Chains* you're in. The first two books are quite independent of one another, and completely independent of the third. This last book is a monograph which explains one way to think about chains with instantaneous states. The results in it are supposed to be new, except where there are specific disclaimers; it's written in the framework of *Markov Chains*.

Most of the proofs in the trilogy are new, and I tried hard to make them explicit. The old ones were often elegant, but I seldom saw what made them go. With my own, I can sometimes show you why things work. And, as I will argue in a minute, my demonstrations are easier technically. If I wrote them down well enough, you may come to agree.

The approach in all three books is constructive: I did not use the notion of separability for stochastic processes and in general avoided the uncountable axiom of choice. Separability is a great idea for dealing with any really large class of processes. For Markov chains I find it less satisfactory. To begin with, a theorem on Markov chains typically amounts to a statement about a probability on a Borel σ-field. It's a shame to have the proof depend on the existence of an unnamable set. Also, separability proofs usually have two parts. There is an abstract part which establishes the existence of a separable version. And there is a combinatorial argument, which establishes some property of the separable version by looking at the behavior of the process on a countable set of times. If you take the constructive approach, the combinatorial argument alone is enough proof.

When I started writing, I believed in regular conditional distributions. To me they're natural and intuitive objects, and the first draft was full of them. I told it like it was, and if the details were a little hard to supply, that was the reader's problem. Eventually I got tired of writing a book intelligible only to me. And I came to believe that in most proofs, the main point is estimating a probability number: the fewer complicated intermediaries, the better. So I switched to computing integrals by Fubini. This is a more powerful technique than you might think and it makes for proofs that can be checked. Virtually all the conditional distributions were banished to the Appendix. The major exception is Chapter 4 of *Markov Chains*, where the vividness of the conditional distribution language compensates for its technical difficulty.

In *Markov Chains*, Chapters 3 to 6 and 8 cover material not usually available in textbooks — for instance: invariance principles for functionals of a Markov chain; Kolmogorov's inequality on the concentration function; the boundary, with examples; and the construction of a variety of continuous-time chains from their jump processes and holding times. Some of these constructions are part of the folklore, but I think this is the first careful public treatment.

Brownian Motion and Diffusion dispenses with most of the customary transform apparatus, again for the sake of computing probability numbers more directly. The chapter on Brownian motion emphasizes topics which haven't had much textbook coverage, like square variation, the reflection principle, and the invariance principle. The chapter on diffusion shows how to obtain the process from Brownian motion by changing time.

I studied with the great men for a time, and saw what they did. The trilogy is what I learned. All I can add is my recommendation that you buy at least one copy of each book.

User's guide to *Markov Chains*

In one semester, you can cover Sections 1.1-9, 5.1-3, 7.1-3 and 9.1-3. This gets you the basic results for both discrete and continuous time. In one year you could do the whole book, provided you handle Chapters 4, 6, and 8 lightly. Chapters 2-4, 6 and 8 are largely independent of one another, treat specialized topics, and are more difficult; Section 8.5 is particularly hard. I do recommend looking at Section 6.6 for some extra grip on Markov times.

Sections 10.1-3 explain the cruel and unusual notation, and the reference system; 10.4-9 review probability theory quickly; 10.10-17 do the more exotic analyses which I've found useful at various places in the trilogy; and a few things are in 10.10-17 just because I like them.

Chapter 10 is repeated in *B & D;* Chapters 1, 5, 7 and 10 are repeated in *ACM*. The three books have a common preface and bibliography. Each has its own index and symbol finder.

Acknowledgments

Much of the trilogy is an exposition of the work of other mathematicians, who sometimes get explicit credit for their ideas. Writing *Markov Chains* would have been impossible without constant reference to Chung (1960). Doob (1953) and Feller (1968) were also heavy involuntary contributors. The diffusion part of *Brownian Motion and Diffusion* is a peasant's version of Itô and McKean (1965).

The influence of David Blackwell, Lester Dubins and Roger Purves will be found on many pages, as will that of my honored teacher, William Feller. Ronald Pyke and Harry Reuter read large parts of the manuscript and made an uncomfortably large number of excellent suggestions, many of which I was forced to accept. I also tested drafts on several generations of graduate students, who were patient, encouraging and helpful. These drafts were faithfully typed from the cuneiform by Gail Salo.

The Sloan Foundation and the US Air Force Office of Scientific Research supported me for various periods, always generously, while I did the writing. I finished two drafts while visiting the Hebrew University in Jerusalem, Imperial College in London, and the University of Tel Aviv. I am grateful to the firm of Cohen, Leithman, Kaufman, Yarosky and Fish, criminal lawyers and xerographers in Montreal. And I am still nostalgic for Cohen's Bar in Jerusalem, the caravansary where I wrote the first final draft of *Approximating Countable Markov Chains*.

David Freedman

Berkeley, California
July, 1970

Preface to the Springer edition

My books on *Markov Chains*, *Brownian Motion and Diffusion*, and *Approximating Countable Markov Chains*, were first published in the early 1970's, and have not been readily available since then. However, there still seems to be some substantial interest in them, perhaps due to their constructive and set-theoretic flavor, and the extensive use of concrete examples. I am pleased that Springer-Verlag has agreed to reprint the books, making them available again to the scholarly public. I have taken the occasion to correct many small errors, and to add a few references to new work.

David Freedman

Berkeley, California

September, 1982

TABLE OF CONTENTS

Part I. Discrete time

1. INTRODUCTION TO DISCRETE TIME

2. RATIO LIMIT THEOREMS

1

INTRODUCTION TO DISCRETE TIME

1. FOREWORD

Consider a stochastic process which moves through a countable set I of states. At stage n, the process decides where to go next by a random mechanism which depends only on the current state, and not on the previous history or even on the time n. These processes are called *Markov chains with stationary transitions and countable state space*. They are the object of study in the first part of this book. More formally, there is a countable set of states I, and a stochastic process $X_0, X_1, \ldots$ on some probability triple $(\mathscr{X}, \mathscr{F}, \mathscr{P})$, with $X_n(x) \in I$ for all nonnegative integer n and $x \in \mathscr{X}$. Moreover, there is a function P on $I \times I$ such that

$$\mathscr{P}\{X_{n+1} = j \mid X_0, \ldots, X_n\} = P(X_n, j).$$

That is, the conditional distribution of X_{n+1} given $X_0, \ldots, X_n$ depends on X_n, but not on n or on $X_0, \ldots, X_{n-1}$. The process X is said to be *Markov with stationary transitions P*, or to have *transitions P*. Suppose I is reduced to the *essential range*, namely the set of j with $\mathscr{P}\{X_n = j\} > 0$ for some n. Then the transitions P are unique, and form a stochastic matrix. Here is an equivalent characterization: X is Markov with stationary transitions P iff

$$\mathscr{P}\{X_n = j_n \text{ for } n = 0, \ldots, N\} = \mathscr{P}\{X_0 = j_0\} \prod_{n=0}^{N-1} P(j_n, j_{n+1})$$

for all N and $j_n \in I$. If $\mathscr{P}\{X_0 = j\} = 1$ for some $j \in I$, then X is said to *start*

I want to thank Richard Olshen for checking the final draft of this chapter.

from j or to have *starting state* j. This involves no real loss in generality, as one sees by conditioning on X_0.

(1) Definition. *A stochastic matrix* P *on* I *is a function on* $I \times I$, *such that:*

$$P(i,j) \geqq 0 \quad \textit{for all } i \textit{ and } j \textit{ in } I;$$

and

$$\Sigma_{j \in I}\, P(i,j) = 1 \quad \textit{for all } i \textit{ in } I.$$

If P and Q are stochastic matrices on I, so is PQ, where

$$(PQ)(i,k) = \Sigma_{j\in I}\, P(i,j)Q(j,k).$$

And so are P^n, where $P^1 = P$ and $P^{n+1} = PP^n$.

Here are three examples: let Y_n be independent and identically distributed, taking the values 1 and -1 with equal probability $\frac{1}{2}$.

(2) Example. Let $X_0 = 1$. For $n = 1, 2, \ldots$, let $X_n = Y_n$. Then $\{X_n\}$ is a Markov chain with state space $I = \{-1, 1\}$ and stationary transitions P, where $P(i,j) = \frac{1}{2}$ for all i and j in I. The starting state is 1.

(3) Example. Let $X_0 = 0$. For $n = 1, 2, \ldots$, let $X_n = X_{n-1} + Y_n$. Then $\{X_n\}$ is a Markov chain with the integers for state space and stationary transitions P, where

$$P(n, n+1) = P(n, n-1) = \tfrac{1}{2}$$

$$P(n, m) = 0 \quad \text{when } |n - m| \neq 1.$$

The starting state is 0.

(4) Example. Let $X_n = (Y_n, Y_{n+1})$ for $n = 0, 1, \ldots$. Then $\{X_n\}$ is a Markov chain with state space I and stationary transitions P, where I is the set of pairs (a, b) with $a = \pm 1$ and $b = \pm 1$, and

$$P[(a,b),(c,d)] = 0 \quad \text{when } b \neq c$$

$$= \tfrac{1}{2} \quad \text{when } b = c.$$

By contrast, let $\hat{X}_n = Y_n + Y_{n+1}$. Now $\hat{X}_n$ is a function of X_n. But $\{\hat{X}_n\}$ is not Markov.

Return to the general Markov chain X with stationary transitions. For technical reasons, it is convenient to study the distribution of X rather than X itself. The formal exposition begins in Section 3 by describing these distributions. This will be repeated here, with a brief explanation of how to translate the results back into statements about X. Introduce the space I^∞ of I-sequences. That is, I^∞ is the set of functions ω from the nonnegative integers to I. For $n = 0, 1, \ldots$, define the coordinate function ξ_n on I^∞ by

$$\xi_n(\omega) = \omega(n) \quad \text{for } \omega \in \Omega.$$

Then $\xi_0, \xi_1, \ldots$ is the coordinate process. Give I^∞ the smallest σ-field $\sigma(I^\infty)$ over which each coordinate function is measurable. Thus, $\sigma(I^\infty)$ is generated by the cylinders

$$\{\xi_0 = i_0, \ldots, \xi_n = i_n\}.$$

For any $i \in I$ and stochastic matrix P on I, there is one and only one probability P_i on I^∞ making the coordinate process Markov with stationary transitions P and starting state i. In other terms:

$$P_i\{\xi_n = i_n \text{ for } n = 0, \ldots, N\} = \Pi_{n=0}^{N-1} P(i_n, i_{n+1}),$$

for all N and $i_n \in I$ with $i_0 = i$. The probability P_i really does depend only on P and i.

Now I^∞ is the sample space for X, namely the space of all realizations. More formally, there is a mapping M from $\mathscr{X}$ to I^∞, uniquely defined by the relation

$$\xi_n(Mx) = X_n(x) \quad \text{for all } n = 0, 1, \ldots \text{ and } x \in \mathscr{X}.$$

That is, the nth coordinate of Mx is $X_n(x)$, and Mx is the sequence of states X passes through at x, namely: $(X_0(x), X_1(x), X_2(x), \ldots)$. Check that M is measurable. Fix $i \in I$ and a stochastic matrix P on I. Suppose X is Markov with stationary transitions P and starting state i. With respect to the *distribution* of X, namely $\mathscr{P}M^{-1}$, the coordinate process is Markov with stationary transitions P and starting state i. Therefore $\mathscr{P}M^{-1} = P_i$. Conversely, $\mathscr{P}M^{-1} = P_i$ implies that X is Markov with stationary transitions P and starting state i. Now probability statements about X can be translated into statements about P_i. For example, the following three assertions are all equivalent:

(5a) $$P_i\{\xi_n = i \text{ for infinitely many } n\} = 1.$$

(5b) For some Markov chain X with stationary transitions P and starting state i,

$$\mathscr{P}\{X_n = i \text{ for infinitely many } n\} = 1.$$

(5c) For all Markov chains X with stationary transitions P and starting state i,

$$\mathscr{P}\{X_n = i \text{ for infinitely many } n\} = 1.$$

Indeed, the set talked about in (5b) is the M-inverse image of the set talked about in (5a); and $P_i = \mathscr{P}M^{-1}$.

The basic theory of these processes is developed in a rapid but complete way in Sections 3–9; Sections 10, 12, and 14 present some examples, while Sections 11 and 13 cover special topics. Readers who want a more leisurely discussion of the intuitive background should look at (Feller, 1968, XV) or (Kemeny and Snell, 1960). Here is a summary of Sections 3–9.

2. SUMMARY

The main result in Section 3 is the strong Markov property. To state the best case of it, let the random variable τ on I^∞ take only the values $0, 1, \ldots, \infty$. Suppose the set $\{\tau = n\}$ is in the σ-field spanned by $\xi_0, \ldots, \xi_n$ for $n = 0, 1, \ldots$, and suppose

$$P_i\{\tau < \infty \text{ and } \xi_\tau = j\} = 1 \quad \text{for some } j \in I.$$

Then the fragment

$$(\xi_0, \ldots, \xi_\tau)$$

and the process

$$\xi_\tau, \xi_{\tau+1}, \xi_{\tau+2}, \ldots$$

are P_i-independent; the P_i-distribution of the process is P_j. This is a special case of the *strong Markov property*.

(6) Illustration. Let τ be the least n with $\xi_n = j$, and $\tau = \infty$ if there is no such n; the assumption above is $P_i\{\tau < \infty\} = 1$.

To state the results of Section 4, write:

$$i \to j \quad \text{iff} \quad P^n(i,j) > 0 \quad \text{for some } n = 1, 2, \ldots;$$

$$i \leftrightarrow j \quad \text{iff} \quad i \to j \text{ and } j \to i$$

$$i \text{ is } essential \quad \text{iff} \quad i \to j \quad \text{implies } j \to i.$$

(7) Illustration. Suppose $I = \{1, 2, 3, 4\}$ and P is this matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Then 1,2,3 are essential and 4 is inessential. Moreover, $1 \leftrightarrow 1$ while $2 \leftrightarrow 3$.

For the rest of this summary,

$$\text{suppose} \quad \text{all } i \in I \text{ are essential.}$$

Then $\leftrightarrow$ is an equivalence relation. For the rest of this summary, suppose that I consists of one equivalence class, namely,

$$\text{suppose} \quad i \to j \text{ and } j \to i \quad \text{for all } i \text{ and } j \text{ in } I.$$

Let *period* i be the greatest common denominator (g.c.d.) of the set of $n > 0$ with $P^n(i, i) > 0$. Then period i does not depend on i; say it is d. And I is the disjoint union of sets $C_0, C_1, \ldots, C_{d-1}$, such that

$$i \in C_n \quad \text{and} \quad P(i, j) > 0 \quad \text{imply } i \in C_{n \oplus 1},$$

where $\oplus$ means addition modulo d.

(8) Illustration. Suppose $I = \{1, 2, 3, 4\}$ and P is this matrix:

$$\begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

Then I has period 2, and $C_0 = \{1, 2\}$ and $C_1 = \{3, 4\}$.

For the rest of the summary,

$$\text{suppose} \quad \text{period } i = 1 \quad \text{for all } i \in I.$$

To state the result of Section 5, say

$$i \text{ is } \textit{recurrent} \quad \text{iff} \quad P_i\{\xi_n = i \text{ for infinitely many } n\} = 1$$
$$i \text{ is } \textit{transient} \quad \text{iff} \quad P_i\{\xi_n = i \text{ for infinitely many } n\} = 0.$$

This classification is exhaustive. Namely, the state i is either recurrent or transient, according as $\Sigma_n P^n(i, i)$ is infinite or finite. And all $i \in I$ are recurrent or transient together. These results follow from the strong Markov property. Parenthetically, under present assumptions: if I is finite, all $i \in I$ are recurrent.

(9) Example. Suppose $I = \{0, 1, 2, \ldots\}$. Let $0 < p_n < 1$. Suppose $P(0, 1) = 1$ and for $n = 1, 2, \ldots$ suppose $P(n, n + 1) = p_n$ and $P(n, 0) = 1 - p_n$. Suppose all other entries in P vanish; see Figure 1. The states are recurrent or transient according as $\Pi\, p_n$ is zero or positive.

HINT. See (16) below. ★

For the rest of this summary,

$$\text{suppose} \quad \text{all } i \in I \text{ are recurrent.}$$

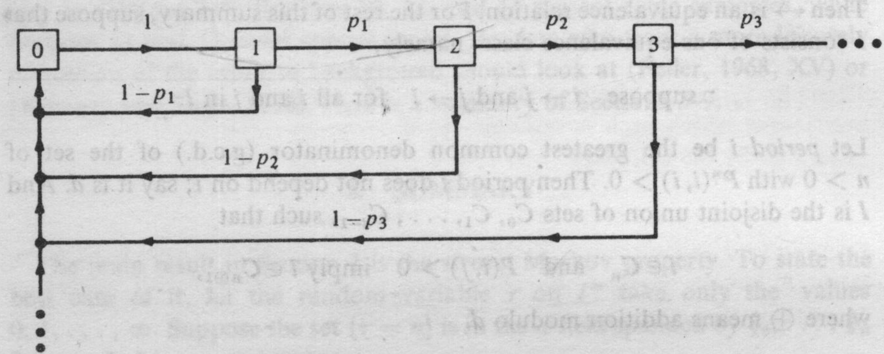

Figure 1

To state the result of Section 6, let $Y_1, Y_2, \ldots$ be a sequence of independent, identically distributed random variables, taking only the values $1, 2, 3, \ldots$ with probabilities $p_1, p_2, p_3, \ldots$. Let $\mu = \Sigma\, np_n$, and suppose

$$\text{g.c.d. } \{n : p_n > 0\} = 1.$$

Let $U(m)$ be the probability that

$$Y_1 + \cdots + Y_n = m \quad \text{for some } n = 0, 1, 2, \ldots.$$

Then

$$\lim_{m\to\infty} U(m) = 1/\mu.$$

This result is called the *renewal theorem*. It is used in Section 7, together with strong Markov, to show that

$$\lim_{n\to\infty} P^n(i, j) = \pi(j),$$

where $1/\pi(j)$ is the P_j-expectation of the least $m > 0$ with $\xi_m = j$.

To state the result of Section 8, say

j is *positive recurrent* iff $\pi(j) > 0$

j is *null recurrent* iff $\pi(j) = 0$.

Then all $i \in I$ are either positive recurrent or null recurrent together.

(10) Example. Let $I = \{0, 1, 2, \ldots\}$. Let $p_n > 0$ and $\Sigma_{n=1}^{\infty} p_n = 1$. Let $P(0, n) = p_n$ and $P(n, n-1) = 1$ for $n = 1, 2, \ldots$. See Figure 2. The states are positive recurrent or null recurrent according as $\Sigma_{n=1}^{\infty} np_n$ is finite or infinite.