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.

MARKOV CHAINS

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.

© 1971 by Holden-Day Inc.

© 1983 by David A. Freedman

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, N.Y. 10010, U.S.A.

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

987654321

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

The source of the bill three beenlest is constructive. I did not use the notion

able axiom of choice. Separability is a great idea for degling with any really

PREFACE SANTAGE OF THE 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:

reader's problem. Eventually I got fixed of writing a book intelligible only

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

PREFACE viii

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 continuoustime 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.

ew. Yerk Heidelberg fixthin

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 X PREFACE

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.

The influence of David Blackwell Lester Digities and Roger Purves with

Romand Pyko and Herry Repter read large parts of the manuscrips and made

textissium of Cohon Leidmen, Kaufman, Zurosky and Fish, orini at law-

David Freedman

Berkeley, California
September, 1982

ome invariance principles

An example of Dyson

Introduction M. ton

59

Part I. Discrete time

1	INTRODUCTION	TO	DISCOUTE	TIME

5.

Restricting the range

1.	Foreword		1
2.	Summary		4
3.	The Markov and strong Markov properties		7
4.	Classification of states	1	6
5.	Recurrence	1	9
6.	The renewal theorem	2	2
7.	The limits of Pn solling and polylen A	2	5
8.	Positive recurrence	. 2	6
9.	Invariant probabilities and an additional and an additional and a second a second and a second a	2	9
10.	The Bernoulli walk	3	2
11.	Forbidden transitions	3	4
12.	The Harris walk	. 3	6
13.	The tail σ-field and a theorem of Orey	3	9
14.	Examples 38AO A.MATE UNI 904 23.19MA	4.	5
RAT	TIO LIMIT THEOREMS noithertence tend of T		
1.	Examples on the first construction . notice that	£ 4	7
2.		4	-
3.	Proofs of Derman and Doeblin	. 5	
4	Variations Samu voltam	.0 5	

	6.	Proof of Kingman-Orey	64 70
	7.	An example of Dyson	73
	8.	Almost everywhere ratio limit theorems	75
	9.	The sum of a function over different j-blocks	13
3.	SOM	IE INVARIANCE PRINCIPLES	
	1.	Introduction	82
	2.	Estimating the partial sums	83
	3.	The number of positive sums	87
	4.	Some invariance principles	95
	5.	The concentration function	99
4.	THE	BOUNDARY	
	1.	Introduction	111
	2.	Proofs	113
	3.	A convergence theorem	121
	4.	Examples	124
	5.	The last visit to i before the first visit to $J\setminus\{i\}$	132
		brownio 4 - 4	
		Part II. Continuous time	
5.	INT	RODUCTION TO CONTINUOUS TIME	100
	1.	Semigroups and processes	138
	. 2.	Analytic properties	142
	3.	Uniform semigroups	150
	4.	Uniform substochastic semigroups	150
	5.	The exponential distribution	154
DE.	6.	The step function case	165
	7.	The uniform case	103
6.	FY	AMPLES FOR THE STABLE CASE	AL.
0.	LA		172
	1.	Introduction EMSSOSETTEMATON	172
	2.	The first construction	173
	3.	Examples on the first construction	181
	4.	The second construction	197
	5.	Examples on the second construction	203
	6.	Markov times another Y	210
02	7.	Crossing the infinities	210

CONTENTS	XIII

7.	THE	STABLE CASE		
	1.	Introduction insument argument - Introduction		216
	2.	Regular sample functions	. G 3	217
	3.	The post-exit process		223
360	4.	The strong Markov property	Ài	229
361	5.	The minimal solution	7	237
365	6.	The backward and forward equations		243
8.	MOI	RE EXAMPLES FOR THE STABLE CASE		Idia
			X	252
373	1.	An oscillating semigroup		260
379	2.	A semigroup with an infinite second derivative		266
	3. 4.	Large oscillations in $P(t,1,1)$ An example of Speakman		271
	5.	The embedded jump process is not Markov		273
	6.	Isolated infinities		292
	7.	The set of infinities is bad		295
		The set of minutes to sur		
9.	THI	E GENERAL CASE		
	1.	An example of Blackwell		297
	2.	Quasiregular sample functions		299
	3.	The sets of constancy		308
	4.	The strong Markov property		315
	5.	The post-exit process		323
	6.	The abstract case		326
		and a stochastic process X_0, X_1, \dots on some probably with $X_0(x) \in \mathcal{X}$ for all nonnegative integer n and $x \in \mathcal{X}$.		EGVez
		function P on I x 1 Part III.		
10.	API	PENDIX		
	1.	Notation		329
	2.	Numbering		
	3.	Ribliography		
	4.	The abstract Lebesgue integral		331
The	5.	Atoms		334
	6.	Independence		337
	7.	Conditioning		338
	8.	Martingales		339
	9.	Metric spaces		346
	10.	Regular conditional distributions		347

CONTENTS

11. The Kolmogorov consistency theorem 12. The diagonal argument 13. Classical Lebesgue measure 14. Real variables 15. Absolute continuity 16. Convex functions 17. Complex variables	1 1 2 2 4 4 5 5 6 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6	353 354 356 357 360 361 365
RE EXAMPLES FOR THE STABLE CASE	OM	367
An oscillating semigroup A semigroup with an infinite second derivative Large oscillations in P(4,1,1) REDUIT LORGE OSCILLATION OF THE SECOND OF THE SECO	1. 2. 3.	373 379
An example of Speakman The embedded jump process is not Markove Isolated infinities The set of infinities is bad		161
An example of Blackwell Ouasiregular sample functions The sets of constantly strong Markov property The strong Markov property The post-exit process OUVII FOOD OF MOTORS 233	1.5.4.2.	
The abstract case essessors been sainted as a sessor of be		138 142 147 150 162
Notation Notation Numbering Numbering Bibliography Bibliography Bibliography	1. 2. 3.	
The abstract Lebesgue integral Atoms Atoms Independence doi.surmess Conditioning and active and active and active and active and active ac	4. 6. 5. 7. 6. 9. 7. 9. 9. 9. 9. 9. 9. 9. 9. 9. 9. 9. 9. 9.	

INTRODUCTION TO DISCRETE TIME

And so aret Phywhole P! \$ P. and P! 1 TPP" A haw 13 A bac William for

taking the values of anti-offwith equal-probability & see as the value

Example. Let $X_0 = 1$. For n = 1, 2. Let $X_n = Y_n$. Then (X_n) is a Markov chain with state spanning (X_n) and state in (X_n) and state is (X_n) .

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 $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, with $X_n(x) \in I$ for all nonnegative integer n and $x \in \mathcal{X}$. Moreover, there is a function P on $I \times I$ such that

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

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 $\mathcal{P}\{X_n=j\}>0$ 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

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

for all N and $j_n \in I$. If $\mathcal{P}\{X_0 = j\} = 1$ 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:

$$P(i,j) \ge 0$$
 for all i and j in I;

and

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

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

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

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

Here are three examples: let Y, be independent and identically distributed, taking the values 1 and -1 with equal probability 1/2.

- (2) Example. Let $X_0 = 1$. For $n = 1, 2, ..., let <math>X_n = Y_n$. 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 $n = 1, 2, ..., let <math>X_n = X_{n-1} + Y_n$. Then $\{X_n\}$ is a Markov chain with the integers for state space and stationary transitions P, where the state themes out no who abnoach dolder mains

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

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

The starting state is 0.

 $(\mathcal{Z}, \mathcal{F}, \mathcal{P})$, with $X_n(x) \in I$ for all nonnegative integer n and (4) Example. Let $X_n = (Y_n, Y_{n+1})$ for n = 0, 1, ... 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

That is, the condition
$$p(a,b)$$
, $p(a,b)$, $p(a,b)$, when $p(a,b)$ when $p(a,b)$ when $p(a,b)$ when $p(a,b)$ when $p(a,b)$ where $p(a,b)$ is said to be Markov Markov.

 $= \frac{1}{2} \quad \text{when } b = c.$ 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 Then the transitions P are unique, and form a stochastic mair volam ton

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 Isequences. 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

$$\xi_n(\omega) = \omega(n)$$
 for $\omega \in \Omega$.

Then ξ_0, ξ_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:

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

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

$$\xi_n(Mx) = X_n(x)$$
 for all $n = 0, 1, \dots$ and $x \in \mathcal{X}$.

That is, the *n*th 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 $\mathcal{P}M^{-1}$, the coordinate process is Markov with stationary transitions P and starting state i. Therefore $\mathcal{P}M^{-1} = P_i$. Conversely, $\mathcal{P}M^{-1} = P_i$ 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 $P_i = \mathcal{P}M^{-1}$.

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 ξ_0, \ldots, ξ_n for $n = 0, 1, \ldots$, and suppose

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

Then the fragment space of the same with the space store and all T wow

More formally, there is a man
$$(1, \frac{3}{2}, \dots, \frac{3}{6})$$
 . \mathcal{L} to \mathcal{L}^{∞} , uniquely defined by

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:

$$i \rightarrow j$$
 iff $P^n(i,j) > 0$ for some $n = 1, 2, ...$; $i \leftrightarrow j$ iff $i \rightarrow j$ and $j \rightarrow i$
 i is essential iff $i \rightarrow j$ implies $j \rightarrow i$.

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

(5b) For some Markov chain
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & \sqrt{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} &$$

Then 1,2,3 are essential and 4 is inessential. Moreover, $1 \leftrightarrow 1$ while $2 \leftrightarrow 3$. For the rest of this summary,

suppose all $i \in I$ are essential.

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

suppose
$$i \rightarrow j$$
 and $j \rightarrow i$ for all i and j 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

$$i \in C_n$$
 and $P(i,j) > 0$ 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:

To state the result of Section
$$\begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & 0 & 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & 0 & 0 & \frac{2}{3} & \frac{2}{3} \end{pmatrix}$$
 to the result of Section $\begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & \frac{2}{3} & \frac{2}$

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

For the rest of the summary,

suppose period i = 1 for all $i \in I$.

Let U(m) be the probability that

To state the result of Section 5, say

i is recurrent iff
$$P_i\{\xi_n = i \text{ for infinitely many } n\} = 1$$

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

This classification is exhaustive. Namely, the state i is either recurrent or transient, according as $\sum_{n} P^{n}(i, i)$ 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 $I = \{0, 1, 2, ...\}$. Let $0 < p_n < 1$. Suppose P(0, 1) = 1 and for n = 1, 2, ... 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 $\prod p_n$ is zero or positive.

HINT. See (16) below.

For the rest of this summary, senses the same and assume a subsequent

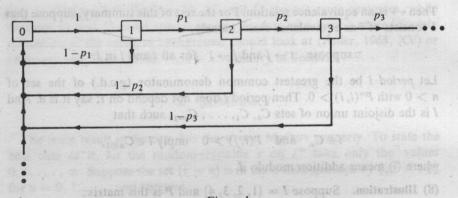


Figure 1

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 $\mu = \sum n p_n$, and suppose

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

Let U(m) be the probability that

$$Y_1 + \cdots + Y_n = m$$
 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

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

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