An Introduction to Stochastic Processes with Applications to Biology

Linda J. S. Allen
Department of Mathematics and Statistics
Texas Tech University
Contents

Preface xi

1 Review of Probability Theory and an Introduction to Stochastic Processes 1
  1.1 Introduction .............................................. 1
  1.2 Brief Review of Probability Theory .................. 3
  1.3 Generating Functions ................................... 18
  1.4 Central Limit Theorem .................................. 22
  1.5 Introduction to Stochastic Processes ............... 24
  1.6 An Introductory Example: A Simple Birth Process .. 27
  1.7 Exercises for Chapter 1 ................................ 32
  1.8 References for Chapter 1 .............................. 35
  1.9 Appendix for Chapter 1 ............................... 37
    1.9.1 MATLAB and FORTRAN Programs ................. 37
    1.9.2 Interevent Time .................................. 38

2 Discrete Time Markov Chains 41
  2.1 Introduction ........................................... 41
  2.2 Definitions and Notation ............................. 42
  2.3 Classification of States .............................. 45
  2.4 First Passage Time ................................... 51
  2.5 Basic Theorems for Markov Chains .................. 56
  2.6 Stationary Probability Distribution ............... 62
  2.7 Finite Markov Chains ................................ 65
    2.7.1 Mean Recurrence Time and Mean First Passage Time 69
  2.8 The n-Step Transition Matrix ......................... 71
  2.9 An Example: Genetics Inbreeding Problem ........... 75
  2.10 Unrestricted Random Walks in Two and Three Dimensions 77
    2.10.1 Two Dimensions .................................. 77
    2.10.2 Three Dimensions ................................ 78
  2.11 Exercises for Chapter 2 ............................. 80
  2.12 References for Chapter 2 ............................ 86
  2.13 Appendix for Chapter 2 ............................. 88
    2.13.1 Power of a Matrix ................................ 88
    2.13.2 Genetics Inbreeding Problem ...................... 89
3 Biological Applications of Discrete Time Markov Chains 91
  3.1 Introduction ........................................... 91
  3.2 Restricted Random Walk Models ......................... 92
  3.3 Gambler's Ruin Problem ................................ 93
    3.3.1 Probability of Absorption ......................... 95
    3.3.2 Expected Time until Absorption .................. 98
    3.3.3 Probability Distribution for Absorption ........ 101
  3.4 Gambler's Ruin Problem on a Semi-Infinite Domain ...... 104
  3.5 General Birth and Death Process ......................... 106
    3.5.1 Expected Time to Extinction ...................... 107
  3.6 Logistic Growth Process ................................ 109
  3.7 Quasistationary Probability Distribution ................ 112
  3.8 SIS Epidemic Model .................................... 115
    3.8.1 Deterministic SIS Epidemic Model ................ 117
    3.8.2 Stochastic SIS Epidemic Model ................... 118
  3.9 Chain Binomial Epidemic Models ......................... 121
    3.9.1 Greenwood Model .................................. 122
    3.9.2 Reed-Frost Model .................................. 124
    3.9.3 Duration and Size of the Epidemic ............... 125
  3.10 Exercises for Chapter 3 ................................ 127
  3.11 References for Chapter 3 .............................. 133
  3.12 Appendix for Chapter 3 ................................ 135
    3.12.1 MATLAB Programs ................................ 135
    3.12.2 Maple Program ................................... 137

4 Discrete Time Branching Processes 139
  4.1 Introduction ........................................... 139
  4.2 Definitions and Notation ................................ 140
  4.3 Probability Generating Function of $X_n$ ............... 143
  4.4 Probability of Population Extinction ................... 145
  4.5 Mean and Variance of $X_n$ ............................ 151
  4.6 Multitype Branching Processes ......................... 155
  4.7 An Example: Age-Structured Model ....................... 159
  4.8 Exercises for Chapter 4 ................................ 164
  4.9 References for Chapter 4 ................................ 169

5 Continuous Time Markov Chains 171
  5.1 Introduction ........................................... 171
  5.2 Definitions and Notation ................................ 172
  5.3 The Poisson Process ................................... 174
  5.4 Generator Matrix $Q$ .................................... 178
  5.5 Embedded Markov Chain and Classification of States ... 181
  5.6 Kolmogorov Differential Equations ....................... 186
  5.7 Finite Markov Chains ................................... 189
  5.8 Generating Function Technique .......................... 194
5.9 Interevent Time and Stochastic Realizations 197
5.10 Review of Method of Characteristics 203
5.11 Exercises for Chapter 5 204
5.12 References for Chapter 5 208
5.13 Appendix for Chapter 5 209
5.13.1 MATLAB Program 209

6 Continuous Time Birth and Death Chains 211
6.1 Introduction 211
6.2 General Birth and Death Process 212
6.3 Stationary Probability Distribution 215
6.4 Simple Birth and Death Processes 218
6.4.1 Simple Birth Process 219
6.4.2 Simple Death Process 222
6.4.3 Simple Birth and Death Process 224
6.4.4 Simple Birth and Death Process with Immigration 228
6.5 Queueing Processes 232
6.6 Probability of Population Extinction 236
6.7 Expected Time to Extinction and First Passage Time 237
6.8 Logistic Growth Process 242
6.9 Quasistationary Probability Distribution 247
6.10 An Explosive Birth Process 249
6.11 Nonhomogeneous Birth and Death Process 252
6.12 Exercises for Chapter 6 254
6.13 References for Chapter 6 261
6.14 Appendix for Chapter 6 263
6.14.2 Proofs of Theorems 6.2 and 6.3 265
6.14.3 Comparison Theorem 268

7 Epidemic, Competition, Predation and Population Genetics Processes 269
7.1 Introduction 269
7.2 Continuous Time Branching Processes 270
7.3 SI and SIS Epidemic Processes 275
7.3.1 Stochastic SI Epidemic Model 277
7.3.2 Stochastic SIS Epidemic Model 280
7.4 Multivariate Processes 281
7.5 SIR Epidemic Process 284
7.5.1 Stochastic SIR Epidemic Model 286
7.5.2 Final Size of the Epidemic 288
7.5.3 Expected Duration of an SIR Epidemic 291
7.6 Competition Processes 293
7.6.1 Stochastic Competition Model 295
## Contents

7.7 Predator-Prey Processes ................................................. 297
  7.7.1 Stochastic Predator-Prey Model ................................... 298

7.8 Other Population Processes ............................................ 300
  7.8.1 SEIR Epidemic Model ............................................. 300
  7.8.2 Spatial Predator-Prey Model ...................................... 302
  7.8.3 Population Genetics Model ....................................... 304

7.9 Exercises for Chapter 7 ................................................ 308

7.10 References for Chapter 7 .............................................. 313

7.11 Appendix for Chapter 7 ............................................... 316
  7.11.1 MATLAB Programs .............................................. 316

8 Diffusion Processes and Stochastic Differential Equations 321
  8.1 Introduction ....................................................... 321
  8.2 Definitions and Notation ........................................... 322
  8.3 Random Walk and Brownian Motion ................................ 324
  8.4 Diffusion Process .................................................. 327
  8.5 Kolmogorov Differential Equations ................................ 328
  8.6 Wiener Process ..................................................... 333
  8.7 Ito Stochastic Integral ............................................. 335
  8.8 Ito Stochastic Differential Equation ................................ 341
  8.9 Numerical Methods for Solving SDEs ............................... 348
  8.10 Ito SDEs for Interacting Populations .............................. 351
  8.11 Epidemic, Competition, and Predation Processes ............... 357
    8.11.1 Competition Model ........................................... 357
    8.11.2 Predator-Prey Model ......................................... 358
    8.11.3 SIR Epidemic Model ......................................... 360
  8.12 Population Genetics Process ...................................... 362
  8.13 Expected Time to Extinction and First Passage Time ........... 365
  8.14 Exercises for Chapter 8 .......................................... 367
  8.15 References for Chapter 8 ......................................... 373
  8.16 Appendix for Chapter 8 ........................................... 376
    8.16.1 Derivation of Kolmogorov Equations ........................ 376
    8.16.2 MATLAB Programs ........................................... 377

Index 381
Chapter 2

Discrete Time Markov Chains

2.1 Introduction

In this chapter, discrete time Markov chains are introduced. Both time and state space are discrete. The theory and application of Markov chains is probably one of the most well-developed theories of stochastic processes. A classic textbook on finite Markov chains (where the state space is finite) is the textbook by Kemeny and Snell (1960). Some additional references on the theory and numerical methods for discrete time Markov chains include *A First Course in Stochastic Processes*, by Karlin and Taylor (1975); *An Introduction to Stochastic Modeling*, by Taylor and Karlin (1998); *Classical and Spatial Stochastic Processes*, by Schinazi (1999); *Markov Chains*, by Norris (1997); and *Introduction to the Numerical Solution of Markov Chains*, by Stewart (1994).

We introduce some basic notation and theory for discrete time Markov chains in this chapter. A discrete time chain can be classified as irreducible or reducible, periodic or aperiodic, and recurrent or transient. These classifications help in determining the behavior of the Markov chain. Basic theorems concerning the asymptotic behavior are stated that apply to particular types of Markov chains. For example, it is shown that a stationary limiting distribution exists for an aperiodic, irreducible, and recurrent Markov chain. A stationary distribution is analogous to a stable equilibrium in a deterministic model. However, in a stochastic model, the “equilibrium” is defined by a probability distribution, known as the stationary probability distribution. Some well-known examples of discrete time Markov chains are discussed in this chapter, including the random walk model in one, two, and three dimensions. In addition, a problem related to genetics inbreeding is discussed.
2.2 Definitions and Notation

Consider a discrete time stochastic process, \( \{X_n\} \), \( n = 0, 1, 2, \ldots \), where the random variable \( X_n \) is a discrete random variable defined on a finite or countably infinite state space. For convenience, we denote the state space as \( \{1, 2, \ldots \} \). However, the set could be finite and could include nonpositive integer values. Also, the variable \( n \) is used instead of \( t \) to denote an element of the index set; this notation is used frequently in discrete time processes. The index set is defined as \( \{0, 1, 2, \ldots \} \), since it often represents the progression of time, which is also the reason for the terminology, discrete time processes. Therefore, the index \( n \) shall be referred to as "time" \( n \).

A Markov stochastic process is a stochastic process in which the future behavior of the system depends only on the present and not on its past history. More formally,

\[ \text{Definition 2.1.} \quad \text{A discrete time stochastic process} \ \{X_n\}_{n=0}^\infty \ \text{is said to have the Markov property if} \]

\[ \text{Prob}\{X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\} = \text{Prob}\{X_n = i_n | X_{n-1} = i_{n-1}\}, \]

where the values of \( i_k \in \{1, 2, \ldots \} \) for \( k = 0, 1, 2, \ldots, n \). The stochastic process is then called a Markov chain or, more specifically, a discrete time Markov chain. It is called a finite state Markov chain or a finite Markov chain if the state space is finite.

The stochastic process is referred to as a chain when the state space is discrete. The name Markov refers to Andrei A. Markov, a Russian probabilist (1856–1922), whose work in Markov chains contributed much to the theory of stochastic processes.

The notation \( \text{Prob} \) is used to denote the induced probability measure, \( \text{Prob}\{\cdot\} = P_{X_n}(\cdot) \), because \( P \) will refer to the transition matrix that is defined below. Denote the probability mass function associated with the random variable \( X_n \) by \( \{p_i(n)\}_{i=0}^{\infty} \), where

\[ p_i(n) = \text{Prob}\{X_n = i\}. \quad (2.1) \]

The state of the process at time \( n \), \( X_n \), is related to the process at time \( n + 1 \) through what is known as the transition probabilities. If the process is in state \( i \) at time \( n \), at the next time step \( n + 1 \), it will either stay in state \( i \) or move or transfer to another state \( j \). The probabilities for these changes in state are defined by the one-step transition probabilities.

\[ \text{Definition 2.2.} \quad \text{The one-step transition probability, denoted as} \ p_{ji}(n), \ \text{is defined as the following conditional probability:} \]

\[ p_{ji}(n) = \text{Prob}\{X_{n+1} = j | X_n = i\}. \]

the probability that the process is in state \( j \) at time \( n + 1 \) given that the process was in state \( i \) at the previous time \( n \), for \( i, j = 1, 2, \ldots \).
Definition 2.3. If the transition probabilities $p_{ji}(n)$ in a Markov chain do not depend on time $n$, they are said to be stationary or time homogeneous or simply homogeneous. In this case, we shall use the notation $p_{ji}$. If the transition probabilities are time dependent, $p_{ji}(n)$, then they are said to be nonstationary or nonhomogeneous.

A Markov chain can have either stationary or nonstationary transition probabilities. Unless stated otherwise, it shall be assumed that the transition probabilities of the Markov chain are stationary. For each state, the one-step transition probabilities satisfy

$$\sum_{j=1}^{\infty} p_{ji} = 1, \quad \text{for} \quad i = 1, 2, \ldots \quad \text{and} \quad p_{ji} \geq 0,$$

meaning that, with probability one, the process in any state $i$ must move or transfer to some other state $j$, $j \neq i$ or stay in state $i$ at the next time interval. This identity also states for a fixed $i$, $\{p_{ji}\}$ is a probability distribution.

The one-step transition probabilities can be expressed in matrix form, which is referred to as the transition matrix.

Definition 2.4. The transition matrix of the discrete time Markov chain \( \{X_n\}_{n=0}^{\infty} \) with state space \( \{1, 2, \ldots\} \) and one-step transition probabilities, \( \{p_{ij}\}_{i,j=1}^{\infty} \), is denoted as \( P = (p_{ij}) \), where

$$P = \begin{pmatrix}
p_{11} & p_{12} & p_{13} & \cdots \\
p_{21} & p_{22} & p_{23} & \cdots \\
p_{31} & p_{32} & p_{33} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}.$$ 

If the set of states is finite, \( \{1, 2, \ldots, N\} \), then \( P \) is an \( N \times N \) matrix. Note that the column elements sum to one since \( \sum_{j=1}^{N} p_{ji} = 1 \). A nonnegative matrix with the property that each column sum equals one is called a stochastic matrix. The transition matrix \( P \) is a stochastic matrix. It is left as an Exercise to show that if \( P \) is a stochastic matrix, then \( P^n \) is a stochastic matrix, for \( n \) any positive integer. If the row sums also equal one, then the matrix is called doubly stochastic.

The notation used here differs from that used in some textbooks in two respects. First, the transition matrix is sometimes defined as the transpose of \( P \), \( P^T \). Then the definition of a stochastic matrix is defined as a nonnegative matrix whose row sums equal one (rather than column sums equal one) (Kemeny and Snell, 1960; Norris, 1997; Stewart, 1994). Second, generally, the one-step transition probability \( p_{ij} \) is defined as the probability of a transition from state \( i \) to state \( j \) rather than a transition from \( j \) to \( i \) as in our notation (Bailey, 1990; Karlin and Taylor, 1975; Kemeny and Snell,
We prefer this notation because it closely resembles the notation used in deterministic models and will allow us to more easily relate deterministic models to stochastic models (see Tuljapurkar, 1997). For example, suppose \( Y_{n+1} = AY_n \) represents the dynamics of a deterministic system that changes over time. Matrix \( A = (a_{ij}) \) and \( Y = (y_1, y_2, \ldots, y_k)^T \). The term \( a_{ij} \) in matrix \( A \) represents the effect variable \( y_j \) has on \( y_i \), \( j \rightarrow i \) during the time interval \([n, n + 1]\). In addition, using our notation, the element \( p_{ij} \) is in the \( i \)th row and \( j \)th column of the transition matrix \( P \), which is the standard notation used to define matrix elements. As an aid in setting up and understanding how the elements of \( P \) are related, note that the nonzero elements in the \( i \)th row of \( P \) represent all those states \( j \) (column \( j \)) that can transfer into state \( i \) in one time step. Next, we define the \( n \)-step transition probabilities.

**Definition 2.5.** The \( n \)-step transition probability, denoted \( p_{ji}^{(n)} \), is the probability of moving or transferring from state \( i \) to state \( j \) in \( n \) time steps,

\[
p_{ji}^{(n)} = \text{Prob}\{X_n = j | X_0 = i\}.
\]

The \( n \)-step transition matrix is denoted as \( P^{(n)} = (p_{ji}^{(n)}) \). For the cases \( n = 0 \) and \( n = 1 \), \( P^{(0)} = I \) and \( P^{(1)} = P \) and

\[
p_{ji}^{(0)} = \delta_{ji} = \begin{cases} 
1, & j = i, \\
0, & j \neq i,
\end{cases}
\]

where \( \delta_{ji} \) represents the Kronecker delta symbol. Then \( P^{(1)} = P \) and \( P^{(0)} = I \), where \( I \) is the identity matrix.

Relationships exist between the \( n \)-step transition probabilities and \( s \)-step and \( (n-s) \)-step transition probabilities. These relationships are known as the Chapman-Kolmogorov equations:

\[
p_{ji}^{(n)} = \sum_{k=1}^{\infty} p_{jk}^{(n-s)} p_{ki}^{(s)}, \quad 0 < s < n.
\]

Verification of the Chapman-Kolmogorov equations can be shown as follows (Stewart, 1994):

\[
p_{ji}^{(n)} = \text{Prob}\{X_n = j | X_0 = i\},
\]

\[
= \sum_{k=1}^{\infty} \text{Prob}\{X_n = j, X_s = k | X_0 = i\}, \quad (2.2)
\]

\[
= \sum_{k=1}^{\infty} \text{Prob}\{X_n = j | X_s = k, X_0 = i\} \text{Prob}\{X_s = k | X_0 = i\}, \quad (2.3)
\]
where equations (2.2)–(2.5) hold for $0 < s < n$. The relationship (2.3) follows from conditional probabilities (see Exercise 2). The relationship (2.4) follows from the Markov property. The preceding identity written in terms of matrix notation yields

$$P^{(n)} = P^{(n-s)} P^{(s)}.$$  

However, because $P^{(1)} = P$, it follows from the Chapman-Kolmogorov equations (2.6) that $P^{(2)} = P^2$ and, in general, $P^{(n)} = P^n$. The $n$-step transition matrix $P^{(n)}$ is just the $n$th power of $P$. The elements of $P^n$ are the $n$-step transition probabilities, $p_{ij}^{(n)}$. Be careful not to confuse the notation $p_{ij}^{(n)}$ with $p_{ij}^{(n)}$, $p_{ij}^{(n)} 
eq p_{ij}^{n}$. The notation $p_{ij}^{n}$ is the $n$th power of the element $p_{ij}$, whereas $p_{ij}^{(n)}$ is the $ij$ element in the $n$th power of $P$.

Let $p(n)$ denote the vector form of the probability mass function associated with $X_n$; that is, $p(n) = (p_1(n), p_2(n), \ldots)^T$, where $p_i(n)$ is defined in (2.1) and the states are arranged in increasing order in the column vector $p(n)$. The probabilities satisfy

$$\sum_{i=1}^{\infty} p_i(n) = 1.$$  

Given the probability distribution associated with $X_n$, the probability distribution associated with $X_{n+1}$ can be found by multiplying the transition matrix $P$ by $p(n)$; that is,

$$p_i(n + 1) = \sum_{j=1}^{\infty} p_{ij} p_j(n)$$

or

$$p(n + 1) = P p(n).$$

In general,

$$p(n + m) = P^{n+m} p(0) = P^n (P^m p(0)) = P^n p(m).$$

### 2.3 Classification of States

Relationships between the states of a Markov chain lead to a classification scheme for the states and ultimately classification for Markov chains.
Definition 2.6. The state \( j \) can be reached from the state \( i \) (or state \( j \)

is accessible from state \( i \)) if there is a nonzero probability, \( p_{ji}^{(n)} > 0 \), for

some \( n \geq 0 \). This relationship is denoted as \( i \rightarrow j \). If \( i \) can be reached from

\( j \), \( j \rightarrow i \), and if \( j \) can be reached from \( i \), \( i \rightarrow j \), then \( i \) and \( j \) are said
to communicate, or to be in the same class, denoted \( i \leftrightarrow j \); that is, there

exists nonnegative integers \( n \) and \( n' \) such that

\[
p_{ji}^{(n)} > 0 \quad \text{and} \quad p_{ij}^{(n')} > 0.
\]

The relation \( i \rightarrow j \) can be represented in graph theory as a directed

graph (see Figure 2.1).

The relation \( i \leftrightarrow j \) is an equivalence relation on the state space \( \{1, 2, \ldots \} \).
The relation satisfies the following three properties (Karlin and Taylor, 1975):

(1) reflexivity: \( i \leftrightarrow i \), because \( p_{ii}^{(0)} = 1 \). Beginning in state \( i \), the system

stays in state \( i \) if there is no time change.

(2) symmetry: \( i \leftrightarrow j \) implies \( j \leftrightarrow i \) follows from the definition.

(3) transitivity: \( i \leftrightarrow j \), \( j \leftrightarrow k \) implies \( i \leftrightarrow k \). To verify this last property,

note that the first two properties imply there exist nonnegative integers \( n \) and \( m \) such that

\( p_{ji}^{(n)} > 0 \) and \( p_{kj}^{(m)} > 0 \). Thus,

\[
p_{ki}^{(n+m)} = \text{Prob}\{X_{n+m} = k|X_0 = i\},
\]

\[
\geq \text{Prob}\{X_{n+m} = k, X_n = j|X_0 = i\},
\]

\[
= \text{Prob}\{X_{n+m} = k|X_n = j\} \text{Prob}\{X_n = j|X_0 = i\},
\]  (2.7)

\[
= p_{kj}^{(m)} p_{ji}^{(n)},
\]

where probability (2.7) follows from conditional probabilities and the

Markov property. Thus, \( p_{ki}^{(n+m)} > 0 \) and \( i \rightarrow k \). Similarly, it can be

shown that \( p_{ik}^{(n+m)} > 0 \), which implies \( k \rightarrow i \).

The equivalence relation on the states of the Markov chain define a set

of equivalence classes. These equivalence classes are known as classes of the

Markov chain.

Definition 2.7. The set of equivalence classes in a discrete time Markov

chain are called the communication classes or, more simply, the classes of

the Markov chain.
If every state in the Markov chain can be reached from every other state, then there is only one communication class (all the states are in the same class).

**Definition 2.8.** If there is only one communication class, then the Markov chain is said to be **irreducible**, but if there is more than one communication class, then the Markov chain is said to be **reducible**.

A communication class may have the additional property that it is **closed**.

**Definition 2.9.** A set of states \( C \) is said to be **closed** if it is impossible to reach any state outside of \( C \) from any state in \( C \) by one-step transitions; \( p_{ji} = 0 \) if \( i \in C \) and \( j \notin C \).

A sufficient condition that shows that a Markov chain is irreducible is the existence of a positive integer \( n \) such that \( p_{ji}^{(n)} > 0 \) for all \( i \) and \( j \); that is, every element in \( P^n \) is positive, \( P^n > 0 \), for some positive integer \( n \). For a finite Markov chain, irreducibility can be checked from the directed graph for that chain. A finite Markov chain with states \( \{1, 2, \ldots, N\} \) is irreducible if there is a directed path from \( i \) to \( j \) for every \( i, j \in \{1, 2, \ldots, N\} \).

The definitions of **irreducible** and **reducible** apply more generally to \( N \times N \) matrices, \( A = (a_{ij}) \). A **directed graph** or **digraph** with \( N \) nodes can be constructed from an \( N \times N \) matrix. There is a single directed path from node \( i \) to node \( j \) if \( a_{ij} \neq 0 \). Then node \( j \) can be reached from node \( i \) in one step. A more general signed digraph can be constructed, where the sign of \( a_{ji} \) is associated with each directed path. Node \( j \) can be reached from node \( i \) in \( n \) steps if \( a_{ji}^{(n)} \neq 0 \), where \( a_{ji}^{(n)} \) is the element in the \( j \)th row and \( i \)th column of \( A^n \). A directed graph with \( N \) nodes constructed from a matrix \( A \) is said to be **strongly connected** if there exists a series of directed paths from \( i \) to \( j \) for every \( i, j \in \{1, 2, \ldots, N\} \) \((i \leftrightarrow j)\). Then a directed graph is strongly connected if it is possible to start from any node \( i \) and reach any other node \( j \) in a finite number of steps. Matrix irreducibility is defined as a strongly connected digraph (Ortega, 1987).

**Definition 2.10.** Matrix \( A \) is said to be **irreducible** if and only if its directed graph is strongly connected. Alternately, matrix \( A \) is said to be **reducible** if and only if its directed graph is not strongly connected.

**Example 2.1** A discrete time Markov chain with four states \( \{1, 2, 3, 4\} \) has the following transition matrix:

\[
P = \begin{pmatrix}
0 & 0 & p_{13} & 0 \\
p_{21} & 0 & p_{23} & p_{24} \\
0 & 0 & 0 & 0 \\
0 & p_{42} & 0 & 0
\end{pmatrix},
\]

where \( p_{ij} \) denotes a positive element. Then it is easy to see that \( 4 \leftrightarrow 2 \leftrightarrow 1 \leftarrow 3 \) and \( 4 \leftrightarrow 2 \leftarrow 3 \) (see Figure 2.2).
Since it is impossible to return to states 1 or 3 after having left them, each of these states forms a single communication class, \( \{1\} \), \( \{3\} \). The set \( \{2, 4\} \) is a third communication class. The Markov chain is reducible. In addition, the set \( \{2, 4\} \) is closed, but the sets \( \{1\} \) and \( \{3\} \) are not closed. If one of the elements, either \( p_{12} \) or \( p_{14} \), is positive, then the communication classes consist of \( \{1, 2, 4\} \) and \( \{3\} \). If one of the elements, either \( p_{32} \) or \( p_{34} \), is positive, then there is a single communication class \( \{1, 2, 3, 4\} \); the directed graph is strongly connected and matrix \( P \) is irreducible. Also, the discrete time Markov chain is irreducible.

The following example illustrates the classical gambler's ruin problem.

**Example 2.2** The state space is the set \( \{0, 1, 2, \ldots, N\} \). The states represent the amount of money of one of the players (gambler). The gambler bets \$1 per game and either wins or loses each game. The gambler is ruined if he/she reaches state 0. The probability of winning (moving to the right) is \( p > 0 \) and the probability of losing (moving to the left) is \( q > 0 \), \( p + q = 1 \) (i.e., \( p_{i,i+1} = q \) and \( p_{i+1,i} = p \), \( i = 1, \ldots, N - 1 \)). In addition, \( p_{00} = 1 \) and \( p_{NN} = 1 \), which are referred to as absorbing boundaries. All other elements of the transition matrix are zero. In general, a state \( i \) is called absorbing if \( p_{ii} = 1 \). See the directed graph in Figure 2.3 and the corresponding \( (N + 1) \times (N + 1) \) transition matrix:

\[
P = \begin{pmatrix}
1 & q & 0 & \cdots & 0 & 0 \\
0 & 0 & q & \cdots & 0 & 0 \\
0 & p & 0 & \cdots & 0 & 0 \\
0 & 0 & p & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & q & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & p & 1
\end{pmatrix}
\]

**Figure 2.3.** The probability of winning is \( p \) and losing is \( q \). The boundaries or end states, 0 and \( N \), are absorbing, \( p_{00} = 1 = p_{NN} \).
2.3. Classification of States

There are three communication classes for the Markov chain graphed in Figure 2.3: \{0\}, \{1, 2, \ldots, N-1\}, and \{N\}. The Markov chain is reducible. The sets \{0\} and \{N\} are closed, but the set \{1, 2, \ldots, N-1\} is not closed. Also, states 0 and N are absorbing; the remaining states are transient. A transient state is defined more formally later. This example is also an illustration of a random walk with absorbing boundaries at 0 and N.

Example 2.3 In an infinite-dimensional random walk or unrestricted random walk, the states are the integers, 0, ±1, ±2, \ldots. Let \( p > 0 \) be the probability of moving to the right and \( q > 0 \) be the probability of moving to the left, \( p + q = 1 \). There are no absorbing boundaries, \( p_{i,i+1} = q \) and \( p_{i+1,i} = p \) for \( i \in \{0, ±1, ±2, \ldots\} \). From the directed graph in Figure 2.4 it is easy to see that the Markov chain is irreducible. Every state in the system communicates with every other state. The set of states forms a closed set. In this case, the transition matrix \( P \) is infinite dimensional. If the states are ordered such that \( \ldots, -1, 0, 1, \ldots \), then matrix \( P \) is an extension of the matrix in Example 2.2 with \( q \) along the superdiagonal and \( p \) along the subdiagonal.

Example 2.4 Suppose the states of the system are \{1, 2, 3, 4, 5\} with directed graph in Figure 2.5 and transition matrix \( P \) given as follows:

\[
P = \begin{pmatrix}
1/2 & 1/3 & 0 & 0 & 0 \\
1/2 & 2/3 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/4 & 0 \\
0 & 0 & 1 & 1/2 & 1 \\
0 & 0 & 0 & 1/4 & 0
\end{pmatrix}.
\]

There are two communication classes, \{1, 2\} and \{3, 4, 5\}. Both classes are closed. The Markov chain is reducible.

---

**Figure 2.4.** Unrestricted random walk; the probability of moving right is \( p \) and left is \( q \).

---

**Figure 2.5.** Directed graph for Example 2.4.
Example 2.5 Suppose the states of the system are \( \{1, 2, \ldots, N\} \) with transition matrix given by \( P \) and directed graph in Figure 2.6. For this example, the Markov chain is irreducible. The set \( \{1, 2, \ldots, N\} \) is closed.

\[
P = \begin{pmatrix}
0 & 0 & \cdots & 0 & 1 \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\]

The chain in Example 2.5 has the property that beginning in state \( i \) it takes exactly \( N \) time steps to return to state \( i \). In addition, \( P^N = I \). The chain is periodic with period equal to \( N \).

Definition 2.11. The period of state \( i \), denoted as \( d(i) \), is the greatest common divisor of all integers \( n \geq 1 \) for which \( p_{ii}^{(n)} > 0 \); that is,

\[
d(i) = \gcd\{n | p_{ii}^{(n)} > 0 \text{ and } n \geq 1\}.
\]

If a state \( i \) has period \( d(i) > 1 \), it is said to be periodic of period \( d(i) \). If the period of a state equals one, it is said to be aperiodic. If \( p_{ii}^{(n)} = 0 \) for all \( n \geq 1 \), we define \( d(i) = 0 \).

It follows from the definition that \( d(i) \) is a nonnegative integer.

Example 2.6 The directed graph of a Markov chain with three states \( \{1, 2, 3\} \) is given in Figure 2.7. The corresponding transition matrix is

\[
P = \begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 1
\end{pmatrix}
\]

It is easy to see that there are three communication classes, \( \{1\} \), \( \{2\} \), and \( \{3\} \). The value of \( d(i) = 0 \) for \( i = 1, 2 \) because \( p_{ii}^{(n)} = 0 \) for \( i = 1, 2 \) and \( n = 1, 2, \ldots \). Also, \( d(3) = 1 \); state 3 is aperiodic.
In the special case \( d(i) = 0 \), it can be shown that the set \( \{i\} \) forms a communication class (see Exercise 5). Also, if \( p_{ii} > 0 \), then \( d(i) = 1 \). Generally, the term periodic is reserved for the case \( d(i) > 1 \).

**Example 2.7** In Example 2.1, the classes are \( \{1\}, \{3\} \) and \( \{2, 4\} \). States 1 and 3 satisfy \( d(1) = d(3) = 0 \). States 2 and 4 are periodic with period 2, \( d(2) = 2 = d(4) \); \( p^{(2n)}_{ii} = 1 \) and \( p^{(2n+1)}_{ii} = 0 \) for \( i = 2, 4 \) and \( n = 1, 2, \ldots \).

Periodicity is a class property; that is, if \( i \leftrightarrow j \), then \( d(i) = d(j) \). All states in one class have the same period. Thus, we can speak of a periodic class or a periodic chain. This result is verified in the next theorem.

**Theorem 2.1.** If \( i \leftrightarrow j \), then \( d(i) = d(j) \).

**Proof.** The case \( d(i) = 0 \) is trivial. Suppose \( d(i) \geq 1 \) and \( p^{(s)}_{ii} > 0 \) for some \( s > 0 \). Then \( d(i) \) divides \( s \). Since \( i \leftrightarrow j \), there exists \( m \) and \( n \) such that \( p^{(m)}_{ij} > 0 \) and \( p^{(n)}_{ji} > 0 \). Then

\[
p^{(n+s+m)}_{jj} \geq p^{(n)}_{jj} p^{(s)}_{ii} p^{(m)}_{ij} > 0.
\]

Also, since \( p^{(2s)}_{ii} > 0 \), \( p^{(n+2s+m)}_{jj} > 0 \). Thus, \( d(j) \) divides \( n + s + m \) and \( n + 2s + m \) and must divide \( (n + 2s + m) - (n + s + m) = s \). Since \( s \) was arbitrary, \( d(j) \leq d(i) \).

Reverse the argument by assuming \( p^{(r)}_{jj} > 0 \). Then it can be shown that \( d(i) \leq d(j) \). Combining these two inequalities gives the desired result, \( d(i) = d(j) \). \( \square \)

In the random walk model with absorbing boundaries, Example 2.2, the classes \( \{0\} \) and \( \{N\} \) are aperiodic. The class \( \{1, 2, \ldots, N - 1\} \) has period 2. In the unrestricted random walk model, Example 2.3, the entire chain is periodic of period 2. In this case, we shall use the notation \( d = 2 \) rather than stating that \( d(i) = 2 \) for each of the \( i \) states. The two classes in Example 2.4 are both aperiodic.

Some additional definitions and notation are needed to define a transient state. This is done in the next section.

### 2.4 First Passage Time

Assume the process begins in state \( i \), \( X_0 = i \). Then we define a first return to state \( i \) and a first passage to state \( j \) for \( j \neq i \).

**Definition 2.12.** Let \( f^{(n)}_{ii} \) denote the probability that, starting from state \( i \), \( X_0 = i \), the first return to state \( i \) is at the \( n \)th time step, \( n \geq 1 \); that is,

\[
f^{(n)}_{ii} = \text{Prob}\{X_n = i, X_m \neq i, m = 1, 2, \ldots, n-1 | X_0 = i\}.
\]
The probabilities \( f_{ii}^{(n)} \) are known as first return probabilities. Define \( f_{ii}^{(0)} = 0 \).

Note that \( f_{ii}^{(1)} = p_{ii} \) but, in general, \( f_{ii}^{(n)} \) is not equal to \( p_{ii}^{(n)} \). The first return probabilities, \( f_{ii}^{(n)} \), represent the first time the chain returns to state \( i \); thus,

\[
0 \leq \sum_{n=1}^{\infty} f_{ii}^{(n)} \leq 1.
\]

A transient state is defined in terms of these first return probabilities.

**Definition 2.13.** State \( i \) is said to be **transient** if \( \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1 \). State \( i \) is said to be **recurrent** if \( \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1 \).

The term **persistent** is sometimes used instead of recurrent (Bailey, 1990). If state \( i \) is recurrent, then the set \( \{f_{ii}^{(n)}\}_{n=0}^{\infty} \) defines a probability distribution for the random variable representing the first return time, which is

\[
T_{ii} = \inf \{m \mid X_m = i \text{ and } X_0 = i\};
\]

that is, \( T_{ii} = n \) with probability \( f_{ii}^{(n)} \), \( n = 0, 1, 2, \ldots \). When state \( i \) is transient, \( \{f_{ii}^{(n)}\}_{n=0}^{\infty} \) does not define a complete set of probabilities necessary to define a probability distribution. However, if \( f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)} < 1 \), then we can define \( 1 - f_{ii} \) as the probability of never returning to state \( i \). The random variable \( T_{ii} \) may be thought of as a “waiting time” until the chain returns to state \( i \).

**Definition 2.14.** The mean of the distribution of \( T_{ii} \) is referred to as the **mean recurrence time** or **mean first return time** for state \( i \) and is denoted as \( \mu_{ii} = E(T_{ii}) \). For a recurrent state \( i \),

\[
\mu_{ii} = \sum_{n=1}^{\infty} nf_{ii}^{(n)}.
\]

Although \( T_{ii} \) is not defined for a transient state, we shall assume the mean recurrence time for a transient state is always infinite (formally \( T_{ii} = \infty \) with probability \( 1 - f_{ii} \)). The mean recurrence time for a recurrent state can be either finite or infinite.

**Definition 2.15.** If a recurrent state \( i \) satisfies \( \mu_{ii} < \infty \), then it is said to be **positive recurrent**, and if it satisfies \( \mu_{ii} = \infty \), then it is said to be **null recurrent**.

Sometimes the term **nonnull recurrent** is used instead of positive recurrent (Bailey, 1990).
Example 2.8 A simple example of a positive recurrent state is an absorbing state. If $i$ is an absorbing state, then $p_{ii} = 1$, so that $f_{ii}^{(1)} = p_{ii} = 1$ and $f_{ii}^{(n)} = 0$ for $n \neq 1$. The mean recurrence time of an absorbing state is $\mu_{ii} = 1$.

Example 2.9 Suppose the transition matrix of a two-state Markov chain satisfies

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

where $0 < p_{ii} < 1$ for $i = 1, 2$. Then all of the elements of matrix $P$ are positive, $p_{ij} > 0$, $i, j = 1, 2$. Hence, $f_{11}^{(1)} = p_{11}$, $f_{11}^{(2)} = p_{12}p_{21}$, $f_{11}^{(3)} = p_{12}p_{22}p_{21}$, and, in general,

$$f_{11}^{(n)} = p_{12}p_{22}^{n-2}p_{21}, \quad n \geq 3.$$

This can be verified easily from the directed graph. Because $p_{22} < 1$, it follows that $\lim_{n \to \infty} f_{11}^{(n)} = 0$ and that

$$\sum_{n=1}^{\infty} f_{11}^{(n)} = p_{11} + p_{12}p_{21} \sum_{n=0}^{\infty} p_{22}^{n} = p_{11} + \frac{p_{12}p_{21}}{1 - p_{22}}.$$

Next, applying the definition of a stochastic matrix, $p_{11} + p_{21} = 1$ and $p_{12} + p_{22} = 1$, it follows that

$$\sum_{n=1}^{\infty} f_{11}^{(n)} = p_{11} + p_{21} = 1,$$

which implies that state 1 is recurrent. Similarly, it can be shown that state 2 is recurrent. In addition, it can be shown that the mean recurrence times are finite; for example,

$$\mu_{11} = p_{11} + p_{12}p_{21} \sum_{n=0}^{\infty} (n + 2)p_{22}^{n} < \infty$$

(see Exercise 6). Therefore, the Markov chain is positive recurrent.

Note that in the definitions of first return probabilities and mean recurrence time, the Markov property was not assumed. These concepts do not require the Markov assumption and are sometimes discussed in the context of renewal processes. These definitions are extended to first passage time probabilities and mean first passage time. Then they are related to Markov chains.

Define the probability $f_{ji}^{(n)}$ for $j \neq i$ in a manner analogous to $f_{ii}^{(n)}$. 


Definition 2.16. Let \( f_{ji}^{(n)} \) denote the probability that, starting from state \( i \), \( X_0 = i \), the first return to state \( j \), \( j \neq i \) is at the \( n \)th time step, \( n \geq 1 \),
\[
f_{ji}^{(n)} = \text{Prob}\{X_n = j, X_m \neq j, m = 1, 2, \ldots, n-1 | X_0 = i\}, \ j \neq i.
\]
The probabilities \( f_{ji}^{(n)} \) are known as first passage time probabilities. Define \( f_{ji}^{(0)} = 0 \).

It follows from the definition that \( 0 \leq \sum_{n=0}^{\infty} f_{ji}^{(n)} \leq 1 \). If \( \sum_{n=0}^{\infty} f_{ji}^{(n)} = 1 \), \( \{f_{ji}^{(n)}\}_{n=0}^{\infty} \) defines a probability distribution for a random variable \( T_{ji} = \inf_{m \geq 1} \{m | X_m = j \text{ and } X_0 = i\} \) known as the first passage to state \( j \) from state \( i \). If \( i = j \), then Definition 2.16 is the same as Definition 2.12.

Definition 2.17. If \( X_0 = i \), then the mean first passage time to state \( j \) is denoted as \( \mu_{ji} = E(T_{ji}) \) and defined as
\[
\mu_{ji} = \sum_{n=1}^{\infty} nf_{ji}^{(n)}, \ j \neq i.
\]
This definition can be extended to include the case \( f_{ji} = \sum_{n=0}^{\infty} f_{ji}^{(n)} < 1 \) by defining the probability of never reaching state \( j \) from state \( i \) as \( 1 - f_{ji} \). If \( f_{ji} < 1 \), then the mean first passage time is infinite.

There exists relationships between the \( n \)-step transition probabilities of a Markov chain and the first return probabilities. The transition from state \( i \) to \( i \) at the \( n \)th step, \( p_{ii}^{(n)} \), may have its first return to state \( i \) at any of the steps \( j = 1, 2, \ldots, n \). It is easy to see that
\[
p_{ii}^{(n)} = f_{ii}^{(0)} p_{ii}^{(n)} + f_{ii}^{(1)} p_{ii}^{(n-1)} + \cdots + f_{ii}^{(n)} p_{ii}^{(0)}
\]
\[
= \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)}, \quad (2.9)
\]
since \( f_{ii}^{(0)} = 0 \) and \( p_{ii}^{(0)} = 1 \). A similar relationship exists for \( f_{ji}^{(n)} \) and \( p_{ji}^{(n)} \):
\[
p_{ji}^{(n)} = \sum_{k=1}^{n} f_{ji}^{(k)} p_{jj}^{(n-k)}, \ j \neq i. \quad (2.10)
\]

Let the generating function for the sequence \( \{f_{ji}^{(n)}\} \) be
\[
F_{ji}(s) = \sum_{n=0}^{\infty} f_{ji}^{(n)} s^n, \ |s| < 1
\]
and the generating function for the sequence \( \{p_{ji}^{(n)}\} \) be
\[
P_{ji}(s) = \sum_{n=0}^{\infty} p_{ji}^{(n)} s^n, \ |s| < 1
\]
for all states $i, j$ of the Markov chain. Note that these functions may not be probability generating functions since the set of probabilities $\{p_{ji}^{(n)}\}_{n=0}^{\infty}$ and $\{p_{ij}^{(n)}\}_{n=0}^{\infty}$ may not represent a probability distribution (the sum may be less than one). However, relationships between these two generating functions are shown and these relationships are used to prove results about Markov chains.

Multiply $F_{ii}(s)$ by $P_{ii}(s)$ using the definition for the product of two series. The product $C(s)$ of two series $A(s)$ and $B(s)$, where $A(s) = \sum_{k=0}^{\infty} a_k s^k$ and $B(s) = \sum_{l=0}^{\infty} b_l s^l$, is

$$C(s) = A(s)B(s) = \sum_{r=0}^{\infty} c_r s^r,$$

where

$$c_r = a_0 b_r + a_1 b_{r-1} + \cdots + a_r b_0 = \sum_{k=0}^{r} a_k b_{r-k}.$$

If $A(s)$ and $B(s)$ converge on the interval $(-1, 1)$, then $C(s)$ also converges on $(-1, 1)$ (Wade, 2000). Identify the coefficient $a_k$ of $A(s)$ with $f_{ii}^{(k)}$ of $F_{ii}(s)$ and the coefficient $b_l$ of $B(s)$ with $p_{ii}^{(l)}$ of $P_{ii}(s)$ and apply equation (2.9) so that $c_r = p_{ii}^{(r)}$. The following relationship between the generating functions is obtained:

$$F_{ii}(s)P_{ii}(s) = \sum_{r=1}^{\infty} p_{ii}^{(r)} s^r = P_{ii}(s) - 1,$$

where $p_{ii}^{(r)} = \sum_{k=1}^{r} f_{ii}^{(k)} p_{ii}^{(r-k)}$ and $|s| < 1$. Note that the number one is subtracted from $P_{ii}(s)$ since the first term $c_0 = f_{ii}^{(0)} p_{ii}^{(0)}$ in the product of $F_{ii}(s)P_{ii}(s)$ is zero but the first term in $P_{ii}(s)$ is $p_{ii}^{(0)} = 1$. Hence,

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)}. \quad (2.11)$$

A similar relationship exists between $P_{ji}(s)$ and $F_{ji}(s)$ that follows from (2.10):

$$F_{ji}(s)P_{ji}(s) = P_{ji}(s), \quad i \neq j. \quad (2.12)$$

For equation (2.12), the number one is not subtracted from $P_{ji}(s)$ since the first term in its series representation is $p_{ij}^{(0)} = 0, i \neq j$, which equals the first term in the series representation of $F_{ji}(s)P_{ji}(s)$; that is, $c_0 = f_{ji}^{(0)} p_{ji}^{(0)} = 0$. The above identities are used to verify some theoretical results on Markov chains.
2.5 Basic Theorems for Markov Chains

The relationships between the generating functions are used to relate a recurrent state \( i \) to the \( n \)-step transition probabilities \( p_{ii}^{(n)} \). In addition, an important result referred to as the basic limit theorem for Markov chains is stated that gives conditions for a Markov chain to have a limiting distribution. First, a lemma is needed.

**Lemma 2.1 (Abel's Convergence Theorem).**

(i) If \( \sum_{k=0}^{\infty} a_k \) converges, then \( \lim_{s \to 1^-} \sum_{k=0}^{\infty} a_k s^k = \sum_{k=0}^{\infty} a_k = a \).

(ii) If \( a_k \geq 0 \) and \( \lim_{s \to 1^-} \sum_{k=0}^{\infty} a_k s^k = a \leq \infty \), then \( \sum_{k=0}^{\infty} a_k = a \).

For a proof of Abel's convergence theorem, consult Karlin and Taylor (1975, pp. 64–65). The lemma is straightforward if the series converges absolutely. Lemma 2.1 is used to verify the following theorem.

**Theorem 2.2.** A state \( i \) is recurrent (transient) if and only if \( \sum_{n=0}^{\infty} p_{ii}^{(n)} \) diverges (converges); that is,

\[
\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty \text{ (or } < \infty \text{).}
\]

**Proof.** We prove the theorem in the case of a recurrent state. The proof in the case of a transient state follows as a direct consequence because if a state \( i \) is not recurrent it is transient. Assume state \( i \) is recurrent; that is,

\[
\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1.
\]

By part (i) of Lemma 2.1,

\[
\lim_{s \to 1^-} \sum_{n=1}^{\infty} f_{ii}^{(n)} s^n = \lim_{s \to 1^-} F_{ii}(s) = 1.
\]

From the identity (2.11), it follows that

\[
\lim_{s \to 1^-} P_{ii}(s) = \lim_{s \to 1^-} \frac{1}{1 - P_{ii}(s)} = \infty.
\]

Because \( P_{ii}(s) = \sum_{n=0}^{\infty} p_{ii}^{(n)} s^n \), it follows from Lemma 2.1 part (ii) that

\[
\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.
\]
The converse of the theorem is proved by contradiction. Assume that
\[ \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty \] and state \( i \) is transient; that is,
\[ \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1. \]

Applying Lemma 2.1 part (i),
\[ \lim_{s \to 1^-} F_{it}(s) = \lim_{s \to 1^-} \sum_{n=0}^{\infty} f_{ii}^{(n)} s^n = \sum_{n=0}^{\infty} f_{ii}^{(n)} = \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1. \]

Now, applying the identity (2.11), it follows that
\[ \lim_{s \to 1^-} P_{ii}(s) = \lim_{s \to 1^-} \frac{1}{1 - F_{it}(s)} < \infty. \]

Finally, Lemma 2.1 part (ii) and the above inequality yield
\[ \sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty, \]
which contradicts the original assumption. The theorem is verified. \( \square \)

Next it is shown that if state \( i \) is recurrent, then all states in the same communicating class are recurrent. Thus, recurrence and transience are class properties. If the chain is irreducible, then the chain is either recurrent or transient.

**Corollary 2.1.** Assume \( i \leftarrow j \). State \( i \) is recurrent (transient) if and only if state \( j \) is recurrent (transient).

**Proof.** Suppose \( i \leftarrow j \) and state \( i \) is recurrent. Then there exists \( n, m \geq 1 \) such that
\[ p_{ij}^{(n)} > 0 \quad \text{and} \quad p_{ji}^{(m)} > 0. \]

Let \( k \) be a nonnegative integer,
\[ p_{ji}^{(m+n+k)} \geq p_{ji}^{(m)} (k) \frac{p_{ij}^{(n)}}{p_{ii}}. \]

Summing on \( k \),
\[ \sum_{k=0}^{\infty} p_{ji}^{(k)} \geq \sum_{k=0}^{\infty} p_{ji}^{(n+m+k)} \geq \sum_{k=0}^{\infty} p_{ji}^{(m)} (k) \frac{p_{ij}^{(n)}}{p_{ii}} \sum_{k=0}^{\infty} p_{ji}^{(k)} = p_{ji}^{(m)} p_{ij}^{(n)} \sum_{k=0}^{\infty} p_{ii}^{(k)}. \]

The right-hand side is infinite by Theorem 2.2 because state \( i \) is recurrent. Thus, \( \sum_{k=0}^{\infty} p_{ji}^{(k)} \) is divergent and state \( j \) is recurrent. The theorem also holds for transient states because if a state is not recurrent it is transient. \( \square \)
An important property about recurrent classes follows from the definition of a recurrent state. The next corollary shows that a recurrent class forms a closed set.

**Corollary 2.2.** Every recurrent class in a discrete time Markov chain is a closed set.

*Proof.* Let $C$ be a recurrent class. Suppose $C$ is not closed. Then for some $i \in C$ and $j \notin C$, $p_{ji} > 0$. Because $j \notin C$, it is impossible to return to the set $C$ from state $j$ (otherwise $i \leftrightarrow j$). Thus, beginning from state $i$, the probability of never returning to $C$ is at least $p_{ji}$ or $\sum_{n=1}^{\infty} p_{ji}^{(n)} \leq 1 - p_{ji} < 1$, a contradiction to the fact that $i$ is a recurrent state. Hence, $C$ must be closed. \qed

**Example 2.10** Suppose the transition matrix of a Markov chain with states $\{1, 2, 3, \ldots\}$ satisfies

$$P = \begin{pmatrix}
    a_1 & 0 & 0 & \cdots \\
    a_2 & a_1 & 0 & \cdots \\
    a_3 & a_2 & a_1 & \cdots \\
    \vdots & \vdots & \vdots & \ddots
\end{pmatrix},$$

where $a_i > 0$ and $\sum_{i=1}^{\infty} a_i = 1$. The communication classes consist of $\{1\}$, $\{2\}$, $\{3\}$, and so on. Each state represents a communication class. In addition, none of the classes are closed. Hence, by Corollary 2.2, it follows that none of the classes are recurrent, they must all be transient. In fact, each class is aperiodic and transient. \qed

We will use Example 2.3, the one-dimensional unrestricted random walk, to illustrate Theorem 2.2. It will be shown that the Markov chain for this process is recurrent if and only if (iff) the probabilities of moving right or left are equal, $p = 1/2 = q$, which means it is a symmetric random walk.

**Example 2.11** Consider the one-dimensional, unrestricted random walk in Example 2.3. The chain is irreducible and periodic of period 2. Let $p$ be the probability of moving to the right and $q$ be the probability of moving left, $p + q = 1$. We verify that the state 0 or the origin is recurrent iff $p = 1/2 = q$. However, if the origin is recurrent, then all states are recurrent because the chain is irreducible. Notice that starting from the origin, it is impossible to return to the origin in an odd number of steps,

$$p_{00}^{(2n+1)} = 0 \quad \text{for} \quad n = 0, 1, 2, \ldots.$$  

The chain has period 2 because only in an even numbers of steps is the transition probability positive. In $2n$ steps, there are a total of $n$ steps to the right and a total of $n$ steps to the left, and the $n$ steps to the left must
be the reverse of those steps taken to the right in order to return to the origin. In particular, in $2n$ steps, there are
\[
\binom{2n}{n} = \frac{(2n)!}{n!n!}
\]
different paths (combinations) that begin and end at the origin. Also, the probability of occurrence of each one of these paths is $p^n q^n$. Thus,
\[
\sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n.
\]

We need an asymptotic formula for $n!$ known as Stirling's formula to verify recurrence. The notation $f(n) \sim g(n)$ means
\[
\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.
\]

Both quantities, $f(n)$ and $g(n)$, grow at the same rate as $n \to \infty$. In Stirling's formula,
\[
\lim_{n \to \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1.
\]

Thus, Stirling's formula is the following asymptotic relation:
\[
[n! \sim n^n e^{-n} \sqrt{2\pi n}]
\]

Verification of Stirling's formula can be found in Feller (1968) or Norris (1997).

Stirling's formula gives the following approximation:
\[
p_{00}^{(2n)} = \frac{(2n)!}{n!n!} p^n q^n
\]
\[
\sim \frac{\sqrt{4\pi n (2n) e^{-2n}}}{2\pi n^{2n+1} e^{-2n}} p^n q^n
\]
\[
= \frac{(4pq)^n}{\sqrt{\pi n}}.
\]

Thus, there exists a positive integer $N$ such that for $n \geq N$,
\[
\frac{(4pq)^n}{2\sqrt{\pi n}} < p_{00}^{(2n)} < \frac{2(4pq)^n}{\sqrt{\pi n}}.
\]

Considered as a function of $p$, the expression $4pq = 4p(1-p)$ has a maximum at $p = 1/2$. If $p = 1/2$, then $4pq = 1$ and if $p \neq 1/2$, then $4pq < 1$. When $pq \neq 1/4$, then
\[
\sum_{n=0}^{\infty} p_{00}^{(2n)} < N + \sum_{n=N}^{\infty} \frac{2(4pq)^n}{\sqrt{\pi n}} < \infty.
\]
The latter series converges by the ratio test. When \( pq = 1/4 \), we have
\[
\sum_{n=0}^{\infty} p_{00}^{(2n)} > \sum_{n=N}^{\infty} \frac{(4pq)^n}{2\sqrt{\pi n}} = \frac{1}{2\sqrt{\pi}} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty.
\]
The latter series diverges because it is just a multiple of a divergent \( p \)-series. Therefore, the series \( \sum_{n=0}^{\infty} p_{00}^{(2n)} \) diverges if \( p = 1/2 = q \), which means the one-dimensional random walk is recurrent if it is a symmetric random walk. If \( p \neq q \), then all states are transient; there is a positive probability that an object starting from the origin will never return to the origin. What happens if an object never returns to the origin? It can be shown that either the object tends to \( +\infty \) or to \( -\infty \).

Before giving additional examples, some important results concerning irreducible and recurrent Markov chains are stated. Verification of these results are quite lengthy and the proofs are not given. They depend on a result from renewal theory known as the discrete renewal theorem. A statement of the discrete renewal theorem, its proof, and proofs of the following theorems can be found in Karlin and Taylor (1975) (also consult Norris, 1997). The first result is known as the basic limit theorem for aperiodic Markov chains, which gives conditions for recurrent, irreducible, and aperiodic Markov chains to have a limiting probability distribution. The second result applies to periodic Markov chains.

**Theorem 2.3 (Basic limit theorem for aperiodic Markov chains).**
Consider a recurrent, irreducible, and aperiodic Markov chain. Then
\[
\lim_{n \to \infty} p_{ij}^{(n)} = \frac{1}{\mu_{ii}},
\]
where \( \mu_{ii} \) is the mean recurrence time for state \( i \) defined by (2.8) and \( i \) and \( j \) are any states of the chain. [If \( \mu_{ii} = \infty \), then \( \lim_{n \to \infty} p_{ij}^{(n)} = 0 \).]

**Theorem 2.4 (Basic limit theorem for periodic Markov chains).**
Consider a recurrent, irreducible, and \( d \)-periodic Markov chain. Then
\[
\lim_{n \to \infty} p_{ii}^{(nd)} = \frac{d}{\mu_{ii}},
\]
and \( p_{ii}^{(m)} = 0 \) if \( m \) is not a multiple of \( d \), where \( \mu_{ii} \) is the mean recurrence time for state \( i \) defined by (2.8). [If \( \mu_{ii} = \infty \), then \( \lim_{n \to \infty} p_{ii}^{(nd)} = 0 \).]

In Theorem 2.4, \( d \geq 1 \), since the chain is irreducible (see Exercise 5).

**Example 2.12** The Markov chain in Example 2.5 is periodic with period \( d = N \). The Markov chain is also irreducible and recurrent. Applying Theorem 2.4, \( \lim_{n \to \infty} p_{ii}^{(nN)} = N/\mu_{ii} \). However, \( P^N = I \), so that \( p_{ii}^{(nN)} = 1 \),
and if $i \neq j$, $p_{ij}^{(n)} = 0$ for $n = 1, 2, \ldots$. Therefore, $1 = N/\mu_{ii}$, which implies that the mean recurrence time is $N$. This result is obvious if one notices that $f_i^{(N)} = 1$ and $f_i^{(n)} = 0$ for $n \neq N$. Then $\mu_{ii} = N f_i^{(N)} = N$. 

Theorems 2.3 and 2.4 apply to recurrent classes as well as to recurrent chains. Suppose $C$ is a recurrent communication class. Then since $C$ is closed, $p_{ki}^{(n)} = 0$ for $i \in C$ and $k \notin C$ for $n \geq 1$. Therefore, the submatrix $P_C$ of $P$ given by $P_C = (p_{ij})_{i,j \in C}$ is a transition matrix for $C$. By Corollary 2.1, the associated Markov chain for $C$ is irreducible and recurrent. Therefore, Theorems 2.3 and 2.4 can be applied to any aperiodic, recurrent class (rather than a chain) or to any periodic, recurrent class. We state the extension of Theorem 2.3 as a corollary.

**Corollary 2.3.** If $i$ and $j$ are any states in a recurrent and aperiodic class of a Markov chain, then

$$\lim_{n \to \infty} p_{ij}^{(n)} = \frac{1}{\mu_{ii}},$$

where $\mu_{ii}$ is defined in (2.8).

**Example 2.13** The Markov chain in Example 2.4 has two aperiodic, recurrent classes, \{1, 2\} and \{3, 4, 5\}. Then $\lim_{n \to \infty} p_{ij}^{(n)} = 1/\mu_{ii}$ for $i, j = 1, 2$ and for $i, j = 3, 4, 5$. Note that $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ when $i \in \{1, 2\}$ and $j \in \{3, 4, 5\}$ or when $i \in \{3, 4, 5\}$ and $j \in \{1, 2\}$. We shall show in the next section how to compute the limit, $1/\mu_{ii}$. 

If $\mu_{ii} = \infty$, then state $i$ is null recurrent and if $0 < \mu_{ii} < \infty$, then state $i$ is positive recurrent. It can be shown that if one state is positive recurrent in a communication class, then all states in that class are positive recurrent. In this case, the entire class is positive recurrent. In addition, it follows that if one state is null recurrent in a communication class, then all states are null recurrent. Verification of these results is left as an exercise (see Exercise 11). Hence, null recurrence and positive recurrence are class properties. Therefore, it follows from the previous results that every irreducible Markov chain can be classified as either periodic or aperiodic and as either transient, null recurrent, or positive recurrent:

(1) periodic or (2) aperiodic.

* (i) transient or (ii) null recurrent or (iii) positive recurrent.

The classifications (1) and (2) are disjoint, and the three classifications, (i), (ii), and (iii), are disjoint. This classification scheme can be applied to communication classes as well, provided the period $d \geq 1$. The special case where $d = 0$ consists of a class with only a single element and the class must be transient (see Example 2.6 and Exercise 5). The term ergodic is used to classify states or irreducible chains that are aperiodic and positive recurrent.
Definition 2.18. A state is ergodic if it is both aperiodic and positive recurrent. An ergodic chain is a Markov chain that is irreducible, aperiodic, and positive recurrent.

When the entire class or chain is ergodic, it is also referred to as strongly ergodic (Karlin and Taylor, 1975). If the ergodic class or chain is null recurrent rather than positive recurrent, then it is said to be weakly ergodic (Karlin and Taylor, 1975).

The next example reconsiders the unrestricted random walk model. It has already been shown that this Markov chain is irreducible and periodic. In the case of a symmetric random walk, it is shown that the chain is null recurrent.

Example 2.14 The unrestricted random walk model is irreducible and periodic with period \( d = 2 \). The chain is recurrent if it is a symmetric random walk, \( p = 1/2 = q \) (Example 2.11). Recall that the \( 2n \)-step transition probability satisfies

\[
P_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}
\]

[see equation (2.13)] and hence, \( \lim_{n \to \infty} P_{00}^{(2n)} = 0 \). It follows from the basic limit theorem for periodic Markov chains that \( d/\mu_{00} = 0 \). Thus, \( \mu_{00} = \infty \); the zero state is null recurrent. Since the chain is irreducible, all states are null recurrent. Thus, when \( p = 1/2 = q \), the chain is periodic and null recurrent and when \( p \neq 1/2 \), the chain is periodic and transient.

\[\blacksquare\]

2.6 Stationary Probability Distribution

A stationary probability distribution represents an "equilibrium" of the Markov chain; that is, a probability distribution that remains fixed in time. For instance, if the chain is initially at a stationary probability distribution, \( p(0) = \pi \), then \( p(n) = P^n \pi = \pi \) for all time \( n \).

Definition 2.19. A stationary probability distribution of a Markov chain with states \( \{1, 2, \ldots\} \) is a nonnegative vector \( \pi = (\pi_1, \pi_2, \ldots)^T \) that satisfies \( P\pi = \pi \) and whose elements sum to one (i.e., \( \sum_{i=1}^{\infty} \pi_i = 1 \)).

Definition 2.19 also applies to a finite Markov chain, where the vector \( \pi = (\pi_1, \pi_2, \ldots, \pi_N)^T \) and \( \sum_{i=1}^{N} \pi_i = 1 \). In the finite case, \( \pi \) is a right eigenvector of \( P \) corresponding to the eigenvalue \( \lambda = 1 \), \( P\pi = \lambda \pi \). There may be one or more than one linearly independent eigenvector corresponding to the eigenvalue \( \lambda = 1 \). In fact, there may be at most \( N \) linearly independent eigenvectors. If there is more than one linearly independent eigenvector corresponding to \( \lambda = 1 \), then the stationary probability distribution of the finite Markov chain is not unique.
Example 2.15 Suppose we have the transition matrix

\[ P = \begin{pmatrix} p & 0 & q \\ q & p & 0 \\ 0 & q & p \end{pmatrix}, \]

where \( p > 0, q > 0 \) and \( p + q = 1 \). To determine \( \pi \), we solve \( P\pi = \pi \) or

\[ (P - I)\pi = 0, \]

where \( I \) is the \( 3 \times 3 \) identity matrix and \( 0 \) is the zero vector. There is only one linearly independent eigenvector corresponding to the eigenvalue \( \lambda = 1 \). The unique stationary probability distribution satisfies \( \pi_1 = \pi_2 = \pi_3 \) so that \( \pi = (1/3, 1/3, 1/3)^T \).

Example 2.16 If the transition matrix \( P \) is the \( N \times N \) identity matrix, then there exist \( N \) linearly independent eigenvectors, \( e_1 = (1, 0, \ldots, 0)^T, \ldots, e_N = (0, 0, \ldots, 1)^T \), corresponding to the eigenvalue \( \lambda = 1 \). Hence, there is an infinite number of stationary probability distributions. Any vector \( \pi = (\pi_1, \pi_2, \ldots, \pi_N)^T \), where \( \pi_i \geq 0 \) for \( i = 1, 2, \ldots, N \), and \( \sum_{i=1}^{N} \pi_i = 1 \) is a stationary probability distribution.

A stationary probability distribution for a finite state Markov chain always exists; although it may not be unique. This is due to the fact that a finite stochastic matrix always has an eigenvalue \( \lambda = 1 \). However, if the state space is infinite, a stationary probability distribution may not exist.

Example 2.17 Consider the transition matrix in Example 2.10,

\[ P = \begin{pmatrix} a_1 & 0 & 0 & \cdots \\ a_2 & a_1 & 0 & \cdots \\ a_3 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \]

where \( a_i > 0 \) and \( \sum_{i=1}^{\infty} a_i = 1 \). There exists no stationary probability distribution because \( P\pi = \pi \) implies \( \pi = 0 \), the zero vector. It is impossible for the sum of the elements of \( \pi \) to equal one.

As illustrated in the previous examples, nonexistence of a stationary distribution only applies to infinite Markov chains. It can be shown that every finite Markov chain has at least one stationary probability distribution (Gantmacher, 1964). In addition, if the finite Markov chain is irreducible, it has a unique stationary probability distribution (Ortega, 1987). The Markov chain in Example 2.15 is irreducible, but the one in Example 2.16 is reducible.

If a Markov chain is irreducible, positive recurrent, and aperiodic, then the next theorem shows that there exists a unique stationary probability
distribution and, in addition, this distribution is the limiting distribution of the Markov chain. It is the property of aperiodicity that is needed for convergence to the stationary probability distribution. A proof of this result can be found in Karlin and Taylor (1975).

**Theorem 2.5.** Suppose a discrete time Markov chain is irreducible, positive recurrent, and aperiodic (strongly ergodic) with states \( \{1, 2, \ldots\} \) and transition matrix \( P \). Then there exists a unique positive stationary probability distribution \( \pi = (\pi_1, \pi_2, \ldots)^T \), \( P\pi = \pi \), such that

\[
\lim_{n \to \infty} p_{ij}^{(n)} = \pi_i \quad \text{for} \quad i, j = 1, 2, \ldots.
\]

Theorem 2.5 gives sufficient conditions on the chain for existence and uniqueness of the limiting probability distribution. The transition matrix of a strongly ergodic chain satisfies

\[
\lim_{n \to \infty} P^n = \begin{pmatrix}
\pi_1 & \pi_1 & \pi_1 & \cdots \\
\pi_2 & \pi_2 & \pi_2 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
\]

Thus,

\[
\lim_{n \to \infty} P^n p(0) = \pi. \quad (2.14)
\]

The limit is independent of the initial distribution and equals the stationary probability distribution. The convergence to a stationary probability distribution is similar to convergence to a stable equilibrium in a deterministic model. Theorem 2.5 applies to finite and infinite Markov chains. The following example shows that a unique stationary probability distribution may exist but that the Markov chain may not converge to that distribution.

**Example 2.18** Suppose the transition matrix of Markov chain satisfies

\[
P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

This chain is irreducible, positive recurrent, and periodic of period 2. There exists a unique stationary distribution, \( \pi = (1/2, 1/2)^T \), but there is no limiting distribution. For any initial distribution \( p(0) \), \( P^{2n} p(0) = p(0) \) and \( P^{2n+1} p(0) = p(1) \). This example shows the necessity of aperiodicity in Theorem 2.5.

Comparing Theorem 2.5 with the basic limit theorem for aperiodic Markov chains, it follows that

\[
\pi_i = \frac{1}{\mu_{ii}} > 0,
\]

where \( \mu_{ii} \) is the mean recurrence time for state \( i \). The mean recurrence time of a positive recurrent, irreducible, and aperiodic chain can be computed from the stationary probability distribution, \( \mu_{ii} = 1/\pi_i \).
Example 2.19 Suppose the transition matrices for two Markov chains are

\[
P_1 = \begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0 & 1/4 & 0 \\ 1 & 1/2 & 1 \\ 0 & 1/4 & 0 \end{pmatrix}.
\]

These Markov chains are strongly ergodic. (In the next section, it is shown that all irreducible finite Markov chains are positive recurrent.) These two Markov chains represent the two recurrent classes in the Markov chain discussed in Examples 2.4 and 2.13.

There exist unique limiting stationary probability distributions \( \pi \) for each matrix \( P_i \). The stationary distribution corresponding to \( P_1 \) satisfies \( \pi_1 = (2/3)\pi_2 \) and \( 1 = \pi_1 + \pi_2 = (2/3)\pi_2 + \pi_2 \), so that \( \pi = (2/5, 3/5)^T \). In addition, the mean recurrence times are \( \mu_{11} = 5/2 \) and \( \mu_{22} = 5/3 \). For matrix \( P_2 \), it can be shown that the stationary probability distribution is \( \pi = (1/6, 2/3, 1/6)^T \). Hence, the mean recurrence times are \( \mu_{11} = 6 \), \( \mu_{22} = 3/2 \), and \( \mu_{33} = 6 \).

It follows from Theorem 2.5 that the columns of \( P_i^n \) approach the stationary probability distribution,

\[
\lim_{n \to \infty} P_1^n = \begin{pmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{pmatrix} \quad \text{and} \quad \lim_{n \to \infty} P_2^n = \begin{pmatrix} 1/6 & 1/6 & 1/6 \\ 2/3 & 2/3 & 2/3 \\ 1/6 & 1/6 & 1/6 \end{pmatrix}.
\]

For example, in the first Markov chain, eventually, 40% of the time is spent in state 1 and 60% of the time is spent in state 2. After leaving state 1, it takes on the average about 2.5 time steps until there is a return to state 1, and, after leaving state 2, it takes about 1.67 time steps until there is a return to state 2.

2.7 Finite Markov Chains

An important property of finite Markov chains is that there are no null recurrent states and not all states can be transient. Therefore, an irreducible finite Markov chain is positive recurrent. The assumption of recurrence is not needed when the basic limit theorems are applied to finite Markov chains. To verify these results, we begin with a lemma.

Lemma 2.2. If \( j \) is a transient state of a Markov chain and \( i \) is any state in the Markov chain, then \( \lim_{n \to \infty} p_{ji}^{(n)} = 0 \).

Actually, Lemma 2.2 applies to finite and infinite Markov chains. The proof is left as an exercise.

Theorem 2.6. In a finite Markov chain, not all states can be transient and no states can be null recurrent. In particular, an irreducible finite Markov chain is positive recurrent.
Proof. From Lemma 2.2, it follows that if \( j \) is transient, then
\[
\lim_{n \to \infty} p_{ji}^{(n)} = 0, \quad \text{for } i = 1, 2, \ldots, N,
\]  
(2.15)

where \( N \) is the number of states. Suppose all states are transient. Then the identity (2.15) holds for all \( i \) and \( j, i, j = 1, 2, \ldots, N \) and
\[
\lim_{n \to \infty} P^n = 0,
\]
the zero matrix. Matrix \( P^n \) is a stochastic matrix (Exercise 1). Hence, 
\[
\sum_{j=1}^{N} p_{ji}^{(n)} = 1; \quad \text{the column sums of } P^n \text{ are one. Taking the limit as } n \to \infty
\]
and interchanging the limit and summation (possible because of the finite sum) leads to 
\[
\sum_{j=1}^{N} \lim_{n \to \infty} p_{ji}^{(n)} = 1, \quad \text{a contradiction to the above limit. Thus, not all states can be transient.}
\]

Suppose there exists a null recurrent state \( i \) and \( i \in C \), where \( C \) is a class of states. The class \( C \) is closed by Corollary 2.2 and all states in \( C \) are null recurrent. (See the remarks following Corollary 2.3 and Exercise 11.)

Suppose the class \( C \) is aperiodic and null recurrent. Then according to the basic limit theorem for aperiodic Markov chains, \( \lim_{n \to \infty} P_{ij}^{(n)} = 0 \) for all \( i, j \in C \). The submatrix \( P_C \) of \( P \) consisting of all states in \( C \) is a stochastic matrix \( (p_{kj}^{(n)} = 0 \text{ for } k \notin C) \). But \( \lim_{n \to \infty} P_C^n = 0 \), an impossibility. Thus, all states are positive recurrent.

Suppose the class \( C \) is periodic and null recurrent. Then according to the basic limit theorem for periodic Markov chains, \( \lim_{n \to \infty} P_{ij}^{(n)} = 0 \) for any \( i \in C \). Furthermore, for any state \( j \in C \), since \( i \leftrightarrow j \), there exists positive integers \( m \) and \( n \) such that \( p_{ij}^{(m)} > 0 \) and \( p_{ji}^{(n)} > 0 \). Therefore,
\[
p_{ii}^{(m+n)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0.
\]

Fix \( n \) and let \( m \to \infty \). Then it follows that \( \lim_{m \to \infty} p_{ji}^{(m)} = 0 \). Also, fix \( m \) and let \( n \to \infty \). Then \( \lim_{n \to \infty} p_{ij}^{(n)} = 0 \). Thus, \( \lim_{n \to \infty} P_C^n = 0 \), where \( P_C \) is the submatrix of \( P \) consisting of states in \( C \). This is an impossibility since \( P_C^n \) is stochastic. Thus, all states are positive recurrent.

In the case that the finite Markov chain is irreducible, there is only one class and all states in that class must be either positive recurrent, null recurrent, or transient. Since they cannot all be transient and there are no null recurrent states, they all must be positive recurrent. \( \square \)

Since there are no null recurrent states in finite Markov chains, there are only four different types of classification schemes based on periodicity and recurrence. The states of a finite Markov chain can be classified as either periodic or aperiodic and either transient or positive recurrent. Recurrence in a finite Markov chain will always mean positive recurrence.
Example 2.20 Let $P$ be the transition matrix of a Markov chain:

$$P = \begin{pmatrix}
1/2 & 0 & 0 & 1/2 \\
1/2 & 1 & 0 & 0 \\
0 & 0 & 0 & 1/2 \\
0 & 0 & 1 & 0 \\
\end{pmatrix}.$$

There are three communication classes, \{1\}, \{2\}, and \{3, 4\}. Class \{1\} is aperiodic and transient. Class \{2\} is aperiodic and recurrent. Class \{3, 4\} is periodic and transient. State 2 is an absorbing state. Matrix $P$ can be partitioned according to the classes,

$$P = \begin{pmatrix}
1/2 & 0 & 0 & 1/2 \\
- & - & - & - \\
- & - & - & - \\
1/2 & 1 & 0 & 0 \\
- & - & - & - \\
0 & 0 & 0 & 1/2 \\
0 & 0 & 1 & 0 \\
\end{pmatrix} = \begin{pmatrix}
T_1 & 0 & A \\
B & 1 & 0 \\
0 & 0 & T_3 \\
\end{pmatrix}.$$

The diagonal matrices $T_1$ and $T_3$ corresponding to the transient classes have the property that

$$\lim_{n \to \infty} T_i^n = 0.$$

In addition, the entire first, third, and fourth rows all tend to zero as $n \to \infty$ (Lemma 2.2). Eventually, all transient classes are absorbed into the recurrent classes. In this example, there is eventual absorption into state 2. ■

An additional property of finite Markov chains is that a communication class that is closed is recurrent. This is verified in the next theorem. See Schinazi (1999).

Theorem 2.7. In a finite Markov chain, a class is recurrent iff it is closed.

Proof. We have already shown that if a class is recurrent, it is closed. The reverse implication is verified by contradiction. Assume a class of states $C$ is closed, but $C$ is not recurrent. Then $C$ is transient. By Lemma 2.2, if $j$ is a transient state, then $\lim_{n \to \infty} p_{ji}^{(n)} = 0$ for all states $i$. In particular, for state $i \in C$, we have

$$\sum_{j \in C} \lim_{n \to \infty} p_{ji}^{(n)} = 0. \quad (2.16)$$

But because $C$ is closed, the submatrix $P_C$ consisting of all states in $C$ is a stochastic matrix. In addition, $P_C^{(n)}$ is a stochastic matrix (Exercise 1). The column sums of $P_C^{(n)}$ equal one and must equal one in the limit as $n \to \infty$, a contradiction to (2.16). Hence, $C$ cannot be transient; $C$ must be recurrent. ■
Example 2.21 The finite Markov chain in Examples 2.4 and 2.13 has two recurrent classes, \{1, 2\} and \{3, 4, 5\}. The transition matrix can be partitioned according to these classes,

\[
P = \begin{pmatrix}
1/2 & 1/3 & 0 & 0 & 0 \\
1/2 & 2/3 & 0 & 0 & 0 \\
- & - & - & - & - \\
0 & 0 & 0 & 1/4 & 0 \\
0 & 0 & 1 & 1/2 & 1 \\
0 & 0 & 0 & 1/4 & 0
\end{pmatrix}.
\]

Since both classes are closed, they are both recurrent. They are also aperiodic. In addition, from the partition, it is easy to see that each of the diagonal submatrices forms a stochastic matrix.

In finite Markov chain theory, a stochastic matrix with the property \(p_{ij}^{(n)} > 0\), for some \(n > 0\) and all \(i, j = 1, 2, \ldots, N\) (\(P^n > 0\)), is often referred to as a regular matrix. If the transition matrix is regular, then the Markov chain is irreducible and aperiodic (Why?). The Markov chain, in this case, is also referred to as regular. Therefore, a regular Markov chain is positive recurrent (strongly ergodic). The theorems of Perron and Frobenius from linear algebra state that a regular matrix \(P\) has a positive dominant eigenvalue \(\lambda\) that is simple and satisfies \(\lambda > |\lambda_i|\), where \(\lambda_i\) is any other eigenvalue of \(P\) (see Gantmacher, 1964, or Ortega, 1987). In addition, the dominant eigenvalue \(\lambda\) has an associated positive eigenvector (see Gantmacher, 1964). The eigenvalue \(\lambda\) of a regular stochastic matrix \(P\) is \(\lambda = 1\), and the associated eigenvector \(\pi\) satisfying \(\sum \pi_j = 1\) defines a stationary probability distribution, \(P\pi = \pi\). If the assumption of regularity is weakened, so that the stochastic matrix \(P\) is irreducible, then the theorems of Perron and Frobenius still imply that \(\lambda = 1\) is a simple eigenvalue satisfying \(\lambda \geq |\lambda_i|\) with associated positive eigenvector \(\pi\) (Ortega, 1987). Therefore, all that is required for existence of a unique stationary probability distribution is that \(P\) be irreducible. However, the additional property of aperiodicity (or regularity) is needed to show convergence to the stationary probability distribution. The following result is a corollary of Theorems 2.5 and 2.6.

Corollary 2.4. Suppose a finite Markov chain is irreducible and aperiodic. Then there exists a unique stationary probability distribution \(\pi = (\pi_1, \ldots, \pi_N)^T\) such that

\[
\lim_{n \to \infty} p_{ij}^{(n)} = \pi_i, \quad i, j = 1, 2, \ldots, N.
\]

The next example is a finite Markov chain that models the dynamics of two squirrel populations.
Example 2.22 The introduction of a new or alien species into an environment will often disrupt the dynamics of native species (Hengeveld, 1989; Shigesada and Kawasaki, 1997; Williamson, 1996). For example, the gray squirrel, *Sciurus carolinensis*, was introduced into Great Britain in the late nineteenth century and it quickly invaded many regions previously occupied by the native red squirrel, *Sciurus vulgaris* (Reynolds, 1985). Data were collected in various regions in Great Britain as to the presence of red squirrels only (R), gray squirrels only (G), both squirrels (B), or absence of both squirrels (A). The data were summarized and a Markov chain model was developed with the four states, \{R, G, B, A\}. The model is reported in Mooney and Swift (1999). Each region was classified as being in one of these states, and the transitions between states over a period of one year were estimated (e.g., \(p_{RR} = 0.8797\), \(p_{RG} = 0.0382\)). If the states 1, 2, 3, and 4 are ordered as R, G, B, and A, respectively, then the transition matrix has the form

\[
P = \begin{pmatrix}
0.8797 & 0.0382 & 0.0527 & 0.0008 \\
0.0212 & 0.8002 & 0.0041 & 0.0143 \\
0.0981 & 0.0273 & 0.8802 & 0.0527 \\
0.0010 & 0.1343 & 0.0630 & 0.9322 \\
\end{pmatrix}.
\]

It is easy to check that the corresponding Markov chain is irreducible, positive recurrent, and aperiodic (\(P\) is regular). There exists a unique, limiting stationary distribution \(\pi\). We calculate this distribution by finding the eigenvector of \(P\) corresponding to the eigenvalue 1,

\[
\pi = (0.1705, 0.0560, 0.3421, 0.4314)^T.
\]

Over the long term, the model predicts that 17.05% of the region will be populated by red squirrels, 5.6% by gray squirrels, 34.21% by both species, and 43.14% by neither species. The mean recurrence times, \(\mu_{ii} = 1/\pi_i\), \(i = 1, 2, 3, 4\), are given by the vector

\[
\mu = (5.865, 17.857, 2.923, 2.318)^T.
\]

For example, a region populated by red squirrels (\(R\)) may change to other states (\(G, B, \text{ or } A\)) but, on the average, it will again be populated by red squirrels after a period of about 5.9 years.

2.7.1 Mean Recurrence Time and Mean First Passage Time

A method is derived for calculating the mean recurrence times and mean first passages in irreducible finite Markov chains. Denote the matrix of
mean recurrence times and mean first passage times by

\[ M = (\mu_{ij}) = \begin{pmatrix}
\mu_{11} & \mu_{12} & \cdots & \mu_{1N} \\
\mu_{21} & \mu_{22} & \cdots & \mu_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_{N1} & \mu_{N2} & \cdots & \mu_{NN}
\end{pmatrix}. \]

Instead of calculating the matrix elements via their definitions, using \( \{ f_{ni}^{(n)} \} \) and \( \{ f_{ji}^{(n)} \} \), an alternate method is applied. A relationship between the mean recurrence and the mean first passage times is derived that defines a linear system whose solution is \( M \).

Consider what happens at the first time step. Either state \( j \) can be reached in one time step with probability \( p_{ji} \) or it takes more than one time step. If it takes more than one time step to reach \( j \), then in one step another state \( k \) is reached, \( k \neq j \), with probability \( p_{ki} \). Then the time it takes to reach state \( j \) is \( 1 + \mu_{jk} \), one time step plus the mean time it takes to reach state \( j \) from state \( k \). This relationship is given by

\[ \mu_{ji} = p_{ji} + \sum_{k=1, k\neq j}^{N} p_{ki}(1 + \mu_{jk}) = 1 + \sum_{k=1, k\neq j}^{N} p_{ki}\mu_{jk}. \] (2.17)

The relationship (2.17) assumes matrix \( P \) is irreducible; every state \( j \) can be reached from any other state \( i \).

The equations in (2.17) can be expressed in matrix form,

\[ M = E + (M - \text{diag}(M))P, \] (2.18)

where \( E \) is an \( N \times N \) matrix of ones. Since \( P \) is irreducible, the Markov chain is irreducible, which means it is positive recurrent, \( 1 \leq \mu_{ii} < \infty \), \( i = 1, 2, \ldots, N \). It follows that \( 1 \leq \mu_{ji} < \infty \) for \( j \neq i \). The system (2.18) can be written as a linear system of equations, \( N^2 \) equations and \( N^2 \) unknowns (the \( \mu_{ij} \)'s). It can be shown that the linear system has a unique solution given by the \( \mu_{ij} \)'s. Stewart (1994) discusses an iterative method based on equation (2.18) to estimate the mean recurrence times and mean first passage times.

**Example 2.23** Suppose

\[ P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \]

Then equation (2.18) can be expressed as

\[ \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} = \begin{pmatrix} 1 + \mu_{12} & 1 \\ 1 & 1 + \mu_{21} \end{pmatrix}. \]

Hence, \( \mu_{12} = 1 = \mu_{21} \) and \( \mu_{11} = 2 \) and \( \mu_{22} = 2 \). This result is obvious once we recognize that the chain is periodic of period 2. It takes two time steps
to return to states 1 or 2 and only one time step to go from state 1 to state 2 or from state 2 to state 1.

The next section illustrates a method for determining the $n$-step transition matrix $P^n$ in the case of a finite Markov chain.

### 2.8 The $n$-Step Transition Matrix

In the case of a finite Markov chain, a general form for the $n$-step transition matrix can be derived. A particularly simple form for $P^n$ can be generated if $P$ can be expressed as

$$P = UDU^{-1},$$

where $D$ is a diagonal matrix and $U$ is a nonsingular matrix. In this case, matrix $P^n$ satisfies

$$P^n = U D^n U^{-1}.$$

An important theorem in linear algebra states that $P$ can be expressed as $P = UDU^{-1}$ iff $P$ is diagonalizable iff $P$ has $n$ linearly independent eigenvectors (Ortega, 1987). Hence, it shall be assumed that $P$ has $n$ linearly independent eigenvectors.

We show how the matrices $U$ and $D$ can be formed. Assume $P$ is an $N \times N$ matrix with $N$ eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_N$. Let $x_j$ be a right eigenvector (column vector) corresponding to $\lambda_j$ and $y_j$ be a left eigenvector (column vector):

$$Px_j = \lambda_j x_j \quad \text{and} \quad y_j^T P = \lambda_j y_j^T. \quad (2.19)$$

Define $N \times N$ matrices

$$H = (x_1, x_2, \ldots, x_N) \quad \text{and} \quad K = (y_1, y_2, \ldots, y_N),$$

where the columns of $H$ are the right eigenvectors and the columns of $K$ are the left eigenvectors. These matrices are nonsingular because the vectors are linearly independent. Because of the identities in (2.19),

$$PH = HD \quad \text{and} \quad K^T P = DK^T,$$

where $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$. Thus,

$$P = HDH^{-1} \quad \text{or} \quad P = (K^T)^{-1}DK^T;$$

$$U = H \quad \text{or} \quad U = (K^T)^{-1}. \quad \text{The $n$-step transition matrix satisfies}$$

$$P^n = HD^n H^{-1} \quad \text{and} \quad P^n = (K^T)^{-1} D^n K^T. \quad (2.20)$$

The identities in (2.20) demonstrate one method that can be used to calculate $P^n$. Another method is also demonstrated below (see Bailey, 1990).
Note that
\[ y_j^T P x_i = y_j^T \lambda_i x_i = y_j^T \lambda_j x_i. \]

If \( \lambda_i \neq \lambda_j \), then \( y_j^T x_i = 0 \); the left and right eigenvectors are orthogonal. Thus, in this method, it is assumed that the eigenvalues are distinct (distinct eigenvalues imply the corresponding eigenvectors are linearly independent). Suppose \( y_j \) and \( x_j \) are chosen to be orthonormal:
\[
y_j^T x_i = \begin{cases} 
0 & i \neq j \\
1 & i = j.
\end{cases}
\]

Then \( K^T H = I \) (identity matrix) or \( K^T = H^{-1} \) and
\[
P = HDH^{-1} = HDK^T \\
= (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_N x_N)(y_1, y_2, \ldots, y_N)^T \\
= \lambda_1 x_1 y_1^T + \lambda_2 x_2 y_2^T + \cdots + \lambda_N x_N y_N^T
\]

Therefore,
\[
P = \sum_{i=1}^{N} \lambda_i x_i y_i^T.
\]

Because the matrix \( x_i y_i^T x_j y_j^T \) is the zero matrix for \( i \neq j \) and the sum \( \sum_{i=1}^{N} x_i y_i^T = H K^T = I \), it follows that \( P^2 \) can be expressed in terms of the matrices \( x_i y_i^T \):
\[
P^2 = \left( \sum_{i=1}^{N} \lambda_i x_i y_i^T \right) \left( \sum_{i=1}^{N} \lambda_i x_i y_i^T \right) = \sum_{i=1}^{N} \lambda_i^2 x_i y_i^T.
\]

In general, the \( n \)-step transition matrix satisfies
\[
P^n = \sum_{i=1}^{N} \lambda_i^n x_i y_i^T. \tag{2.21}
\]

In the case where the Markov chain is regular (or ergodic), which means it is irreducible and aperiodic, then \( P^n \) has a limiting distribution. The limiting distribution is the stationary distribution corresponding to the eigenvalue \( \lambda_1 = 1 \). In this case,
\[
\lim_{n \to \infty} P^n = x_1 y_1^T = \begin{pmatrix} \pi_1 & \pi_1 & \cdots & \pi_1 \\
\pi_2 & \pi_2 & \cdots & \pi_2 \\
\vdots & \vdots & \ddots & \vdots \\
\pi_N & \pi_N & \cdots & \pi_N
\end{pmatrix},
\]

where \( x_1 = \pi \) and \( y_1^T = (1, 1, \ldots, 1) \).

Note that both methods apply to any finite matrix with distinct eigenvalues. The two methods of computing \( P^n \), given in (2.20) and (2.21), are illustrated in the next example.
Example 2.24 Consider a Markov chain with two states \{1, 2\} and transition matrix

\[ P = \begin{pmatrix} 1 - p & p \\ p & 1 - p \end{pmatrix}, \]

where \(0 < p < 1\). Note that this is a doubly stochastic matrix and all states are positive recurrent. Thus, \(\lim_{n\to\infty} P^n p(0) = \pi = (\pi_1, \pi_1)^T\), where \(\pi_2 = \pi_1\) (see Exercise 14).

The eigenvalues and corresponding eigenvectors of \(P\) are \(\lambda_1 = 1, \lambda_2 = 1 - 2p\), \(x_1^T = (1, 1), x_2^T = (1, -1)\), \(y_1^T = (1, 1)\), and \(y_2^T = (1, -1)\). Note that \(x_1/2 = \pi\) is the stationary probability distribution and \(|\lambda_2| = |1 - 2p| < 1\).

Using the first identity in (2.20), an expression for \(P^n\) is given by

\[ P^n = H \Lambda^n H^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - 2p)^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{2}. \]

Multiplication of the three matrices above yields

\[ P^n = \frac{1}{2} \begin{pmatrix} 1 + (1 - 2p)^n & 1 - (1 - 2p)^n \\ 1 - (1 - 2p)^n & 1 + (1 - 2p)^n \end{pmatrix} \]

\[ = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{(1 - 2p)^n}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \]

For example, the probability \(P^{(n)}_{11}\) is \(1/2 + (1 - 2p)^n/2\). Note that \(P^n p(0)\) approaches the stationary probability distribution given by \(\pi = (1/2, 1/2)^T\).

For the second method (2.21), the eigenvectors are normalized. Since \(y_1^T x_1 = y_2^T x_2 = 2\), we divide by 2. Thus,

\[ P^n = \lambda_1^2 \frac{x_1 y_1^T}{2} + \lambda_2^2 \frac{x_2 y_2^T}{2} \]

\[ = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{(1 - 2p)^n}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \]

The two methods give the same expression for \(P^n\). ■

There are other methods for computing \(P^n\) (see Elaydi, 1999; Elaydi and Harris, 1998; Kwapisz, 1998). We shall discuss one additional method for computing \(P^n\), where it is not necessary that \(P\) be diagonalizable. This method is based on the Cayley-Hamilton theorem from linear algebra. Verification of this method is given in the Appendix for Chapter 2.

Suppose the characteristic polynomial of an \(N \times N\) matrix \(P\) is given by

\[ \det(\lambda I - P) = \lambda^N + a_{N-1}\lambda^{N-1} + \cdots + a_0 = 0. \]

This polynomial equation is also the characteristic polynomial of an \(N\)th-order scalar difference equation of the form

\[ x(N + n) + a_{N-1} x(N + n - 1) + \cdots + a_0 x(n) = 0. \]
To find a general formula for $P^n$, it is necessary to find $N$ linearly independent solutions to this $N$th-order scalar difference equation, $x_1(n), x_2(n), \ldots, x_N(n)$, with initial conditions

\[
\begin{align*}
  x_1(0) &= 1 \\
  x_1(1) &= 0 \\
  \vdots \\
  x_1(N - 1) &= 0 \\
  x_2(0) &= 0 \\
  x_2(1) &= 1 \\
  \vdots \\
  x_2(N - 1) &= 0 \\
  \vdots \\
  x_N(0) &= 0 \\
  x_N(1) &= 0 \\
  \vdots \\
  x_N(N - 1) &= 1
\end{align*}
\]

Then

\[P^n = x_1(n)I + x_2(n)P + \cdots + x_N(n)P^{N-1}, \quad n = 0, 1, 2, \ldots \quad (2.22)\]

**Example 2.25** The $n$th power of matrix $P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, given in Example 2.24, is computed using equation (2.22). The characteristic polynomial of $P$ is

\[\lambda^2 - (2 - 2p)\lambda + 1 - 2p = (\lambda - 1)(\lambda - 1 + 2p) = 0.\]

The second-order linear difference equation,

\[x(n + 2) - (2 - 2p)x(n + 1) + (1 - 2p)x(n) = 0,
\]

has two linearly independent solutions, 1 and $(1 - 2p)^n$, for $p \neq 0$. The general solution is a linear combination of these two solutions, $x(n) = c_1 + c_2(1 - 2p)^n$. Next, we find the constants $c_1$ and $c_2$ so that the two solutions $x_1$ and $x_2$ satisfy the required initial conditions. For the first solution, $x_1(0) = c_1 + c_2 = 1$ and $x_1(1) = c_1 + c_2(1 - 2p) = 0$. Solving for $c_1$ and $c_2$ we obtain the first solution:

\[x_1(n) = \frac{2p - 1}{2p} + \frac{(1 - 2p)^n}{2p}.
\]

For the second solution, $x_2(0) = c_1 + c_2 = 0$ and $x_2(1) = c_1 + c_2(1 - 2p) = 1$. Solving for $c_1$ and $c_2$ we obtain the second solution:

\[x_2(n) = \frac{1}{2p} - \frac{(1 - 2p)^n}{2p}.
\]

Then applying the identity (2.22),

\[P^n = x_1(n)I + x_2(n)P = \frac{1}{2} \begin{pmatrix} 1 + (1 - 2p)^n & 1 - (1 - 2p)^n \\ 1 - (1 - 2p)^n & 1 + (1 - 2p)^n \end{pmatrix}.
\]

This latter formula agrees with the one given in Example 2.24.
2.9 An Example: Genetics Inbreeding Problem

Inheritance depends on the information contained in the chromosomes that are passed down from generation to generation. Humans have two sets of chromosomes (diploid), one obtained from each parent. Certain locations along the chromosomes contain the instructions for some physical characteristic. The collections of chemicals at these locations are called genes and their locations are called loci (see, e.g., Hoppensteadt, 1975). At each locus, the gene may take one of several forms referred to as an allele.

Suppose there are only two types of alleles for a given gene, denoted \( a \) and \( A \). A diploid individual could then have one of three different combinations of alleles: \( AA \), \( Aa \), or \( aa \), known as the genotypes of the locus. The combinations \( AA \) and \( aa \) are called homozygous, whereas \( Aa \) is called heterozygous.

Bailey (1990) and Feller (1968) discuss a problem on the genetics of inbreeding and formulate a Markov chain model. We discuss this problem. Assume two individuals are randomly mated. Then, in the next generation, two of their offspring of opposite sex are randomly mated. The process of brother and sister mating or inbreeding continues each year. This process can be formulated as a finite, discrete time Markov chain whose states consist of the six mating types,

1. \( AA \times AA \), 2. \( AA \times Aa \), 3. \( Aa \times Aa \), 4. \( Aa \times aa \), 5. \( AA \times aa \), 6. \( aa \times aa \).

Suppose the parents are of type 1, \( AA \times AA \). Then the next generation of offspring from these parents will be \( AA \) individuals, so that crossing of brother and sister will give only type 1, \( p_{11} = 1 \). Now, suppose the parents are of type 2, \( AA \times Aa \). Offspring of type \( AA \times Aa \) will occur in the following proportions, \( 1/2 \times AA \) and \( 1/2 \times Aa \), so that crossing of brother and sister will give \( 1/4 \times AA \) type 1 (\( AA \times AA \)), \( 1/2 \times 2 \) type 2 (\( AA \times Aa \)), and \( 1/4 \times 3 \) (\( Aa \times Aa \)). If the parents are of type 3, \( Aa \times Aa \), offspring are in the proportions \( 1/4 \times AA \), \( 1/2 \times Aa \), and \( 1/4 \times aa \), so that brother and sister mating will give \( 1/16 \) type 1, \( 1/4 \times 2 \) type 2, \( 1/4 \times 3 \) type 3, \( 1/4 \times 4 \) type 4, \( 1/8 \times 5 \) type 5, and \( 1/16 \) type 6. Continuing in this manner, we can complete the transition probability matrix \( P \):

\[
P = \begin{pmatrix}
1 & 1/4 & 1/16 & 0 & 0 & 0 \\
0 & 1/2 & 1/4 & 0 & 0 & 0 \\
0 & 1/4 & 1/4 & 1/4 & 1 & 0 \\
0 & 0 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 1/8 & 0 & 0 & 0 \\
0 & 0 & 1/16 & 1/4 & 0 & 1
\end{pmatrix}
\]
The Markov chain is reducible and has three communicating classes: \{1\}, \{6\}, and \{2, 3, 4, 5\}. The first two classes are positive recurrent and the third class is transient. States 1 and 6 are absorbing states, \( p_{ii} = 1, \ i = 1, 6 \).

Note that

\[
P^n = \begin{pmatrix}
1 & A_n & 0 \\
0 & T^n & 0 \\
0 & B_n & 1
\end{pmatrix},
\]

where \( A_n \) and \( B_n \) are functions of \( T \), \( A \), and \( B \), \( A_n = A \sum_{i=0}^{n-1} T^i \), and \( B_n = B \sum_{i=0}^{n-1} T^i \). Thus, to determine \( P^n \), we first determine \( T^n \). Because \( T \) corresponds to a transient class, \( \lim_{n \to \infty} T^n = 0 \).

A general formula can be found for \( T^n \). The eigenvalues of \( T \) are \( \lambda_i = 1/2, 1/4, \frac{1}{4}(1 + \sqrt{5}), \frac{1}{4}(1 - \sqrt{5}) \), \( i = 1, 2, 3, 4 \). For example, applying (2.20) or (2.21),

\[
T^n = HD^nH^{-1} \quad \text{or} \quad T^n = \sum_{i=1}^{4} \lambda_i^n x_i y_i^T.
\]

In addition, it can be seen that

\[
\lim_{n \to \infty} B_n = B(I - T)^{-1} \quad \text{and} \quad \lim_{n \to \infty} A_n = A(I - T)^{-1}.
\]

Once \( T^n \) is calculated, various questions can be addressed about the dynamics of the model at the \( n \)th time step. For example, what is the probability of absorption and the proportion of heterozygotes in the population in the \( n \)th generation? Absorption into states 1 or 6 can be calculated as follows. Absorption at step \( n \) into state 1 implies that at the \( (n - 1) \)st step state 2 or 3 is entered. Then state 1 is entered at the next step. Thus, absorption into state 1 at the \( n \)th step is

\[
p_{12}^{(n-1)} p_{2i}^{(n-1)} + p_{13}^{(n-1)} p_{3i}^{(n-1)} = \frac{1}{4} p_{2i}^{(n-1)} + \frac{1}{16} p_{3i}^{(n-1)}.
\]

The values of \( p_{2i}^{(n-1)} \) and \( p_{3i}^{(n-1)} \) can be calculated from \( T^{n-1} \). Absorption into state 6 at the \( n \)th step is

\[
p_{63}^{(n-1)} p_{3i}^{(n-1)} + p_{64}^{(n-1)} p_{4i}^{(n-1)} = \frac{1}{16} p_{3i}^{(n-1)} + \frac{1}{4} p_{4i}^{(n-1)}.
\]
The proportion of heterozygous individuals, $Aa$, at the $n$th time step satisfies

$$h_n = \frac{1}{2} p_2(n) + p_3(n) + \frac{1}{2} p_4(n),$$

where $p_i(n)$ is the proportion of the population in state $i$ at time $n$. Because states 2, 3, and 4 are transient, $\lim_{n \to \infty} h_n = 0$. A simpler method of calculating $h_n$ is discussed in the Appendix for Chapter 2.

2.10 Unrestricted Random Walks in Two and Three Dimensions

The random walk model can be extended to two and three dimensions. It was shown for the unrestricted random walk in one dimension that the chain is null recurrent if and only if $p = 1/2 = q$. For two and three dimensions, it is assumed that the probability of moving in any one direction is the same. Thus, for two dimensions, the probability is $1/4$ of moving in any of the four directions: up, down, right, or left. For three dimensions, the probability is $1/6$ of moving in any of the six directions: up, down, right, left, forward, or backward. It is shown for two dimensions that the chain is null recurrent but for three dimension it is transient. These examples illustrate the distinctly different behavior between one and two dimensions and dimensions greater than two. These examples were first studied by Polya and are discussed in many textbooks (see, e.g., Bailey 1990; Karlin and Taylor, 1975; Norris, 1997; Schinazi, 1999). The verifications are quite lengthy.

The Markov chain represented by this unrestricted random walk is irreducible and periodic of period 2. Therefore, recurrence and transience can be verified by checking recurrence or transience at the origin. Let the origin be denoted as 0 and $p_{00}^{(n)}$ be the probability of returning to the origin after $n$ steps. Note that $p_{00}^{(2n)} > 0$, but $p_{00}^{(2n+1)} = 0$ for $n = 0, 1, 2, \ldots$. It is impossible to begin at the origin and return to the origin in an odd number of steps.

2.10.1 Two Dimensions

In two dimensions, for a path length of $2n$ beginning and ending at 0, if $k$ steps are taken to the right, then $k$ steps must be also taken to the left, and if $n - k$ steps are taken in the upward direction, then $n - k$ steps must be taken downward, $k + k + n - k + n - k = 2n$. There are

$$\sum_{k=0}^{n} \frac{(2n)!}{k!(n-k)!(n-k)!}$$
different paths of length $2n$ that begin and end at the origin. Each of these
paths is equally likely and has a probability of occurring equal to $(1/4)^{2n}$.
Thus,

$$
p_{00}^{(2n)} = \sum_{k=0}^{n} \frac{(2n)!}{k!(n-k)!(n-k)!} \left( \frac{1}{4} \right)^{2n}
= \frac{(2n)!}{(n!)^2} \sum_{k=0}^{n} \binom{n}{k}^2 \left( \frac{1}{4} \right)^{2n}
= \frac{(2n)!}{(n!)^2} \sum_{k=0}^{n} \binom{n}{k}^2 \left( \frac{1}{4} \right)^{2n}.
$$

It can be shown that

$$
\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}
$$

(see, e.g., Bailey, 1990). Hence, $p_{00}^{(2n)}$ can be simplified to

$$
p_{00}^{(2n)} = \frac{(2n)!}{(n!)^2} \binom{2n}{n} \left( \frac{1}{4} \right)^{2n} = \left[ \frac{(2n)!}{n!n!} \right]^2 \frac{1}{4^{2n}}.
$$

Stirling’s formula can be applied to the right side of the above expression
\((n! \sim n^n \sqrt{2\pi ne^{-n}})\) so that

$$
p_{00}^{(2n)} \sim \left[ \frac{(2n)^{2n} \sqrt{4\pi ne^{-2n}}}{n^{2n}2\pi ne^{-2n}} \right]^2 \frac{1}{4^{2n}}
= \left[ \frac{4^n}{\sqrt{\pi n}} \right]^2 \frac{1}{4^{2n}} = \frac{1}{\pi n}.
$$

By comparing $\sum p_{00}^{(2n)}$ with the divergent harmonic series $\sum 1/[\pi n]$, it follows that $\sum p_{00}^{(2n)}$ also diverges. Thus, by Theorem 2.2, the origin is recurrent and all states must be recurrent. In addition, by applying the basic
limit theorem for periodic Markov chains, $\lim_{n \to \infty} p_{00}^{(2n)} = 2/\mu_{00}$. But this limit is zero; thus, $\mu_{00} = \infty$. The zero state is null recurrent and, hence, the Markov chain for the symmetric, two-dimensional random walk is null recurrent.

2.10.2 Three Dimensions

In three dimensions, in a path of length $2n$ beginning and ending at the
origin, if $k$ steps are taken to the right, then $k$ must be taken to the left;
if $j$ steps are taken upward, then $j$ steps must be taken downward; and
if $n - k - j$ steps are taken forward, then $n - k - j$ steps must be taken
backward, \( k + k + j + j + n - k - j + n - k - j = 2n \). The total number of paths of length \( 2n \) is

\[
\sum_{j + k \leq n} \frac{(2n)!}{(k!)^2(j!)^2[(n - k - j)!]^2},
\]

where the sum is over all of the \( j \) and \( k \), \( j + k \leq n \). Because each path has probability \((1/6)^{2n}\), it follows that

\[
P_{00}^{(2n)} = \sum_{j + k \leq n} \frac{(2n)!}{(k!)^2(j!)^2[(n - k - j)!]^2} \left( \frac{1}{6} \right)^{2n}
\]

\[
= \frac{(2n)!}{2^{2n}(n!)^2} \sum_{j + k \leq n} \left( \frac{n!}{j!k!(n - j - k)!} \right)^2 \left( \frac{1}{3} \right)^{2n}.
\]

We use the fact that the trinomial distribution satisfies

\[
\sum_{j + k \leq n} \frac{n!}{j!k!(n - j - k)!} \frac{1}{3^n} = 1.
\]

For convenience, denote the trinomial coefficient as

\[
\frac{n!}{j!k!(n - j - k)!} = \binom{n}{j, k}.
\]

The maximum value of the trinomial distribution can be shown to occur when \( j \approx n/3 \) and \( k \approx n/3 \) and is approximately equal to \( M_n \approx n!(1/3)^n /[\{n/3\}]^3 \) when \( n \) is large. This can be seen as follows. Suppose the maximum value occurs at \( j' \) and \( k' \). Then

\[
\binom{n}{j', k'} \leq \binom{n}{j', k'+1} \leq \binom{n}{j', k+1} \leq \binom{n}{j'+1, k'} \leq \binom{n}{j'+1, k+1}
\]

so that

\[
n - k' - 1 \leq 2j' \leq n - k' + 1, \quad n - j' - 1 \leq 2k' \leq n - j' + 1
\]

or

\[
\frac{n - 1}{n} \leq \frac{2j' + k'}{n} \leq \frac{n + 1}{n}, \quad \frac{n - 1}{n} \leq \frac{2k' + j'}{n} \leq \frac{n - 1}{n}.
\]
Letting $n \to \infty$, then $2j' + k' \sim n$ and $2k' + j' \sim n$, from which it follows that $j' \sim n/3$ and $k' \sim n/3$.

We use the above facts to get an upper bound on $p_{00}^{(2n)}$. First,

$$
p_{00}^{(2n)} \leq \frac{1}{2^{2n} (n!)^2} M_n \left[ \sum_{j+k \leq n} \frac{n!}{j!k!(n-j-k)!} \frac{1}{3^n} \right]
$$

$$
= \frac{1}{2^{2n} (n!)^2} \frac{(2n)!}{[(n/3)!]^3} \frac{1}{3^n},
$$

because the expression in the square brackets is a trinomial distribution whose sum equals one. Next, Stirling's formula is used to approximate the right-hand side of the above inequality for large $n$:

$$
\frac{1}{2^{2n} (n!)^2} \frac{(2n)!}{[(n/3)!]^3} \frac{1}{3^n} \sim \frac{1}{2^{2n}} \frac{(2n)^{2n} \sqrt{4\pi n} e^{-2n}}{n^n \sqrt{2\pi n} e^{-n} (n/3)^n (\sqrt{2\pi n/3})^3 e^{-n/3}} \frac{1}{3^n}
$$

$$
= \frac{c}{n^{3/2}},
$$

where $c = (1/2)(3/\pi)^{3/2}$. Thus, for large $n$, $p_{00}^{(2n)} \leq c/n^{3/2}$. Because $\sum_n c/n^{3/2}$ is a convergent $p$-series, it follows by comparison and Theorem 2.2 that the origin is a transient state. Hence, because the Markov chain is irreducible, all states are transient. The discrete time Markov chain for the symmetric, three-dimensional random walk is transient.

The distinctly different behavior of discrete time Markov chains in three dimensions as opposed to one or two dimensions is not unusual. A path along a line or in a plane is much more restricted than a path in space. This difference in behavior is demonstrated in other models as well (e.g., systems of autonomous differential equations), where the behavior in three or higher dimensions is much more complicated and harder to predict than in one or two dimensions.

### 2.11 Exercises for Chapter 2

1. Suppose $P$ is an $N \times N$ stochastic matrix (column sums equal one),

$$
P = \begin{pmatrix}
    p_{11} & p_{12} & \cdots & p_{1N} \\
    p_{21} & p_{22} & \cdots & p_{2N} \\
    \vdots & \vdots & \ddots & \vdots \\
    p_{N1} & p_{N2} & \cdots & p_{NN}
\end{pmatrix},
$$

(a) Show that $P^2$ is a stochastic matrix. Then show that $P^n$ is a stochastic matrix for all positive integers $n$. 


2.11. Exercises for Chapter 2

(b) Suppose $P$ is a doubly stochastic matrix (row and column sums equal one). Show that $P^n$ is a doubly stochastic matrix for all positive integers $n$.

2. Show that the relation (2.3) follows from conditional probabilities. In particular, show that

$$\text{Prob}\{A \cap B|C\} = \text{Prob}\{A|B \cap C\} \text{Prob}\{B|C\}.$$ 

3. If $j$ is a transient state of a Markov chain with states $\{1, 2, \ldots\}$, prove that for all states $i = 1, 2, \ldots$,

$$\sum_{n=1}^{\infty} p_{ji}^{(n)} < \infty$$

and $\lim_{n \to \infty} p_{ji}^{(n)} = 0$. [Hint: Use the identity $P_{ji}(s) = F_{ji}(s)P_{jj}(s)$ when $s$ equals 1.]

4. Suppose a finite Markov chain has $N$ states. State 1 is absorbing and the remaining states are transient. Use Exercises 3 and 1 to show that

$$\lim_{n \to \infty} P^n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$ 

Then for any initial probability distribution corresponding to $X_0$, $p(0) = (p_1(0), p_2(0), \ldots, p_N(0))^T$, it follows that

$$\lim_{n \to \infty} P^n p(0) = (1, 0, \ldots, 0)^T.$$ 

5. Verify the following two statements.

(a) Assume the period of state $i$ in a discrete time Markov chain model satisfies $d(i) = 0$. Then the set $\{i\}$ is a communication class in the Markov chain.

(b) In an irreducible, discrete time Markov chain, the period $d \geq 1$.

6. Refer to Example 2.9. Show that the mean recurrence times for this example are finite, $\mu_{ii} < \infty$ for $i = 1, 2$.

7. A Markov chain has the following transition matrix:

$$P = \begin{pmatrix} 0 & 1/2 & 0 \\ 1 & 0 & 1 \\ 0 & 1/2 & 0 \end{pmatrix}.$$
(a) Draw a directed graph for the chain.
(b) Identify the communicating classes and classify them as periodic or aperiodic, transient or recurrent.
(c) Calculate the probability of the first return to state \( i \) at the \( n \)th step, \( f_{ii}^{(n)} \), for each state \( i = 1, 2, 3 \) and for each time step \( n = 1, 2, \ldots \).
(d) Use (c) to calculate the mean recurrence times for each state, \( \mu_{ii}, i = 1, 2, 3 \).

8. Three different Markov chains are defined by the following transition matrices:

(i) \[
\begin{pmatrix}
1 & 0 & 1/2 \\
0 & 0 & 1/2 \\
0 & 1 & 0 \\
\end{pmatrix}
\]

(ii) \[
\begin{pmatrix}
1 & 0 & 1/3 \\
0 & 0 & 1/3 \\
0 & 1 & 1/3 \\
\end{pmatrix}
\]

(iii) \[
\begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
\end{pmatrix}
\]

(a) Draw a directed graph for each chain. Is the Markov chain irreducible?
(b) Identify the communicating classes and classify them as periodic or aperiodic, transient or recurrent.

9. Three different Markov chains are defined by the following transition matrices:

(i) \[
\begin{pmatrix}
1/3 & 1/4 & 0 & 1/2 \\
1/3 & 1/4 & 0 & 0 \\
0 & 1/4 & 1 & 0 \\
1/3 & 1/4 & 0 & 1/2 \\
\end{pmatrix}
\]

(ii) \[
\begin{pmatrix}
1/2 & 1/3 & 0 & 0 & 1 \\
0 & 0 & 1/3 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 \\
0 & 1/3 & 1/3 & 1 & 0 \\
1/2 & 0 & 1/3 & 0 & 0 \\
\end{pmatrix}
\]

(iii) \[
\begin{pmatrix}
1/3 & 1/3 & 1/3 \\
1/3 & 2/3 & 0 \\
1/3 & 0 & 2/3 \\
\end{pmatrix}
\]

(a) Draw a directed graph for each chain. Is the Markov chain irreducible?
(b) Identify the communicating classes and classify them as periodic or aperiodic, transient or recurrent.

10. Suppose the states of three different Markov chains are \( \{1, 2, \ldots \} \) and their corresponding transition matrices are

\[
P_1 = \begin{pmatrix}
a_1 & 0 & 0 & \cdots \\
a_2 & a_1 & 0 & \cdots \\
a_3 & a_2 & a_1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{pmatrix},
\]

\[
P_2 = \begin{pmatrix}
0 & 0 & 0 & \cdots \\
a_1 & 0 & 0 & \cdots \\
a_2 & a_1 & 0 & \cdots \\
a_3 & a_2 & a_1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{pmatrix}
\]
and

\[ P_3 = \begin{pmatrix}
1 & 1/2 & 1/3 & \cdots \\
0 & 1/2 & 1/3 & \cdots \\
0 & 0 & 1/3 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}. \]

The elements \( a_i \) of \( P_1 \) and \( P_2 \) are positive and \( \sum_{i=1}^{\infty} a_i = 1. \)

(a) Draw a directed graph for each chain. Is the Markov chain irreducible?

(b) Identify the communicating classes, find their period, then classify them as transient, null recurrent, or positive recurrent.

11. Assume \( i \leftrightarrow j \) for states \( i \) and \( j \) in a Markov chain. Prove the following: State \( i \) is positive recurrent if and only if state \( j \) is positive recurrent. Show that the same result holds if positive recurrent is replaced by null recurrent. \( (\text{Hint: Apply the basic limit theorems for periodic and aperiodic Markov chains and use the relation } p_{jj}^{(m+n)} \geq p_{ji}^{(m)} p_{ij}^{(n)} > 0 \text{ for some } m \text{ and } n.) \)

12. The transition matrix for a three-state Markov chain is

\[ P = \begin{pmatrix}
1 & q & 0 \\
0 & r & q \\
0 & p & p + r
\end{pmatrix}, \]

\( p, q > 0, r \geq 0, \) and \( p + q + r = 1. \)

(a) Draw the directed graph of the chain.

(b) Is the set \( \{2, 3\} \) closed? Why or why not?

(c) Find an expression for \( p_{11}^{(n)} \). Then verify that state \( 1 \) is positive recurrent.

(d) Show that the process has a unique stationary probability distribution, \( \pi = (\pi_1, \pi_2, \pi_3)^T \).

13. The transition matrix for a four-state Markov chain is

\[ P = \begin{pmatrix}
0 & 1/4 & 0 & 1/2 \\
1/2 & 0 & 3/4 & 0 \\
0 & 3/4 & 0 & 1/2 \\
1/2 & 0 & 1/4 & 0
\end{pmatrix}. \]

(a) Show that the chain is irreducible, positive recurrent, and periodic. What is the period?

(b) Find the unique stationary probability distribution.
14. Suppose the transition matrix $P$ of a finite Markov chain is doubly stochastic; that is, row and column sums equal one, $p_{ij} \geq 0$,

$$
\sum_{i=1}^{N} p_{ij} = 1, \quad \text{and} \quad \sum_{j=1}^{N} p_{ij} = 1.
$$

Prove the following: If an irreducible, aperiodic finite Markov chain (ergodic chain) has a doubly stochastic transition matrix, then all stationary probabilities are equal, $\pi_1 = \pi_2 = \cdots = \pi_N$.

15. The transition matrix for a three-state Markov chain is

$$
P = \begin{pmatrix}
0 & 0 & 1/2 \\
0 & 0 & 1/2 \\
1 & 1 & 0
\end{pmatrix}.
$$

(a) Draw a directed graph of the chain and show that $P$ is irreducible.

(b) Show that $P$ is periodic of period 2 and find $P^{2n}$, $n = 1, 2, \ldots$.

(c) Use the identity (2.18) to find the mean recurrence times and mean first passage times. Show that the mean recurrence times agree with the formula given in the basic limit theorem for periodic Markov chains.

16. Suppose that the transition matrix of a two-state Markov chain is

$$
P = \begin{pmatrix}
1 - a & b \\
a & 1 - b
\end{pmatrix},
$$

(2.23)

where $0 < a < 1$ and $0 < b < 1$. Use the identity (2.18) to find a general formula for the mean recurrence times and mean first passage times.

17. Suppose that the transition matrix of a two-state Markov chain is given by equation (2.23) in Exercise 16. Use the identity $P^n = UD^nU^{-1}$ to show that $P^n$ can be expressed as follows:

$$
P^n = \frac{1}{a+b} \begin{pmatrix} b & b \\ a & a \end{pmatrix} + \frac{(1-a-b)^n}{a+b} \begin{pmatrix} a & -b \\ -a & b \end{pmatrix}.
$$

18. Let $P = \begin{pmatrix} 1 & 1/4 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 1/4 & 1/2 \end{pmatrix}$. A general formula for $P^n$ will be derived using the method of Example 2.25.
(a) Show that the characteristic polynomial of $P$ is $\lambda^3 - 2\lambda^2 + (5/4)\lambda - 1/4 = (\lambda - 1)(\lambda - 1/2)^2 = 0$. Therefore, three linearly independent solutions of the third order linear difference equation, $x(n+3) - 2x(n+2) + (5/4)x(n+1) - (1/4)x(n) = 0$, are $y_1(n) = 1$, $y_2(n) = 1/2^n$, and $y_3(n) = n/2^n$.

(b) Use the three linearly independent solutions, $y_i(n)$, $i = 1, 2, 3$, to find three solutions $x_i(n)$, $i = 1, 2, 3$ that satisfy the initial conditions.

(c) Use the identity (2.22) to find a general expression for $P^n$.

19. Suppose that two unbiased coins are tossed repeatedly and after each toss the accumulated number of heads and tails that have appeared on each coin is recorded. Let the random variable $X_n$ denote the difference in the accumulated number of heads on coin 1 and coin 2 after the $n$th toss [e.g., (Total # Heads Coin 1) − (Total # Heads Coin 2)]. Thus, the state space is $\{0, \pm 1, \pm 2, \ldots\}$. Show that the zero state, where the total number of heads is equal on each coin, is null recurrent. Hint: Show that

$$P_{00}^{(n)} = \frac{1}{2^{2n}} \sum_{k=0}^{n} \binom{n}{k}^2 \sim \frac{1}{\sqrt{n\pi}}$$

(Bailey, 1990).

20. Consider the genetics inbreeding problem. Let

$$p(0) = (0, 1/4, 1/4, 1/4, 1/4, 0)^T.$$

(a) Find a general formula for the proportion of heterozygotes $h_n$ in terms of the eigenvalues:

$$h_n = a\lambda_1^n + b\lambda_2^n + c\lambda_3^n + d\lambda_4^n.$$

(b) Find $h_{20}$ and $h_{40}$ (see the Appendix).

21. A Markov chain model for the growth and replacement of trees assumes that there are three stages of growth based on the size of the tree: young tree, mature tree, and old tree. When an old tree dies, it is replaced by a young tree with probability $1 - p$. Order the states numerically, 1, 2, and 3. State 1 is a young tree, state 2 is a mature tree, and state 3 is an old tree. A Markov chain model for the transitions between each of the states over a period of eight years has the following transition matrix:

$$P = \begin{pmatrix}
1/4 & 0 & 1 - p \\
3/4 & 1/2 & 0 \\
0 & 1/2 & p
\end{pmatrix}.$$
Transitions occur over an eight-year period. For example, after a period of eight years the probability that a young tree becomes a mature tree is $3/4$ and the probability it remains a young tree is $1/4$.

(a) Suppose $0 \leq p < 1$. Show that the Markov chain is irreducible and aperiodic. Find the unique limiting stationary distribution.

(b) Suppose $p = 7/10$. Find the mean recurrence time for $i = 1, 2, 3$ (i.e., the mean number of eight-year periods it will take a tree in stage $i$ to be replaced by another tree of stage $i$).

22. A Markov chain model for the growth and replacement of trees assumes that there are four stages of growth based on the size of the tree: seedling, young tree, mature tree, and old tree. When an old tree dies, it is replaced by a seedling. Order the states numerically, 1, 2, 3, and 4. State 1 is a seedling, state 2 is a young tree, and so on. A Markov chain model for the transitions between each state over a period of five years has the following transition matrix:

$$P = \begin{pmatrix}
p_{11} & 0 & 0 & 1 - p_{44} \\
1 - p_{11} & p_{22} & 0 & 0 \\
0 & 1 - p_{22} & p_{33} & 0 \\
0 & 0 & 1 - p_{33} & p_{44}
\end{pmatrix}.$$ 

Transition $p_{ii}$ is the probability that a tree remains in the same state for five years and $1 - p_{ii}$ is the probability a tree is at the next stage after five years of growth.

(a) Suppose $0 < p_{ii} < 1$ for $i = 1, 2, 3, 4$. Show that the Markov chain is irreducible and aperiodic. Find the unique limiting stationary distribution.

(b) Suppose $p_{44} = 1$ and $0 < p_{ii} < 1$ for $i = 1, 2, 3$. What do these assumptions imply about the growth and replacement of trees? Show that $\lim_{n \to \infty} p_{ii}^{(n)} = p_{ii}^\circ$. Identify the communicating classes and determine if they are transient or recurrent.

### 2.12 References for Chapter 2


2.12. References for Chapter 2


### 2.13 Appendix for Chapter 2

#### 2.13.1 Power of a Matrix

Suppose $P$ is an $N \times N$ matrix with characteristic polynomial

$$c(\lambda) = \det(\lambda I - P) = \lambda^N + a_{N-1}\lambda^{N-1} + \cdots + a_0 = 0.$$ 

Note that matrix $Z(n) = P^n$ is the unique solution to the matrix difference equation,

$$Z(N + n) + a_{N-1}Z(n) = Z(N + n - 1) + \cdots + a_0Z(n) = 0,$$  \hspace{1cm} (2.24)

with initial conditions $Z(0) = I$, $Z(1) = P$, $\ldots$, and $Z(N-1) = P^{N-1}$. This follows from the Cayley-Hamilton theorem from linear algebra that states a matrix $P$ satisfies its characteristic polynomial, $c(P) = 0$. The following theorem is due to Kwapisz (1998). It is based on a similar theorem for matrix exponentials by Leonard (1996).

**Theorem 2.8.** Suppose $x_1(n), x_2(n), \ldots, x_N(n)$ are solutions of the $N$th-order scalar difference equation,

$$x(N + n) + a_{N-1}x(N + n - 1) + \cdots + a_0x(n) = 0,$$

with initial conditions

$$\begin{align*}
x_1(0) &= 1 \\
x_1(1) &= 0 \\
\vdots \\
x_1(N-1) &= 0 \\
x_2(0) &= 0 \\
x_2(1) &= 1 \\
\vdots \\
x_2(N-1) &= 0 \\
x_N(0) &= 0 \\
x_N(1) &= 0 \\
\vdots \\
x_N(N-1) &= 1
\end{align*}$$

Then

$$P^n = x_1(n)I + x_2(n)P + \cdots + x_N(n)P^{N-1}, \quad n = 0, 1, 2, \ldots.$$ 

**Proof.** Let $Z(n) = x_1(n)I + x_2(n)P + \cdots + x_N(n)P^{N-1}$ for $n = 0, 1, 2, \ldots$. Then substitution of $Z(n)$ into the difference equation (2.24) shows that $Z(n)$ satisfies the equation. In addition, $Z(0) = I$, $Z(1) = P$, $\ldots$, and $Z(N-1) = P^{N-1}$. Because the solution of (2.24) is unique it follows that $Z(n) = P^n$ for $n = 0, 1, 2, \ldots$. 

\hfill $\Box$
2.13.2 Genetics Inbreeding Problem

In the genetics inbreeding problem, \( h_n \) is the proportion of heterozygotes at time \( n \),

\[
h_n = \frac{1}{2} p_2(n) + p_3(n) + \frac{1}{2} p_4(n),
\]

where \( p_i(n) \) is the proportion of the population in state \( i \) at time \( n \) (Bailey, 1990). The three states are elements of the matrix \( T^n \). Let \( p(0) = (p_2(0), p_3(0), p_4(0), p_5(0))^T \). Then \( T^n p(0) = p(n) = \sum_{i=1}^{4} \lambda_i^n x_i y_i^T p(0) \). It follows that the \( p_i(n) \) satisfy

\[
p_i(n) = c_{i1} \lambda_1^n + c_{i2} \lambda_2^n + c_{i3} \lambda_3^n + c_{i4} \lambda_4^n, \quad i = 2, 3, 4, 5.
\]

Hence,

\[
h_n = a \lambda_1^n + b \lambda_2^n + c \lambda_3^n + d \lambda_4^n,
\]

where \( a, b, c, d \) are combinations of the \( c_{ij} \). The coefficients \( a, b, c, d \) can be found by solving the following four linear equations (linear in \( a, b, c, d \)):

\[
h_i = a \lambda_1^i + b \lambda_2^i + c \lambda_3^i + d \lambda_4^i, \quad i = 0, 1, 2, 3.
\]

Suppose, initially, the entire population is of type 2, \( AA \times Aa \), \( p^{(0)} = (0, 1, 0, 0, 0, 0)^T \). Then \( h_0 = 1/2 \) and \( P^0 p^{(0)} = (1/4, 1/2, 1/4, 0, 0, 0)^T \), so that \( h_2 = (1/2)(1/2) + (1/2)(1/2) + (1/2)(0) = 1/2 \). By computing \( P^2 p^{(0)} \) and \( P^3 p^{(0)} \), values for \( h_2 \) and \( h_3 \) can be calculated, \( h_2 = 3/8 \) and \( h_3 = 5/16 \). The following Maple program was used to calculate \( h_{20} \) and \( h_{30} \) and a general formula was obtained for \( h_n \)

\[
h_{20} = 0.008445, \quad h_{30} = 0.001014,
\]

and

\[
h_n = (1/4 + 3\sqrt{5}/20) \lambda_3^n + (1/4 - 3\sqrt{5}/20) \lambda_4^n.
\]

```maple
> with(linalg):
> P:=matrix(6,6,[1,1/4,1/16,0,0,0,1/2,1/4,0,0,0,0,1/4,1/4,1/4,1/4,1/4,1/2,0,0,0,0,1/8,0,0,0,0,0,1/16,1/4,0,1]):
> T:=matrix(4,4,[1/2,1/4,0,0,1/4,1/4,1,0,1,0,1/4,1/2,0,0,1,8,0,0,0,0,0,0,1/16,1/4,0,1]):
> p0:=vector([0,1,0,0,0,0]):
> h0:=1/2+p0[2]+p0[3]+1/2*p0[4]:
> p:=n->evalm(P^n*p0):
> h:=n->1/2*p(n)[2]+p(n)[3]+1/2*p(n)[4]:
> evalm(p0); h0;
```

\[
[0, 1, 0, 0, 0, 0]
\]
> evalm(p(1)); h(1);

\[ \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0, 0, 0 \]

\[ \frac{1}{2} \]

> evalm(p(2)); h(2);

\[ \frac{25}{64}, \frac{5}{16}, \frac{3}{16}, 1, \frac{11}{16}, 1, \frac{1}{32}, \frac{1}{4} \]

\[ \frac{3}{8} \]

> evalm(p(3)); h(3);

\[ \frac{123}{256}, \frac{13}{64}, \frac{11}{64}, 5, \frac{64}{3}, \frac{128}{3}, \frac{11}{256} \]

\[ \frac{5}{16} \]

> ll:=[eigenvals(P)];

\[ ll := \left[ \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, -\frac{1}{4}, \sqrt{5}, 1, 1 \right] \]


> solve({f(0)=q0,f(1)=q(1),f(2)=q(2),f(3)=q(3)},{a,b,c,d});

\[ \left\{ a = 0, b = 0, c = \frac{1}{4} + \frac{3}{20} \sqrt{5}, d = \frac{1}{4} - \frac{3}{20} \sqrt{5} \right\} \]

> f(20):=evalf(subs({a = 0, b = 0, c = 1/4+3/20*sqrt(5), d = 1/4-3/20*sqrt(5)},f(20)));

\[ f(20) := 0.008445262939 \]

> f(30):=evalf(subs({a = 0, b = 0, c = 1/4+3/20*sqrt(5), d = 1/4-3/20*sqrt(5)},f(30)));

\[ f(30) := 0.001014354178 \]

> evalf(h(20));

0.008445262909

> evalf(h(30));

0.001014354173