A SECOND COURSE IN STOCHASTIC PROCESSES

Samuel Karlin
Howard M. Taylor

A SECOND COURSE IN STOCHASTIC PROCESSES

This page intentionally left blank

A SECOND COURSE IN STOCHASTIC PROCESSES

SAMUEL KARLIN
STANFORD UNIVERSITY

HOWARD M. TAYLOR
CORNELL UNIVERSITY

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
New York London Toronto Sydney San Francisco

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Karlin, Samuel, Date.
A second course in stochastic processes.

Continues A first course in stochastic processes.
Includes bibliographical references and index.
1. Stochastic processes. I. Taylor, Howard M., joint author. II. Title.
QA274.K372 519.2 80–533
ISBN 0–12–398650–8

PRINTED IN THE UNITED STATES OF AMERICA

81 82 83 84 9 8 7 6 5 4 3 2 1

CONTENTS

Chapter 10

ALGEBRAIC METHODS IN MARKOV CHAINS

Chapter 11

RATIO THEOREMS OF TRANSITION PROBABILITIES AND APPLICATIONS

Chapter 12

SUMS OF INDEPENDENT RANDOM VARIABLES AS A MARKOV CHAIN

Chapter 13

ORDER STATISTICS, POISSON PROCESSES, AND APPLICATIONS

Chapter 14

CONTINUOUS TIME MARKOV CHAINS

Chapter 15

DIFFUSION PROCESSES

Chapter 16

COMPOUNDING STOCHASTIC PROCESSES

Chapter 17

FLUCTUATION THEORY OF PARTIAL SUMS OF INDEPENDENT IDENTICALLY DISTRIBUTED RANDOM VARIABLES

Chapter 18

QUEUEING PROCESSES

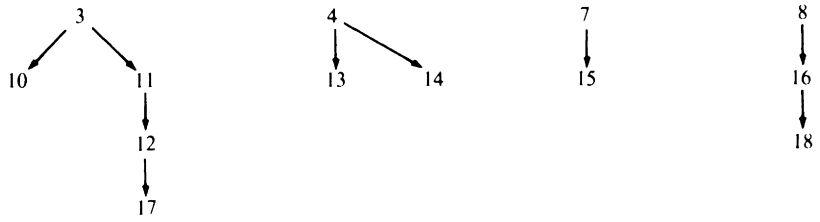
PREFACE

This *Second Course* continues the development of the theory and applications of stochastic processes as promised in the preface of *A First Course.* We emphasize a careful treatment of basic structures in stochastic processes in symbiosis with the analysis of natural classes of stochastic processes arising from the biological, physical, and social sciences.

Apart from expanding on the topics treated in the first edition of this work but not incorporated in *A First Course,* this volume presents an extensive introductory account of the fundamental concepts and methodology of diffusion processes and the closely allied theory of stochastic differential equations and stochastic integrals. A multitude of physical, engineering, biological, social, and managerial phenomena are either well approximated or reasonably modeled by diffusion processes; and modern approaches to diffusion processes and stochastic differential equations provide new perspectives and techniques impinging on many subfields of pure and applied mathematics, among them partial differential equations, dynamical systems, optimal control problems, statistical decision procedures, operations research, studies of economic systems, population genetics, and ecology models.

A new chapter discusses the elegant and far-reaching distributional formulas now available for a wide variety of functionals (e.g., first-passage time, maximum, order statistics, occupation time) of the process of sums of independent random variables. The identities, formulas, and results in this chapter have important applications in queueing and renewal theory, for statistical decision procedures, and elsewhere.

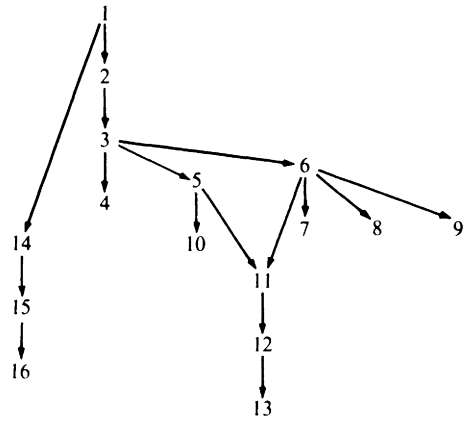
The logical dependence of the chapters in *A Second Course* is shown by the diagram below (consult also the preface to *A First Course* on the relationships of Chapters 1–9).



The book can be coupled to *A First Course* in several ways, depending on the background and interests of the students. The discussion of Markov chains in *A First Course* can be supplemented with parts of the more advanced Chapters 10–12, and 14. The material on fluctuation theory of sums of independent random variables (Chapters 12 and 17), perhaps supplemented by some parts of the chapter on queueing processes (Chapter 18), may be attractive and useful to students of operations research and statistics. Chapter 16, on compounding stochastic processes, is designed as an enticing introduction to a hierarchy of relevant models, including models of multiple species population growth, of migration and demographic structures, of point processes, and compositions of Poisson processes (Lévy processes).

We strongly recommend devoting a semester to diffusion processes (Chapter 15). The dependence relationships of the sections of Chapter 15 are diagrammed below. Section 1 provides a generalized description of various characterizations of diffusion. The examples of Section 2, which need not be absorbed in their totality, are intended to hint at the rich diversity of natural models of diffusion processes; the emphasis on biological examples reflects the authors' personal interests, but diffusion models abound in other sciences as well. Sections 3–5 point up the utility and tractability of diffusion process analysis. Section 6 takes up the boundary classification of one-dimensional diffusion processes; Section 7, on the same topic, is more technical. Section 8 provides constructions of diffusions with different types of boundary behavior. Sections 9 and 10 treat a number of topics motivated by problems of population genetics and statistics. The formal (general) theory of Markov processes with emphasis on applications to diffusions is elaborated in Sections 11 and 12. Section 13 exhibits the spectral representations for several classical diffusion models, which are of some interest because of their connections with classical special functions. The key concepts, a host of examples, and some methods of stochastic differential equations and stochastic integrals are introduced in Sections 14–16.

Chapter 15



As noted in the earlier prefaces, we have drawn freely on the thriving literature of applied and theoretical stochastic processes without citing specific articles. A few representative books are listed at the end of each chapter and may be consulted profitably as a guide to more advanced material.

We express our gratitude to Stanford University, the Weizmann Institute of Science in Israel, and Cornell University for providing a rich intellectual environment and facilities indispensable for the writing of this text. The first author is grateful for the continuing grant support provided by the National Science Foundation and the National Institutes of Health that permitted an unencumbered concentration on a number of the concepts of this book and on its various drafts. We are also happy to acknowledge our indebtedness to many colleagues who have offered constructive criticisms, among them Professor M. Taqqu of Cornell, Dr. S. Tavaré of the University of Utah, Professors D. Iglehart and M. Harrison of Stanford, and Professor J. Kingman of Oxford. Finally, we thank our students P. Glynn, E. Cameron, J. Raper, R. Smith, L. Tierney, and P. Williams for their assistance in checking the problems, and for their helpful reactions to early versions of Chapter 15.

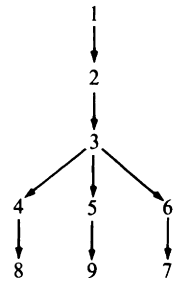
PREFACE TO *A FIRST COURSE*

The purposes, level, and style of this new edition conform to the tenets set forth in the original preface. We continue with our tack of developing simultaneously theory and applications, intertwined so that they refurbish and elucidate each other.

We have made three main kinds of changes. First, we have enlarged on the topics treated in the first edition. Second, we have added many exercises and problems at the end of each chapter. Third, and most important, we have supplied, in new chapters, broad introductory discussions of several classes of stochastic processes not dealt with in the first edition, notably martingales, renewal and fluctuation phenomena associated with random sums, stationary stochastic processes, and diffusion theory.

Martingale concepts and methodology have provided a far-reaching apparatus vital to the analysis of all kinds of functionals of stochastic processes. In particular, martingale constructions serve decisively in the investigation of stochastic models of diffusion type. Renewal phenomena are almost equally important in the engineering and managerial sciences especially with reference to examples in reliability, queueing, and inventory systems. We discuss renewal theory systematically in an extended chapter. Another new chapter explores the theory of stationary processes and its applications to certain classes of engineering and econometric problems. Still other new chapters develop the structure and use of diffusion processes for describing certain biological and physical systems and fluctuation properties of sums of independent random variables useful in the analyses of queueing systems and other facets of operations research.

The logical dependence of chapters is shown by the diagram below. Section 1 of Chapter 1 can be reviewed without worrying about details. Only Sections 5 and 7 of Chapter 7 depend on Chapter 6. Only Section 9 of Chapter 9 depends on Chapter 5.



An easy one-semester course adapted to the junior-senior level could consist of Chapter 1, Sections 2 and 3 preceded by a cursory review of Section 1, Chapter 2 in its entirety, Chapter 3 excluding Sections 5 and/or 6, and Chapter 4 excluding Sections 3, 7, and 8. The content of the last part of the course is left to the discretion of the lecturer. An option of material from the early sections of any or all of Chapters 5–9 would be suitable.

The problems at the end of each chapter are divided into two groups: the first, more or less elementary; the second, more difficult and subtle.

The scope of the book is quite extensive, and on this account, it has been divided into two volumes. We view the first volume as embracing the main categories of stochastic processes underlying the theory and most relevant for applications. In *A Second Course* we introduce additional topics and applications and delve more deeply into some of the issues of *A First Course*. We have organized the edition to attract a wide spectrum of readers, including theorists and practitioners of stochastic analysis pertaining to the mathematical, engineering, physical, biological, social, and managerial sciences.

The second volume of this work, *A Second Course in Stochastic Processes,* will include the following chapters: (10) Algebraic Methods in Markov Chains; (11) Ratio Theorems of Transition Probabilities and Applications; (12) Sums of Independent Random Variables as a Markov Chain; (13) Order Statistics, Poisson Processes, and Applications; (14) Continuous Time Markov Chains; (15) Diffusion Processes; (16) Compounding Stochastic Processes; (17) Fluctuation Theory of Partial Sums of Independent Identically Distributed Random Variables; (18) Queueing Processes.

As noted in the first preface, we have drawn freely on the thriving literature of applied and theoretical stochastic processes. A few representative references are included at the end of each chapter; these may be profitably consulted for more advanced material.

We express our gratitude to the Weizmann Institute of Science, Stanford University, and Cornell University for providing a rich intellectual environment and facilities indispensable for the writing of this text. The first author is grateful for the continuing grant support provided by the Office of Naval Research that permitted an unencumbered concentration on a number of the concepts and drafts of this book. We are also happy to acknowledge our indebtedness to many colleagues who have offered a variety of constructive criticisms. Among others, these include Professors P. Brockwell of La Trobe, J. Kingman of Oxford, D. Iglehart and S. Ghurye of Stanford, and K. Itô and S. Stidham, Jr. of Cornell. We also thank our students M. Nedzela and C. Macken for their assistance in checking the problems and help in reading proofs.

SAMUEL KARLIN
HOWARD M. TAYLOR

PREFACE TO FIRST EDITION

Stochastic processes concern sequences of events governed by probabilistic laws. Many applications of stochastic processes occur in physics, engineering, biology, medicine, psychology, and other disciplines, as well as in other branches of mathematical analysis. The purpose of this book is to provide an introduction to the many specialized treatises on stochastic processes. Specifically, I have endeavored to achieve three objectives: (1) to present a systematic introductory account of several principal areas in stochastic processes, (2) to attract and interest students of pure mathematics in the rich diversity of applications of stochastic processes, and (3) to make the student who is more concerned with application aware of the relevance and importance of the mathematical subleties underlying stochastic processes.

The examples in this book are drawn mainly from biology and engineering but there is an emphasis on stochastic structures that are of mathematical interest or of importance in more than one discipline. A number of concepts and problems that are currently prominent in probability research are discussed and illustrated.

Since it is not possible to discuss all aspects of this field in an elementary text, some important topics have been omitted, notably stationary stochastic processes and martingales. Nor is the book intended in any sense as an authoritative work in the areas it does cover. On the contrary, its primary aim is simply to bridge the gap between an elementary probability course and the many excellent advanced works on stochastic processes.

Readers of this book are assumed to be familiar with the elementary theory of probability as presented in the first half of Feller's classic *Introduction to*

Probability Theory and Its Applications. In Section 1, Chapter 1 of my book the necessary background material is presented and the terminology and notation of the book established. Discussions in small print can be skipped on first reading. Excercises are provided at the close of each chapter to help illuminate and expand on the theory.

This book can serve for either a one-semester or a two-semester course, depending on the extent of coverage desired.

In writing this book, I have drawn on the vast literature on stochastic processes. Each chapter ends with citations of books that may profitably be consulted for further information, including in many cases bibliographical listings.

I am grateful to Stanford University and to the U.S. Office of Naval Research for providing facilities, intellectual stimulation, and financial support for the writing of this text. Among my academic colleagues I am grateful to Professor K. L. Chung and Professor J. McGregor of Stanford for their constant encouragement and helpful comments; to Professor J. Lamperti of Dartmouth, Professor J. Kiefer of Cornell, and Professor P. Ney of Wisconsin for offering a variety of constructive criticisms; to Dr. A. Feinstein for his detailed checking of substantial sections of the manuscript, and to my students P. Milch, B. Singer, M. Feldman, and B. Krishnamoorthi for their helpful suggestions and their assistance in organizing the exercises. Finally, I am indebted to Gail Lemmond and Rosemarie Stampfel for their superb technical typing and all-around administrative care.

SAMUEL KARLIN

CONTENTS OF *A FIRST COURSE*

Chapter 10

ALGEBRAIC METHODS IN MARKOV CHAINS

Many important results concerning Markov chains can be obtained by using either purely algebraic methods or a combination of probabilistic and algebraic techniques. We will develop a number of these techniques in the present chapter. In order not to disrupt the continuity of presentation, we present here only a brief summary of some basic facts of matrix theory needed immediately. A fairly complete discussion of these results is given in the Appendix to *A First Course*.

1: Preliminaries

Of fundamental importance in considerations of Markov chains is the computation of the n-step transition probabilities. (Special methods are developed in Sections 4–6 applicable in the case where the Markov chain is a random walk.) To this end, we develop the necessary machinery involving the theory of eigenvalues and eigenvectors.†

(a) *Spectral Representation*

Let $\mathbf{A}$ be an $n \times n$ matrix. A nonzero n-dimensional vector $\mathbf{x}$ which satisfies the relation $\mathbf{Ax} = \lambda\mathbf{x}$ for some number λ is called a right eigenvector of $\mathbf{A}$, with corresponding eigenvalue λ. If $\mathbf{xA} = \lambda\mathbf{x}$, we call $\mathbf{x}$ a left eigenvector of $\mathbf{A}$. If there exists a complete linearly independent family $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ of right (or, alternatively, left) eigenvectors of $\mathbf{A}$, then there exists a linearly independent family $\boldsymbol{\phi}^{(1)}, \ldots, \boldsymbol{\phi}^{(n)}$ of right eigenvectors of $\mathbf{A}$ and a linearly independent family $\boldsymbol{\psi}^{(1)}, \ldots, \boldsymbol{\psi}^{(n)}$ of left eigenvectors of $\mathbf{A}$ which are biorthogonal. This means that

$$(\boldsymbol{\phi}^{(i)}, \boldsymbol{\psi}^{(j)})^{\ddagger} \equiv \sum_{k=1}^{n} \varphi_{ik}\overline{\psi}_{jk} = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

† The reader unfamiliar with the basic theory of eigenvalues and eigenvectors of matrices should consult the Appendix of *A First Course in Stochastic Processes* at this point.

‡ $(\mathbf{a}, \mathbf{b})$ denotes the inner product of the vectors $\mathbf{a}$ and $\mathbf{b}$.

where $\boldsymbol{\phi}^{(i)} = (\varphi_{i1}, \ldots, \varphi_{in})$, $\boldsymbol{\psi}^{(j)} = (\psi_{j1}, \ldots, \psi_{jn})$, and $\overline{\psi}_{jh}$ denotes the complex conjugate of ψ_{jh}. In this case, the matrix $\mathbf{A}$ is said to be *diagonalizable*. Let

$$\boldsymbol{\Phi} = \begin{Vmatrix} \varphi_{11} & \cdots & \varphi_{n1} \\ \vdots & & \vdots \\ \varphi_{1n} & \cdots & \varphi_{nn} \end{Vmatrix}, \qquad \boldsymbol{\Psi} = \begin{Vmatrix} \overline{\psi}_{11} & \cdots & \overline{\psi}_{1n} \\ \vdots & & \vdots \\ \overline{\psi}_{n1} & \cdots & \overline{\psi}_{nn} \end{Vmatrix},$$

$$\boldsymbol{\Lambda} = \begin{Vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{Vmatrix},$$

where $\lambda_1, \ldots, \lambda_n$ are the (not necessarily distinct) eigenvalues associated with the eigenvectors $\boldsymbol{\phi}^{(1)}, \ldots, \boldsymbol{\phi}^{(n)}$. (Notice that we have not labeled the element of $\boldsymbol{\Phi}$ in the usual order.) Then $\mathbf{A}$ possesses a spectral representation as a product of three special matrices:

$$\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi}.$$

Using the relation $(\boldsymbol{\phi}^{(i)}, \boldsymbol{\psi}^{(j)}) = \delta_{ij}$, we can verify by direct calculation that $\boldsymbol{\Psi}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Psi} = \mathbf{I}$ ($\mathbf{I}$ = the identity matrix). Then $\mathbf{A}^2 = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi} = \boldsymbol{\Phi}\boldsymbol{\Lambda}^2\boldsymbol{\Psi}$ and generally

$$\mathbf{A}^m = \boldsymbol{\Phi}\boldsymbol{\Lambda}^m\boldsymbol{\Psi}, \tag{1.1}$$

where obviously

$$\boldsymbol{\Lambda}^m = \begin{Vmatrix} \lambda_1^m & 0 & \cdots & 0 \\ 0 & \lambda_2^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^m \end{Vmatrix}.$$

When $\mathbf{A}$ is a Markov matrix, formula (1.1) provides a convenient representation of the mth-step transition probability matrix. Its effective use requires determining a complete set of left and right eigenvectors.

(b) *Positive Matrices*

Let $\mathbf{A}$ be a real matrix which has at least one positive element and no negative elements; we write $\mathbf{A} > \mathbf{0}$ and call $\mathbf{A}$ positive. If every element of $\mathbf{A}$ is positive, we write $\mathbf{A} \gg \mathbf{0}$ and call $\mathbf{A}$ strictly positive. The following results are known.

To each $\mathbf{A} > \mathbf{0}$ there corresponds a number $r(\mathbf{A}) \geq 0$, the spectral radius of $\mathbf{A}$, which is zero if and only if $\mathbf{A}^m = \mathbf{0}$ for some integer $m > 0$. In any case there are positive vectors $\mathbf{f}$, $\mathbf{x} > 0$, such that $\mathbf{A}\mathbf{x} = r(\mathbf{A})\mathbf{x}$, $\mathbf{f}\mathbf{A} = r(\mathbf{A})\mathbf{f}$. If λ is any eigenvalue of $\mathbf{A}$, then $|\lambda| \leq r(\mathbf{A})$; if $|\lambda| = r(\mathbf{A})$, then $\eta = \lambda/r(\mathbf{A})$ is a root of unity, i.e., $\eta^k = 1$ for some integer k, and $\eta^m r(\mathbf{A})$ is an eigenvalue of $\mathbf{A}$ for $m = 1, 2, \ldots$. Finally, suppose that $\mathbf{A}^m \gg \mathbf{0}$ for some $m > 0$; then $\mathbf{x}$ and $\mathbf{f}$ are strictly positive

vectors and uniquely determined up to a constant factor. Moreover $|\lambda| < r(\mathbf{A})$ if λ is an eigenvalue of $\mathbf{A}$ different from $r(\mathbf{A})$.

2: *Relations of Eigenvalues and Recurrence Classes*

The foregoing results have immediate application to the study of finite-state Markov chains.

Let $\mathbf{P} = \|P_{ij}\|$, $i, j = 1, \ldots, n$, be a matrix of transition probabilities. Evidently $\mathbf{P} > \mathbf{0}$. Let $\mathbf{x}$ be any vector satisfying $\sum_{i=1}^{n} x_i = 1$. Then

$$\mathbf{xP} = \left(\sum_{i=1}^{n} x_i P_{i1}, \sum_{i=1}^{n} x_i P_{i2}, \ldots, \sum_{i=1}^{n} x_i P_{in} \right).$$

Now

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n} x_i P_{ij} \right) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} P_{ij} = \sum_{i=1}^{n} x_i = 1. \tag{2.1}$$

We claim that $\mathbf{xP} \geq \lambda \mathbf{x}$ cannot hold with $\mathbf{x} > \mathbf{0}$ for any value of $\lambda > 1$, so that $r(\mathbf{P}) \leq 1$. In fact, summing the components of both sides in $\mathbf{xP} \geq \lambda \mathbf{x}$ as in (2.1) yields $\sum_{i=1}^{n} x_i \geq \lambda \sum_{i=1}^{n} x_i$. Since $\sum_{i=1}^{n} x_i > 0$, we can cancel this factor, which implies that $\lambda \leq 1$.

On the other hand, the vector $(1, \ldots, 1)$ is immediately seen to be a right eigenvector of $\mathbf{P}$ with eigenvalue 1; thus $r(\mathbf{P}) = 1$.

The property that 1 is an eigenvalue with a corresponding positive left eigenvector for any finite Markov matrix can also be deduced from Theorem 1.3 of Chapter 3.† We know that in a finite state Markov chain at least one state (and therefore at least one class) is positive recurrent. Relabeling the states if necessary, we may assume that the states $i = 1, \ldots, s$ form a positive recurrent class. Therefore $P_{ij} = 0$ for any pair i, j for which $i \in \{1, \ldots, s\}$ and $j \in \{s + 1, \ldots, n\}$. Thus, $\mathbf{P}$ has the form

$$\mathbf{P} = \begin{Vmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{B} & \mathbf{C} \end{Vmatrix} \tag{2.2}$$

and $\mathbf{P}_1$ forms an $s \times s$ Markov matrix. Now the basic limit theorem of Markov chains (see Theorem 1.3 of Chapter 3) asserts the existence of $\pi_1, \ldots, \pi_s$ such that $\pi_i > 0$,

$$\sum_{i=1}^{s} \pi_i P_{ij} = \pi_j, \qquad j = 1, \ldots, s,$$

and

$$\sum_{i=1}^{s} \pi_i = 1.$$

† Chapters 1–9 are in *A First Course in Stochastic Processes* (Second Edition, 1975).

Let $\mathbf{x}^0 = (\pi_1, \ldots, \pi_s, 0, \ldots, 0)$; we may verify at once because of the special structure of $\mathbf{P}$ as displayed in (2.2) that $\mathbf{x}^0\mathbf{P} = \mathbf{x}^0$. A slightly more detailed analysis yields the following:

***Theorem* 2.1.** *If* $\mathbf{P}$ *is a finite Markov matrix, then the multiplicity of the eigenvalue* 1 *is equal to the number of recurrent classes associated with* $\mathbf{P}$.

Proof. We have seen above that if C_1 is a recurrent class of states, then there exists a left eigenvector $\mathbf{x}^{(1)} > 0$ for the eigenvalue 1 such that $x_i^{(1)} = 0$ if $i \notin C_1$. Similarly, with each recurrent class $C_2, C_3, \ldots$ there is associated a positive eigenvector $\mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \ldots$, with eigenvalue 1 such that $x_i^{(h)} = 0$ if $i \notin C_h$. Since distinct classes are disjoint, it is clear that $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ are linearly independent vectors, and so the multiplicity of the eigenvalue 1 is at least the number of distinct recurrent classes. To prove the reverse inequality suppose that $\mathbf{xP} = \mathbf{x}$. Then $\mathbf{xP}^m = \mathbf{x}$ for $m = 1, 2, \ldots$, i.e.,

$$\sum_{i=1}^{n} x_i P_{ij}^m = x_j, \qquad j = 1, \ldots, n, \quad m = 1, 2, \ldots.$$

But if j is a transient state, we know that $\lim_{m\to\infty} P_{ij}^{(m)} = 0$ for all i. It follows that $x_j = 0$ for every transient state j, and so we can write

$$\sum_{h=1}^{r} \sum_{i \in C_h} x_i P_{ij} = x_j, \qquad j \in \bigcup_{h=1}^{r} C_h,$$

where $C_1, \ldots, C_r$ are the recurrent classes. Also $P_{ij} = 0$ if i and j are in distinct recurrent classes; therefore, we have

$$\sum_{i \in C_h} x_i P_{ij} = x_j \qquad \text{for} \quad j \in C_h, \quad h = 1, \ldots, r.$$

If $x_i \neq 0$ for some $i \in C_h$, then by Theorem 1.3 of Chapter 3 there exists a constant a_h such that

$$x_i = a_h x_i^{(h)}, \qquad i \in C_h.$$

Thus

$$\mathbf{x} = \sum_{h=1}^{r} a_h \mathbf{x}^{(h)},$$

from which we see that the $\mathbf{x}^{(h)}$ form a basis for the manifold of left eigenvectors with eigenvalue 1. ■

PROBABILISTIC INTERPRETATION OF EIGENVALUES AND EIGENVECTORS

Let us now consider the manifold of right eigenvectors of $\mathbf{P}$ with eigenvalue 1. It turns out that there is a basis for this manifold which has a very simple probabilistic interpretation. In fact, if $C_1, \ldots, C_r$ are the recurrent classes associated

with $\mathbf{P}$, we define $p_i^{(h)}$ to be the probability that starting from i the state of the system will eventually lie in C_h, i.e.,

$$p_i^{(h)} = \Pr\{X_n \in C_h \text{ for some } n = 1, 2, \ldots \mid X_0 = i\}.$$

Clearly

$$p_i^{(h)} = \begin{cases} 1 & \text{for } i \in C_h, \\ 0 & \text{for } i \in C_j, \quad j \neq h, \end{cases} \tag{2.3}$$

since it is not possible to leave a recurrent class. If we define $\mathbf{p}^{(h)} = (p_1^{(h)}, \ldots, p_n^{(h)})$, $h = 1, \ldots, r$, the preceding equations show at once that the vectors $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(r)}$ are linearly independent. Furthermore, the $p_i^{(h)}$ satisfy the equations

$$p_i^{(h)} = \sum_{j=1}^{n} P_{ij} p_j^{(h)}, \qquad i = 1, \ldots, n, \quad h = 1, \ldots, r,$$

[Eq. (3.4) of Chapter 3], which shows that $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(r)}$ are right eigenvectors of $\mathbf{P}$ with eigenvalue 1. As the $\mathbf{p}^{(i)}$ are linearly independent and their number r is the multiplicity of the eigenvalue 1, they form a basis for the right eigenmanifold corresponding to the eigenvalue 1. Finally, we observe by direct evaluation with the aid of (2.3) that

$$(\mathbf{x}^{(i)}, \mathbf{p}^{(j)}) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

since the only nonzero components of $\mathbf{x}^{(i)}$ are those whose indices are in C_i, and their sum is just 1.

Let us assume now that $\mathbf{P}$ has a spectral representation, and that the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are labeled so that $1 = \lambda_1 = \cdots = \lambda_r \geq |\lambda_{r+1}| \geq |\lambda_{r+2}| \geq \cdots$ and $\lambda_{r+1} \neq 1$. Then we can take $\boldsymbol{\phi}^{(1)} = \mathbf{p}^{(1)}, \ldots, \boldsymbol{\phi}^{(r)} = \mathbf{p}^{(r)}$ and $\boldsymbol{\psi}^{(1)} = \mathbf{x}^{(1)}, \ldots, \boldsymbol{\psi}^{(r)} = \mathbf{x}^{(r)}$ (see Appendix to *A First Course*). From

$$\mathbf{P}^m = \boldsymbol{\Phi} \boldsymbol{\Lambda}^m \boldsymbol{\Psi}$$

we obtain

$$P_{ij}^m = \sum_{h=1}^{n} \varphi_{hi} \lambda_h^m \bar{\psi}_{hj} = \varphi_{1i} \bar{\psi}_{1j} + \cdots + \varphi_{ri} \bar{\psi}_{rj} + \sum_{h=r+1}^{n} \varphi_{hi} \lambda_h^m \bar{\psi}_{hj}.$$

Suppose that $\mathbf{P}$ has no eigenvalue, different from 1, whose modulus equals 1; then $|\lambda_h| < 1$, $h = r + 1, \ldots, n$, and as $m \to \infty$,

$$\sum_{h=r+1}^{n} \varphi_{hi} \lambda_h^m \bar{\psi}_{hj} \to 0$$

and the rate of convergence is of the order at least $|\lambda_{r+1}|^m$. We shall see shortly that $|\lambda_h| < 1$, $h = r + 1, \ldots, n$, if and only if $\mathbf{P}$ has no periodic recurrent classes (Theorem 3.1 below). Assuming that $\mathbf{P}$ has no periodic recurrent classes and

recalling that $x_j^{(h)} = \psi_{hj}$, $h = 1, \ldots, r$, $j = 1, \ldots, n$, is different from zero if and only if $j \in C_h$, we see that

$$\varphi_{1i}\bar{\psi}_{1j} = \varphi_{2i}\bar{\psi}_{2j} = \cdots = \varphi_{ri}\bar{\psi}_{rj} = 0 \qquad \text{for} \quad j \text{ transient.}$$

Thus, if j is transient, $P_{ij}^m = \sum_{h=r+1}^n \varphi_{hi}\lambda_h^m\bar{\psi}_{hj}$ and this tends to zero at the rate $|\lambda_{r+1}|^m$ as $m \to \infty$. Now if $i, j \in C_h$, then among the first r terms in the expression for P_{ij}^m the only nonvanishing one is $\varphi_{hi}\bar{\psi}_{hj}$; but $\varphi_{hi} = 1$ (recall that $\varphi_{hi} = p_i^{(h)}$) and $\psi_{hj} = \pi_j = \lim_{m\to\infty} P_{ij}^m$. We see generally for all states j that $\pi_j - P_{ij}^m$ goes to zero at least as fast as $|\lambda_{r+1}|^m$ as $m \to \infty$.

Now let us assume that in addition to $|\lambda_{r+1}| < 1$ we have the special situation that $|\lambda_{r+2}| < |\lambda_{r+1}|$. Let, as usual, T denote the set of all transient states, $i, j \in T$; we wish to find the following limit:

$$\lim_{m\to\infty} \Pr\{X_m = j | X_0 = i, X_m \in T\},$$

i.e., the limiting value ($m \to \infty$) of the probability that starting from state i the process is in the transient state j, given that at time m, X_m is in a transient state. We have

$$\Pr\{X_m = j | X_0 = i, X_m \in T\} = \frac{P_{ij}^m}{\sum_{j\in T} P_{ij}^m}.$$

As we have seen before, for j transient $P_{ij}^m = \sum_{h=r+1}^n \varphi_{hi}\lambda_h^m\bar{\psi}_{hj}$.

Since $|\lambda_{r+1}| > |\lambda_{r+2}|$, we readily find that

$$\lim_{m\to\infty} \frac{P_{ij}^m}{\sum_{j\in T} P_{ij}^m} = \frac{\varphi_{r+1,i}\bar{\psi}_{r+1,j}}{\sum_{j\in T} \varphi_{r+1,i}\bar{\psi}_{r+1,j}} = \frac{\bar{\psi}_{r+1,j}}{\sum_{j\in T} \bar{\psi}_{r+1,j}},$$

assuming that the denominator does not vanish. If the denominator does vanish, we have to examine the terms in $\sum_{h=r+1}^n \varphi_{hi}\lambda_h^m\bar{\psi}_{hj}$ containing λ_{r+2} and other eigenvalues whose modulus equals $|\lambda_{r+2}|$, and so forth.

3: Periodic Classes

We wish to give a more complete description of the structure of a periodic chain. The simplest class with period d is clearly one in which there are d states $1, \ldots, d$ and

$$P_{12} = P_{23} = \cdots = P_{d-1,d} = P_{d1} = 1, \qquad \mathbf{P} = \begin{Vmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \end{Vmatrix}.$$

A less trivial example may be formed by replacing the individual states $1, \ldots, d$ by disjoint families $C_1, \ldots, C_d$ of states and defining the P_{ij} in such a way that

$P_{ij} \neq 0$ only if $i \in C_1, j \in C_2$, or $i \in C_2, j \in C_3, \ldots,$ or $i \in C_d, j \in C_1$. The matrix $\mathbf{P}$ then takes the form

$$\mathbf{P} = \begin{Vmatrix} \mathbf{0} & \mathbf{P}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_2 & \cdots \\ \vdots & \vdots & \vdots & \\ \mathbf{P}_d & \mathbf{0} & \mathbf{0} & \cdots \end{Vmatrix}.$$

At the same time, we can define P_{ij} so that every two states communicate. We will now prove that every periodic class is of this form. Let d be the period of the class W and assume that the states are labeled by $1, 2, \ldots, M$. Let C_1 consist of the states of W which can be reached from state 1 in some multiple of d transitions; i.e., $j \in C_1$ if and only if $P_{1j}^{nd} > 0$ for some integer $n > 0$. For each $r = 1, \ldots, d-1$, we define C_{r+1} to consist of those states which can be reached from state 1 in r plus some multiple of d transitions; i.e., $j \in C_{r+1}$ if and only if $P_{1j}^{nd+r} > 0$ for some integer $n \geq 0$.

First we show that if $j \in C_1$ then $P_{j1}^{h} > 0$ implies $h = md$ for some $m > 0$. In fact, since $j \in C_1$ implies that $P_{1j}^{nd} > 0$ for some $n > 0$, it follows that $P_{jj}^{nd+h} \geq P_{j1}^{h} P_{1j}^{nd} > 0$ (cf. Chapter 2, Theorem 3.1), and so by the definition of period $nd + h$ must be divisible by d; hence h is. Next we show that if $i \in C_1, j \in C_{r+1}$ then $P_{ij}^{h} > 0$ implies that $h = nd + r$ for some $n \geq 0$. In fact, let $P_{ji}^{s} > 0$ for some $s > 0$; $P_{i1}^{qd} > 0$ for some $q > 0$; and $P_{1j}^{md+r} > 0$ for some $m \geq 0$. Thus, if $w = s + dq + md + r$ then $P_{11}^{w} \geq P_{1j}^{md+r} P_{ji}^{s} P_{i1}^{qd} > 0$, and w is a multiple of d; therefore $s + r$ also is a multiple of d. But $P_{jj}^{h+s} \geq P_{ji}^{s} P_{ij}^{h} > 0$, so that $h + s$ is divisible by d. Combining these two results, we infer that $h - r$ is divisible by d, and therefore, $h = nd + r$ for some $n \geq 0$.

We leave it to the reader to verify that the above results imply that $C_1, \ldots, C_d$ are disjoint and nonempty, that $\bigcup_{i=1}^{d} C_i = W$, and that $i \in C_r$ requires $P_{ij} = 0$ for every $j \notin C_{r+1}$ where $C_{d+1} = C_1$.

Having thus analyzed the matrix of a periodic class, we can now demonstrate an earlier assertion concerning the occurrence of eigenvalues of modulus 1 of a Markov transition matrix.

Theorem 3.1. *If* $\mathbf{P}$ *is the transition matrix of a finite irreducible periodic Markov chain with period* d, *then the* dth *roots of unity are eigenvalues of* $\mathbf{P}$, *each with multiplicity* 1, *and there are no other eigenvalues of modulus* 1.

Proof. Let $D_1, \ldots, D_d$ be the "moving classes" of the process as established above, i.e., $i \in D_r$ implies $P_{ij} = 0$ for every $j \notin D_{r+1}$. It is no loss of generality to assume that $D_1 = \{1, \ldots, n_1\}$, $D_2 = \{n_1 + 1, \ldots, n_1 + n_2\}, \ldots, D_d = \{M - n_d + 1, \ldots, M\}$. From the definition of the moving class it follows that

$$\mathbf{P}^d = \begin{Vmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_d \end{Vmatrix},$$

where $\mathbf{A}_i$ is an $n_i \times n_i$ Markov matrix. Furthermore, for each i, $\mathbf{A}_i^m \gg \mathbf{0}$ for some integer $m > 0$ (see Problem 5, Chapter 2). Thus, $\mathbf{A}_i$ has a strictly positive left eigenvector $\boldsymbol{\mu}^{(i)}$, with eigenvalue 1, of algebraic multiplicity 1. Owing to the form of $\mathbf{P}^d$, it is clear that, by adjoining an appropriate number of zeros on one or both sides of each $\boldsymbol{\mu}^{(i)}$, we determine linearly independent vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(d)}$ such that

$$\mathbf{x}^{(i)} = \mathbf{x}^{(i)}\mathbf{P}^d, \qquad i = 1, \ldots, d.$$

Let us consider the vectors $\mathbf{y}^{(1)} = \mathbf{x}^{(1)}$, $\mathbf{y}^{(2)} = \mathbf{x}^{(1)}\mathbf{P}, \ldots, \mathbf{y}^{(d)} = \mathbf{x}^{(1)}\mathbf{P}^{d-1}$. Since the only nonzero components of $\mathbf{x}^{(1)}$ are those with indices $1, 2, \ldots, n_1$, and observing that $P_{ij}^{(h)}$ may differ from zero only if the moving class in which i lies agrees with that of j after precisely h steps, we see that the only nonzero components of $\mathbf{y}^{(i)}$ are those with indices $(n_1 + \cdots + n_{i-1} + 1, \ldots, n_1 + \cdots + n_i)$. This implies that the vectors $\mathbf{y}^{(i)}$ $(i = 1, \ldots, d)$ are linearly independent. Furthermore

$$\mathbf{y}^{(i)}\mathbf{P}^d = \mathbf{x}^{(1)}\mathbf{P}^{i-1}\mathbf{P}^d = \mathbf{x}^{(1)}\mathbf{P}^d\mathbf{P}^{i-1} = \mathbf{x}^{(1)}\mathbf{P}^{i-1} = \mathbf{y}^{(i)}.$$

It follows that if we restrict attention to the n_i-dimensional linear space obtained by considering only those components of $\mathbf{y}^{(i)}$ whose indices lie in D_i, we obtain a left eigenvector with eigenvalue 1 for $\mathbf{A}_i$.

Because the eigenvalue 1 has simple multiplicity for $\mathbf{A}_i$, it follows that each $\mathbf{y}^{(i)}$ is a constant multiple of $\mathbf{x}^{(i)}$. Actually, if we normalize each $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(d)}$ by the condition $\sum_{i=1}^{n} x_i^{(h)} = 1$, $h = 1, \ldots, d$, then, in fact, $\mathbf{y}^{(h)} = \mathbf{x}^{(h)}$, $h = 1, \ldots, d$. Accordingly, we may write $\mathbf{x}^{(2)} = \mathbf{x}^{(1)}\mathbf{P}$, $\mathbf{x}^{(3)} = \mathbf{x}^{(2)}\mathbf{P}, \ldots, \mathbf{x}^{(1)} = \mathbf{x}^{(d)}\mathbf{P}$.

Let $\omega = e^{2\pi i/d}$. Combining the above equations in the indicated manner, we obtain

$$\begin{aligned}
&(\mathbf{x}^{(1)} + \mathbf{x}^{(2)} + \mathbf{x}^{(3)} + \cdots + \mathbf{x}^{(d)})\mathbf{P} \\
&\quad = \mathbf{x}^{(1)} + \mathbf{x}^{(2)} + \cdots + \mathbf{x}^{(d)}, \\
&(\mathbf{x}^{(1)} + \omega\mathbf{x}^{(2)} + \omega^2\mathbf{x}^{(3)} + \cdots + \omega^{d-1}\mathbf{x}^{(d)})\mathbf{P} \\
&\quad = \omega^{-1}(\mathbf{x}^{(1)} + \omega\mathbf{x}^{(2)} + \cdots + \omega^{d-1}\mathbf{x}^{(d)}), \\
&(\mathbf{x}^{(1)} + \omega^2\mathbf{x}^{(2)} + \omega^4\mathbf{x}^{(3)} + \cdots + \omega^{2(d-1)}\mathbf{x}^{(d)})\mathbf{P} \\
&\quad = \omega^{-2}(\mathbf{x}^{(1)} + \omega^2\mathbf{x}^{(2)} + \cdots + \omega^{2(d-1)}\mathbf{x}^{(d)}), \\
&\quad \vdots \\
&(\mathbf{x}^{(1)} + \omega^{(d-1)}\mathbf{x}^{(2)} + \omega^{2(d-1)}\mathbf{x}^{(3)} + \cdots + \omega^{(d-1)^2}\mathbf{x}^{(d)})\mathbf{P} \\
&\quad = \omega^{-(d-1)}(\mathbf{x}^{(1)} + \omega^{(d-1)}\mathbf{x}^{(2)} + \cdots + \omega^{(d-1)^2}\mathbf{x}^{(d)}).
\end{aligned}$$

The linear independence of the $\mathbf{x}^{(i)}$ ensures that none of the vectors appearing above are zero. These relations exhibit the property that the dth roots of unity are all eigenvalues of $\mathbf{P}$.

Suppose next that $\mathbf{xP} = \lambda\mathbf{x}$ for some nonzero $\mathbf{x}$. Then $\mathbf{xP}^d = \lambda^d\mathbf{x}$. Looking at the contracted vectors $\mathbf{z}^{(1)} = (x_1, \ldots, x_{n_1})$, $\mathbf{z}^{(2)} = (x_{n_1+1}, \ldots, x_{n_1+n_2})$, $\ldots$, $\mathbf{z}^{(d)} = (x_{M-n_d+1}, \ldots, x_M)$, we see that

$$\mathbf{z}^{(i)}\mathbf{A}_i = \lambda^d\mathbf{z}^{(i)}, \qquad i = 1, \ldots, d.$$

Since at least one of the $\mathbf{z}^{(i)}$ is nonzero, and for each $\mathbf{A}_i$ there is an m such that $\mathbf{A}_i^m \gg 0$, either $\lambda^d = 1$ or $|\lambda^d| < 1$. If $\lambda^d = 1$, then there are constants $c_1, \ldots, c_d$ such that

$$\mathbf{z}^{(i)} = c_i\mathbf{x}^{(i)}, \qquad i = 1, \ldots, d,$$

and so we see that $\mathbf{x} = c_1\mathbf{x}^{(1)} + \cdots + c_d\mathbf{x}^{(d)}$.

Now

$$\lambda\mathbf{x} = \mathbf{xP} = c_1\mathbf{x}^{(2)} + c_2\mathbf{x}^{(3)} + \cdots + c_d\mathbf{x}^{(1)}$$

or

$$\lambda c_1\mathbf{x}^{(1)} + \cdots + \lambda c_d\mathbf{x}^{(d)} = c_d\mathbf{x}^{(1)} + c_1\mathbf{x}^{(2)} + \cdots + c_{d-1}\mathbf{x}^{(d)}.$$

Since the $\mathbf{x}^{(i)}$ are linearly independent, we have

$$\lambda c_1 = c_d, \qquad \lambda c_2 = c_1, \qquad \ldots, \qquad \lambda c_d = c_{d-1},$$

or

$$c_{d-1} = \lambda c_d = (\lambda^{-1})^{d-1}c_d, \qquad c_{d-2} = \lambda^2 c_d = (\lambda^{-1})^{d-2}c_d, \qquad \ldots,$$
$$c_1 = \lambda^{d-1}c_d = \lambda^{-1}c_d$$

since $\lambda^d = 1$, and this means that $\mathbf{x}$ is plainly a constant multiple of one of the eigenvectors of $\mathbf{P}$ already constructed. ■

The case of an arbitrary Markov matrix $\mathbf{P}$ follows easily from the preceding theorem. We have

***Theorem* 3.2.** *If* $\mathbf{P}$ *is a finite Markov matrix, then any eigenvalue of* $\mathbf{P}$ *of modulus* 1 *is a root of unity. The* dth *roots of unity are eigenvalues of* $\mathbf{P}$ *if and only if* $\mathbf{P}$ *has a recurrent class with period* d. *The multiplicity of each* dth *root of unity is just the number of recurrent classes of period* d.

The proof is essentially identical with that of Theorem 3.1. Since $\lambda\mathbf{x} = \mathbf{xP}$ implies

$$\lambda^m\mathbf{x} = \mathbf{xP}^m$$

or

$$\lambda^m x_j = \sum_{i=1}^{n} x_i P_{ij}^m,$$

then, letting $m \to \infty$, we see that $x_j = 0$ if j is transient. We may therefore restrict attention to the recurrent states, and the theorem immediately reduces to the case considered in the previous theorem.

4: Special Computational Methods in Markov Chains

Let $\mathbf{P}$ be the transition probability matrix of a random walk on the nonnegative integers with probability $\frac{1}{2}$ of going to each of the two neighboring states from state k ($k \geq 1$) and with a reflecting barrier at the origin; that is,

$$\mathbf{P} = \begin{Vmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix}.$$

To obtain the probability of reaching state l from state k in n steps we could multiply matrix $\mathbf{P}$ by itself n times and seek out the element P_{kl}^n in the kth row and lth column of the matrix $\mathbf{P}^n$. This method, however, is very cumbersome and lengthy in practice.

A second approach is to attempt to generalize the method of eigenvalues and eigenvectors as developed in Section 2. In the case of infinite matrices this cannot always be done. However, for matrices of the same form as $\mathbf{P}$ above or, more generally, transition probability matrices corresponding to random walks, there is available an infinite analog to the representation formula (1.1).

We proceed to obtain P_{kl}^n in a manner which will illustrate a general method applicable to arbitrary random walks.

Adding the two trigonometric identities

$$\cos(\alpha \pm \beta) = \cos\alpha \cos\beta \mp \sin\alpha \sin\beta$$

leads to the identity

$$\cos\alpha \cos\beta = \tfrac{1}{2}\cos(\alpha + \beta) + \tfrac{1}{2}\cos(\alpha - \beta). \qquad (*)$$

Let $\alpha = \theta$ and $\beta = k\theta$ ($k = 1, 2, \ldots$). We get

$$\cos\theta \cos k\theta = \tfrac{1}{2}\cos(k+1)\theta + \tfrac{1}{2}\cos(k-1)\theta. \qquad (4.1)$$

Since the elements in the kth row of matrix $\mathbf{P}$ are

$$P_{k,0} = P_{k,1} = 0, \ldots, \quad P_{k,k-2} = 0, \quad P_{k,k-1} = \tfrac{1}{2}, \quad P_{k,k} = 0,$$
$$P_{k,k+1} = \tfrac{1}{2}, \quad P_{k,k+2} = 0, \ldots, \qquad k = 2, 3, \ldots,$$
$$P_{1,0} = \tfrac{1}{2}, \quad P_{1,1} = 0, \quad P_{1,2} = \tfrac{1}{2}, \quad P_{1,3} = 0, \ldots,$$
$$P_{0,0} = 0, \quad P_{0,1} = 1, \quad P_{0,2} = 0, \ldots,$$