2

MARKOV CHAINS: FIRST STEPS

> Let us finish the article and the whole book with a good example of dependent trials, which approximately can be considered as a simple chain.
>
> –Andrei Andreyevich Markov

2.1 INTRODUCTION

Consider a game with a playing board consisting of squares numbered 1–10 arranged in a circle. (Think of miniature Monopoly.) A player starts at square 1. At each turn, the player rolls a die and moves around the board by the number of spaces shown on the face of the die. The player keeps moving around and around the board according to the roll of the die. (Granted, this is not a very exciting game.)

Let X_k be the number of squares the player lands on after k moves, with $X_0 = 1$. Assume that the player successively rolls 2, 1, and 4. The first four positions are

$$(X_0, X_1, X_2, X_3) = (1, 3, 4, 8).$$

Given this information, what can be said about the player's next location X_4? Even though we know the player's full past history of moves, the only information relevant

Introduction to Stochastic Processes with R, First Edition. Robert P. Dobrow.

for predicting their future position is their most recent location X_3. Since $X_3 = 8$, then necessarily $X_4 \in \{9, 10, 1, 2, 3, 4\}$, with equal probability. Formally,

$$P(X_4 = j|X_0 = 1, X_1 = 3, X_2 = 4, X_3 = 8) = P(X_4 = j|X_3 = 8) = \frac{1}{6},$$

for $j = 9, 10, 1, 2, 3, 4$. Given the player's most recent location X_3, their future position X_4 is independent of past history X_0, X_1, X_2.

The sequence of player's locations $X_0, X_1, \ldots$ is a stochastic process called a *Markov chain*. The game illustrates the essential property of a Markov chain: the future, given the present, is independent of the past.

Markov Chain

Let $\mathcal{S}$ be a discrete set. A Markov chain is a sequence of random variables $X_0, X_1, \ldots$ taking values in $\mathcal{S}$ with the property that

$$\begin{aligned} P(X_{n+1} = j|X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i) \\ = P(X_{n+1} = j|X_n = i), \end{aligned} \tag{2.1}$$

for all $x_0, \ldots, x_{n-1}, i, j \in \mathcal{S}$, and $n \geq 0$. The set $\mathcal{S}$ is the *state space* of the Markov chain.

We often use descriptive language to describe the evolution of a Markov chain. For instance, if $X_n = i$, we say that the chain *visits* state i, or *hits* i, at *time* n.

A Markov chain is *time-homogeneous* if the probabilities in Equation (2.1) do not depend on n. That is,

$$P(X_{n+1} = j|X_n = i) = P(X_1 = j|X_0 = i), \tag{2.2}$$

for all $n \geq 0$. Unless stated otherwise, the Markov chains in this book are time-homogeneous.

Since the probabilities in Equation (2.2) only depend on i and j, they can be arranged in a matrix $\boldsymbol{P}$, whose ijth entry is $P_{ij} = P(X_1 = j|X_0 = i)$. This is the *transition matrix*, or *Markov matrix*, which contains the one-step transition probabilities of moving from state to state.

If the state space has k elements, then the transition matrix is a square $k \times k$ matrix. If the state space is countably infinite, the transition matrix is infinite.

For the simple board game Markov chain, the sample space is

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\},$$

with transition matrix

$$
\boldsymbol{P} = \begin{array}{c c}
 & \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} &
\begin{pmatrix}
0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\
0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 \\
0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 \\
0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 16 & 1/6 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}.
$$

The entries of every Markov transition matrix $\boldsymbol{P}$ are nonnegative, and each row sums to 1, as

$$\sum_j P_{ij} = \sum_j P(X_1 = j | X_0 = i) = \sum_j \frac{P(X_1 = j, X_0 = i)}{P(X_0 = i)} = \frac{P(X_0 = i)}{P(X_0 = i)} = 1,$$

for all rows i. A nonnegative matrix whose rows sum to 1 is called a *stochastic matrix*.

Stochastic Matrix

A stochastic matrix is a square matrix $\boldsymbol{P}$, which satisfies

1. $P_{ij} \geq 0$ for all i, j.
2. For each row i,

$$\sum_j P_{ij} = 1.$$

2.2 MARKOV CHAIN CORNUCOPIA

Following is a taste of the wide range of applications of Markov chains. Many of these examples are referenced throughout the book.

Example 2.1 (Heads you win) Successive coin flips are the very model of independent events. Yet in a fascinating study of how people actually flip coins, Diaconis (2007) shows that vigorously flipped coins are ever so slightly biased to come up the same way they started. “For natural flips,” Diaconis asserts, “the chance of coming up as started is about 0.51.”

In other words, successive coin flips are not independent. They can be described by a Markov chain with transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ H \\ T \end{array} \begin{array}{c} \begin{array}{cc} H & \;\;\;\; T \end{array} \\ \begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix} \end{array}.$$

■

Note on Notation

In Example 2.1, the state space is $S = \{H, T\}$, and the transition matrix is labeled with row and column identifiers. For many Markov chains, such as this, random variables take values in a discrete state space whose elements are not necessarily numbers.

For matrix entries, we can use suitable labels to identify states. For instance, for this example $P_{HH} = P_{TT} = 0.51$ and $P_{HT} = P_{TH} = 0.49$.

Example 2.2 (Poetry and dependent sequences) Andrei Andreyevich Markov, the Russian mathematician who introduced Markov chains over 100 years ago, first applied them in the analysis of the poem *Eugene Onegin* by Alexander Pushkin. In the first 20,000 letters of the poem, Markov counted (by hand!) 8,638 vowels and 11,362 consonants. He also tallied pairs of successive letters. Of the 8,638 pairs that start with vowels, 1,104 pairs are vowel–vowel. Of the 11,362 pairs that start with consonants, 3,827 are consonant–consonant. Markov treated the succession of letter types as a random sequence. The resulting transition matrix is

$$\boldsymbol{P} = \begin{array}{c} \\ \text{v} \\ \text{c} \end{array} \begin{array}{c} \begin{array}{cc} \text{vowel} & \qquad\qquad \text{consonant} \end{array} \\ \begin{pmatrix} 1,104/8,638 & 7,534/8,638 \\ 7,535/11,362 & 3,827/11,362 \end{pmatrix} \end{array} = \begin{pmatrix} 0.175 & 0.825 \\ 0.526 & 0.474 \end{pmatrix}.$$

Markov showed that the succession of letter types was not an independent sequence. For instance, if letter types were independent, the probability of two successive consonants would be $(11,362/20,000)^2 = 0.323$, whereas from Pushkin's poem the probability is $P_{cc} = 0.474$.

Markov's work was a polemic against a now obscure mathematician who argued that the law of large numbers only applied to independent sequences. Markov disproved the claim by showing that the Pushkin letter sequence was a dependent sequence for which the law of large numbers applied. ■

Example 2.3 (Chained to the weather) Some winter days in Minnesota it seems like the snow will never stop. A Minnesotan's view of winter might be described by the following transition matrix for a weather Markov chain, where r, s, and c denote rain, snow, and clear, respectively.

$$\boldsymbol{P} = \begin{array}{c} \\ r \\ s \\ c \end{array}\!\!\begin{array}{c} \begin{array}{ccc} r & s & c \end{array} \\ \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{pmatrix} \end{array}.$$

For this model, no matter what the weather today, there is always at least a 60% chance that it will snow tomorrow. ■

Example 2.4 (I.i.d. sequence) An independent and identically distributed sequence of random variables is trivially a Markov chain. Assume that $X_0, X_1, \ldots$ is an i.i.d. sequence that takes values in $\{1, \ldots, k\}$ with

$$P(X_n = j) = p_j, \text{ for } j = 1, \ldots, k, \quad \text{and} \quad n \geq 0,$$

where $p_1 + \cdots + p_k = 1$. By independence,

$$P(X_1 = j | X_0 = i) = P(X_1 = j) = p_j.$$

The transition matrix is

$$\boldsymbol{P} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ k \end{array}\!\!\begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & k \end{array} \\ \begin{pmatrix} p_1 & p_2 & \cdots & p_k \\ p_1 & p_2 & \cdots & p_k \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_k \end{pmatrix} \end{array}.$$

The matrix for an i.i.d. sequence has identical rows as the next state of the chain is independent of, and has the same distribution as, the present state. ■

Example 2.5 (Gambler's ruin) Gambler's ruin was introduced in Chapter 1. In each round of a gambling game a player either wins \$1, with probability p, or loses \$1, with probability $1 - p$. The gambler starts with \$$k$. The game stops when the player either loses all their money, or gains a total of \$$n$ ($n > k$).

The gambler's successive fortunes form a Markov chain on $\{0, 1, \ldots, n\}$ with $X_0 = k$ and transition matrix given by

$$P_{ij} = \begin{cases} p, & \text{if } j = i + 1, \ 0 < i < n, \\ 1 - p, & \text{if } j = i - 1, \ 0 < i < n, \\ 1, & \text{if } i = j = 0, \ \text{ or } i = j = n, \\ 0, & \text{otherwise.} \end{cases}$$

Here is the transition matrix with $n = 6$ and $p = 1/3$.

$$
\boldsymbol{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
2/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\
0 & 2/3 & 0 & 1/3 & 0 & 0 & 0 \\
0 & 0 & 2/3 & 0 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 2/3 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 2/3 & 0 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Gambler's ruin is an example of *simple random walk with absorbing boundaries.* Since $P_{00} = P_{nn} = 1$, when the chain reaches 0 or n, it stays there forever. ■

Example 2.6 (Wright–Fisher model) The Wright–Fisher model describes the evolution of a fixed population of k genes. Genes can be one of two types, called alleles: A or a. Let X_n denote the number of A alleles in the population at time n, where time is measured by generations. Under the model, the number of A alleles at time $n + 1$ is obtained by drawing with replacement from the gene population at time n. Thus, conditional on there being i alleles of type A at time n, the number of A alleles at time $n + 1$ has a binomial distribution with parameters k and $p = i/k$. This gives a Markov chain with transition matrix defined by

$$
P_{ij} = \binom{k}{j} \left(\frac{i}{k}\right)^j \left(1 - \frac{i}{k}\right)^{k-j}, \quad \text{for } 0 \le i, j \le k.
$$

Observe that $P_{00} = P_{kk} = 1$. As the chain progresses, the population is eventually made up of all a alleles (state 0) or all A alleles (state k). A question of interest is what is the probability that the population evolves to the all-A state? ■

Example 2.7 (Squash) Squash is a popular racket sport played by two or four players in a four-walled court. In the international scoring system, the first player to score nine points is the winner. However, if the game is tied at 8-8, the player who reaches 8 points first has two options: (i) to play to 9 points (set one) or to play to 10 points (set two). Set one means that the next player to score wins the game. Set two means that the first player to score two points wins the game. Points are only scored by the player who is serving. A player who wins a rally serves the next rally. Thus, if the game is tied 8-8, the player who is not serving decides. Should they choose set one or set two? This endgame play is modeled by a Markov chain in Broadie and Joneja (1993). The two players are called A and B and a score of xy means that A has scored x points and B has scored y points. The states of the chain are defined by the score and the server. The authors let p be the probability that A wins a rally given that A is serving, and q be the probability that A wins a rally given that B is serving. Assumptions are that p and q are constant over time and independent of the current score. Following is the transition matrix of the Markov chain for the set two options. The chain is used to solve for the optimal strategy.

$$\boldsymbol{P} = \begin{array}{cc} & \begin{array}{cccccccccc} 88B & 88A & 89B & 89A & 98B & 98A & 99B & 99A & A\text{ loses} & A\text{ wins} \end{array} \\ \begin{array}{r} 88B \\ 88A \\ 89B \\ 89A \\ 98B \\ 98A \\ 99B \\ 99A \\ A\text{ loses} \\ A\text{ wins} \end{array} & \begin{pmatrix} 0 & q & 1-q & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-p & 0 & 0 & 0 & 0 & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q & 0 & 0 & 0 & 0 & 1-q & 0 \\ 0 & 0 & 1-p & 0 & 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q & 1-q & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q & 1-q & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

■

Example 2.8 (Random walk on a graph) A *graph* is a set of vertices and a set of edges. Two vertices are *neighbors* if there is an edge joining them. The *degree* of vertex v is the number of neighbors of v. For the graph in Figure 2.1, $\deg(a) = 1$, $\deg(b) = \deg(c) = \deg(d) = 4$, $\deg(e) = 3$, and $\deg(f) = 2$.

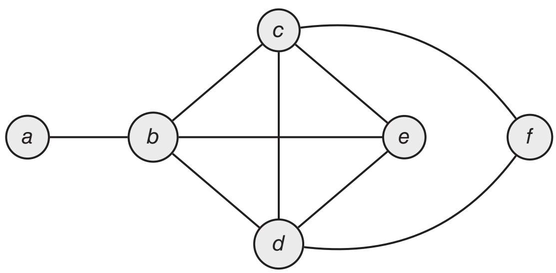

Figure 2.1 Graph on six vertices.

Imagine the vertices as lily pads on a pond. A frog is sitting on one lily pad. At each discrete unit of time, the frog hops to a neighboring lily pad chosen uniformly at random. For instance, if the frog is on lily pad c, it jumps to b, d, e, or f with probability 1/4 each. If the frog is on f, it jumps to c or d with probability 1/2 each. If the frog is on a, it always jumps to b.

Let X_n be the frog's location after n hops. The sequence $X_0, X_1, \ldots$ is a Markov chain. Given a graph G such a process is called *simple random walk on G*.

For vertices i and j, write $i \sim j$ if i and j are neighbors. The one-step transition probabilities are

$$P(X_1 = j | X_0 = i) = \begin{cases} \dfrac{1}{\deg(i)}, & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

The transition matrix for simple random walk on the graph in Figure 2.1 is

$$\boldsymbol{P} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \end{array} \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \end{pmatrix} \end{array}.$$

Of particular interest is the long-term behavior of the random walk. What can be said of the frog's position after it has been hopping for a long time?

(a) The *cycle graph* on nine vertices is shown in Figure 2.2. Simple random walk on the cycle moves left or right with probability 1/2. Each vertex has degree two. The transition matrix is defined using clock arithmetic. For a cycle with k vertices,

$$P_{ij} = \begin{cases} 1/2, & \text{if } j \equiv i \pm 1 \mod k, \\ 0, & \text{otherwise.} \end{cases}$$

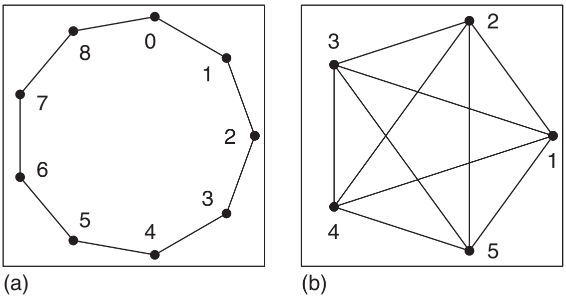

Figure 2.2 (a) Cycle graph on nine vertices. (b) Complete graph on five vertices.

(b) In the *complete graph*, every pair of vertices is joined by an edge. The complete graph on five vertices is shown in Figure 2.2. The complete graph on k vertices has $\binom{k}{2}$ edges. Each vertex has degree $k - 1$. The entries of the transition matrix are

$$P_{ij} = \begin{cases} 1/(k-1), & \text{if } i \neq j \\ 0, & \text{if } i = j. \end{cases}$$

(c) The *k-hypercube graph* has vertex set consisting of all k-element sequences of 0s and 1s. Two vertices (sequences) are connected by an edge if they differ in exactly one coordinate. The graph has 2^k vertices and $k2^{k-1}$ edges. Each vertex has degree k. See Figure 2.3.

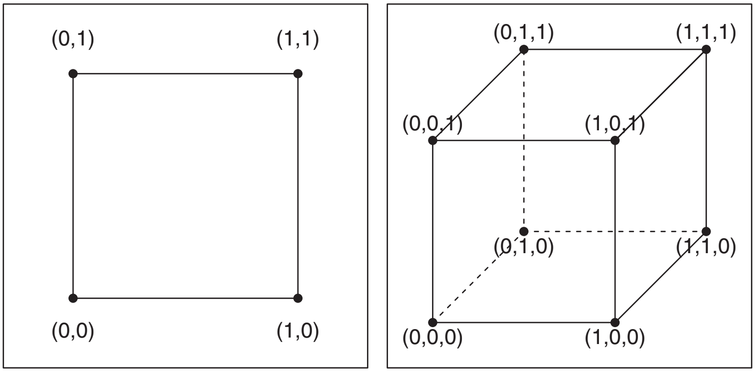

Figure 2.3 The k-hypercube graphs for $k = 2$ and $k = 3$.

Random walk on the k-hypercube can be described as follows. Assume that the walk is at a particular k-element 0–1 sequence. The next sequence of the walk is found by picking one of the k coordinates uniformly at random and *flipping the bit* at that coordinate. That is, switch from 0 to 1, or from 1 to 0. Here is the transition matrix for random walk on the 3-hypercube graph.

$$\boldsymbol{P} = \begin{array}{c} \\ 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 100 & 010 & 110 & 001 & 101 & 011 & 111 \end{array} \\ \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \end{pmatrix} \end{array}.$$

■

Example 2.9 (Birth-and-death chain) A birth-and-death Markov chain is a process with countably infinite state space and two types of transitions: *births* from i to $i + 1$ and *deaths* from i to $i - 1$. Define transition probabilities

$$P_{ij} = \begin{cases} q_i, & \text{if } j = i - 1, \\ p_i, & \text{if } j = i + 1, \\ 1 - p_i - q_i, & \text{if } j = i, \\ 0, & \text{otherwise,} \end{cases}$$

for $0 \leq p_i, q_i$ and $p_i + q_i \leq 1$. The infinite, tri-diagonal transition matrix is

$$\boldsymbol{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} 1 - p_0 & p_0 & 0 & 0 & \cdots \\ q_1 & 1 - q_1 - p_1 & p_1 & 0 & \cdots \\ 0 & q_2 & 1 - q_2 - p_2 & p_2 & \cdots \\ 0 & 0 & q_3 & 1 - q_3 - p_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

Birth-and-death chains are used to model population size, spread of disease, and the number of customers in line at the supermarket. The case when $p_i = p$ and $q_i = q$ are constant gives random walk on the non-negative integers. ■

Example 2.10 (The lazy librarian and move-to-front) A library has k books and one very long bookshelf. Each book's popularity is measured by a probability. The chance that book i will be checked out is p_i, with $p_1 + \cdots + p_k = 1$. When patrons look for books they go to the bookshelf, start at the front (left end), and scan down the bookshelf, left to right, until they find the book they want.

The library wants to organize the bookshelf in such a way as to minimize patrons' search time, measured by how many books they need to scan before they find the one they want. If the probabilities p_i are known then the best way to organize the bookshelf is by order of popularity, with the most popular, high probability books at the front. However, the actual probabilities are not known.

A lazy librarian uses the following method for organizing the books. When a patron returns a book the librarian simply puts the book at the front of the shelf. All the other books move down the shelf to the right. For instance, assume that the library has six books labeled a to f. If the bookshelf is currently ordered (b, c, a, f, e, d) and book e is chosen, the new ordering of the shelf after the book is returned is (e, b, c, a, f, d). We assume that one book is chosen and returned at a time.

Such a scheme has the advantage that over time, as books get taken out and returned, the most popular books will gravitate to the front of the bookshelf and the least popular books will gravitate to the back. The process is known as the move-to-front self-organizing scheme. Move-to-front is studied in computer science as a dynamic data structure for maintaining a linked list.

Let X_n be the order of the books after n steps. Then, the move-to-front process $X_0, X_1, \ldots$ is a Markov chain whose state space is the set of all permutations (orderings) of k books. The transition matrix has dimension $k! \times k!$. Let $\sigma = (\sigma_1, \ldots, \sigma_k)$ and $\tau = (\tau_1, \ldots, \tau_k)$ denote permutations. Then,

$$P_{\sigma,\tau} = P(X_1 = \tau | X_0 = \sigma) = p_x,$$

if $\tau_1 = x$ and τ can be obtained from σ by moving item x to the front of σ. Here is the transition matrix for move-to-front for a $k = 3$ book library.

$$\boldsymbol{P} = \begin{array}{c} \\ abc \\ acb \\ bac \\ bca \\ cab \\ cba \end{array} \begin{array}{c} \begin{array}{cccccc} abc & acb & bac & bca & cab & cba \end{array} \\ \begin{pmatrix} p_a & 0 & p_b & 0 & p_c & 0 \\ 0 & p_a & p_b & 0 & p_c & 0 \\ p_a & 0 & p_b & 0 & 0 & p_c \\ p_a & 0 & 0 & p_b & 0 & p_c \\ 0 & p_a & 0 & p_b & p_c & 0 \\ 0 & p_a & 0 & p_b & 0 & p_c \end{pmatrix} \end{array}.$$

Move-to-front is related to a card-shuffling scheme known as *random-to-top*. Given a standard deck of cards, pick a card uniformly at random and move it to the top of the deck. Random-to-top is obtained from move-to-front by letting $k = 52$

and $p_i = 1/52$, for all i. Of interest is how many such shuffles will mix up the deck of cards. ■

Example 2.11 (Weighted, directed graphs) A *weighted graph* associates a positive number (weight) with every edge. An example is shown in Figure 2.4. The graph contains *loops* at vertices b, c, and f, which are edges joining a vertex to itself.

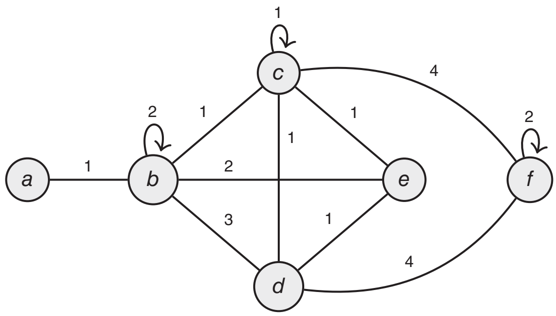

Figure 2.4 Weighted graph with loops.

For random walk on a weighted graph, transition probabilities are proportional to the sum of the weights. If vertices i and j are neighbors, let $w(i,j)$ denote the weight of the edge joining i and j. Let $w(i) = \sum_{i\sim k} w(i,k)$ be the sum of the weights on all edges joining i to its neighbors. The transition matrix is given by

$$P_{ij} = \begin{cases} \dfrac{w(i,j)}{w(i)}, & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

For the weighted graph in Figure 2.4, the transition matrix is

$$\boldsymbol{P} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \end{array} \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/9 & 2/9 & 1/9 & 3/9 & 2/9 & 0 \\ 0 & 1/8 & 1/8 & 1/8 & 1/8 & 4/8 \\ 0 & 3/9 & 1/9 & 0 & 1/9 & 4/9 \\ 0 & 2/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 4/10 & 4/10 & 0 & 2/10 \end{pmatrix} \end{array}.$$

A *directed graph* is a graph where edges have an associated direction. For every pair of vertices i and j one can have an edge from i to j and an edge from j to i. In a weighted, directed graph, there is a weight function $w(i,j)$ which gives the weight for the directed edge from i to j.

Every Markov chain can be described as a random walk on a weighted, directed graph whose vertex set is the state space of the chain. We call such a graph a *transition graph* for the Markov chain.

To create a transition graph from a transition matrix $\boldsymbol{P}$, for each pair of vertices i and j such that $P_{ij} > 0$, put a directed edge between i and j with edge weight P_{ij}.

Conversely, given a weighted, directed graph with non-negative weight function $w(i,j)$, to obtain the corresponding transition matrix let

$$P_{ij} = \frac{w(i,j)}{\sum_{i \sim k} w(i,k)} = \frac{w(i,j)}{w(i)}, \quad \text{for all } i,j.$$

Observe that matrix entries are non-negative and rows sum to 1, as for each i,

$$\sum_j P_{ij} = \sum_j \frac{w(i,j)}{w(i)} = \frac{1}{w(i)} \sum_{j \sim i} w(i,j) = \frac{w(i)}{w(i)} = 1.$$

For example, consider the transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 0.1 & 0.2 & 0.7 \\ 0.4 & 0 & 0.6 \end{pmatrix} \end{array}.$$

Two versions of the transition graph are shown in Figure 2.5. Note that multiplying the weights in the transition graph by a constant does not change the resulting transition matrix. ■

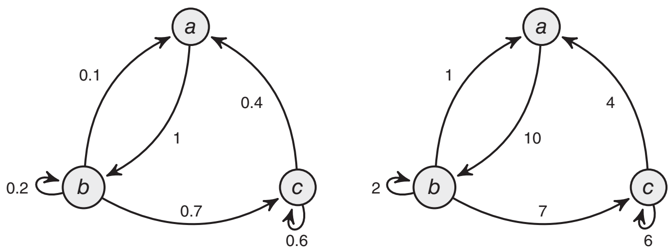

Figure 2.5 Markov transition graphs.

Example 2.12 The metastatic progression of lung cancer throughout the body is modeled in Newton et al. (2012). The 50 possible metastatic locations for cancer spread are the state space for a Markov chain model. Matrix entries were estimated from autopsy data extracted from 3,827 patients. The progress of the disease is observed as a random walk on the weighted, directed graph in Figure 2.6. Site 23 represents the lung.

An important quantity associated with this model is the *mean first-passage time*, the average number of steps it takes to move from the lung to each of the other locations in the body. ■

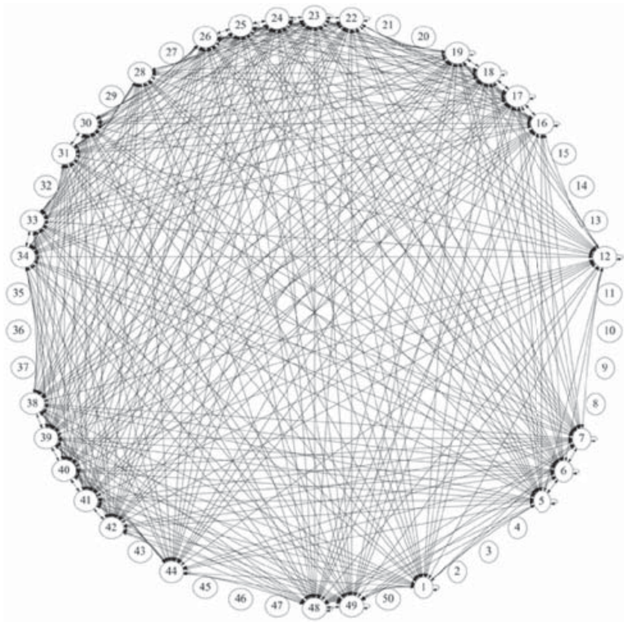

Figure 2.6 Lung cancer network as a weighted, directed graph (weights not shown). *Source:* Newton et al. (2012).

2.3 BASIC COMPUTATIONS

A powerful feature of Markov chains is the ability to use matrix algebra for computing probabilities. To use matrix methods, we consider probability distributions as vectors.

A *probability vector* is a row vector of non-negative numbers that sum to 1. Bold Greek letters, such as $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\pi}$, are used to denote such vectors.

Assume that X is a discrete random variable with $P(X = j) = \alpha_j$, for $j = 1, 2, \ldots$. Then, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots)$ is a probability vector. We say that *the distribution of* X is $\boldsymbol{\alpha}$. For matrix computations we will identify discrete probability distributions with row vectors.

For a Markov chain $X_0, X_1, \ldots$, the distribution of X_0 is called the *initial distribution* of the Markov chain. If $\boldsymbol{\alpha}$ is the initial distribution, then $P(X_0 = j) = \alpha_j$, for all j.

n-Step Transition Probabilities

For states i and j, and $n \geq 1$, $P(X_n = j | X_0 = i)$ is the probability that the chain started in i hits j in n steps. The n-step transition probabilities can be arranged in a matrix. The matrix whose ijth entry is $P(X_n = j | X_0 = i)$ is the *n-step transition matrix* of the Markov chain. Of course, for $n = 1$, this is just the usual transition matrix $\boldsymbol{P}$.

For $n \geq 1$, one of the central computational results for Markov chains is that the n-step transition matrix is precisely $\boldsymbol{P}^n$, the nth matrix power of $\boldsymbol{P}$. To show that

$P(X_n = j|X_0 = i) = (P^n)_{ij}$, condition on X_{n-1}, which gives

$$\begin{aligned} P(X_n = j|X_0 = i) &= \sum_k P(X_n = j|X_{n-1} = k, X_0 = i)P(X_{n-1} = k|X_0 = i) \\ &= \sum_k P(X_n = j|X_{n-1} = k)P(X_{n-1} = k|X_0 = i) \\ &= \sum_k P_{kj}P(X_{n-1} = k|X_0 = i), \end{aligned}$$

where the second equality is by the Markov property, and the third equality is by time-homogeneity.

For $n = 2$, this gives

$$P(X_2 = j|X_0 = i) = \sum_k P_{kj}P(X_1 = k|X_0 = i) = \sum_k P_{kj}P_{ik} = \left(P^2\right)_{ij}.$$

Hence, the two-step transition matrix is $\boldsymbol{P}^2$. Similarly, for $n = 3$,

$$P(X_3 = j|X_0 = i) = \sum_k P_{kj}P(X_2 = k|X_0 = i) = \sum_k P_{kj}P^2_{ik} = \left(P^3\right)_{ij}.$$

The three-step transition matrix is $\boldsymbol{P}^3$. Induction establishes the general result. (See Section 2.6 for an introduction to mathematical induction.)

n-Step Transition Matrix

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix $\boldsymbol{P}$. The matrix $\boldsymbol{P}^n$ is the n-step transition matrix of the chain. For $n \geq 0$,

$$P^n_{ij} = P(X_n = j|X_0 = i), \quad \text{for all } i, j.$$

Note that $P^n_{ij} = (P^n)_{ij}$. Do not confuse this with $(P_{ij})^n$, which is the number P_{ij} raised to the nth power. Also note that $\boldsymbol{P}^0$ is the identity matrix. That is,

$$P^0_{ij} = P(X_0 = j|X_0 = i) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Example 2.13 Consider random walk on the cycle graph consisting of five vertices $\{0, 1, 2, 3, 4\}$. Describe the six-step transition probabilities of the chain.

Solution The transition matrix is

$$
\begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\
\boldsymbol{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}. \end{array}
$$

The six-step transition matrix is

$$
\begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\
\boldsymbol{P}^6 = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 5/16 & 7/64 & 15/64 & 15/64 & 7/64 \\ 7/64 & 5/16 & 7/64 & 15/64 & 15/64 \\ 15/64 & 7/64 & 5/16 & 7/64 & 15/64 \\ 15/64 & 15/64 & 7/64 & 5/16 & 7/64 \\ 7/64 & 15/64 & 15/64 & 7/64 & 5/16 \end{pmatrix}. \end{array}
$$

This matrix reveals that starting from any vertex on the cycle, in six steps the walk either (i) visits itself, with probability 5/16, (ii) visits a neighbor, with probability $7/64 + 7/64 = 7/32$, or (iii) visits a non-neighbor, with probability $15/64 + 15/64 = 15/32$. ■

Example 2.14 For gambler's ruin, assume that the gambler's initial stake is \$3 and the gambler plays until either gaining \$8 or going bust. At each play the gambler wins \$1, with probability 0.6, or loses \$1, with probability 0.4. Find the gambler's expected fortune after four plays.

Solution The transition matrix is

$$
\begin{array}{c c} & \begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\boldsymbol{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} & \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}, \end{array}
$$

with

$$
\boldsymbol{P}^4 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.496 & 0.115 & 0 & 0.259 & 0 & 0.130 & 0 & 0 & 0 \\
0.237 & 0 & 0.288 & 0 & 0.346 & 0 & 0.130 & 0 & 0 \\
0.064 & 0.115 & 0 & 0.346 & 0 & 0.346 & 0 & 0.130 & 0 \\
0.026 & 0 & 0.154 & 0 & 0.346 & 0 & 0.346 & 0 & 0.130 \\
0 & 0.026 & 0 & 0.154 & 0 & 0.346 & 0 & 0.259 & 0.216 \\
0 & 0 & 0.026 & 0 & 0.154 & 0 & 0.288 & 0 & 0.533 \\
0 & 0 & 0 & 0.026 & 0 & 0.115 & 0 & 0.115 & 0.744 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

The gambler's expected fortune is

$$
\begin{aligned}
E(X_4|X_0 = 3) &= \sum_{j=0}^{8} jP(X_4 = j|X_0 = 3) = \sum_{j=0}^{8} jP^4_{3,j} \\
&= 0(0.064) + 1(0.115) + 3(0.346) + 5(0.346) + 7(0.130) \\
&= \$3.79.
\end{aligned}
$$

■

Chapman–Kolmogorov Relationship

For $m, n \geq 0$, the matrix identity $\boldsymbol{P}^{m+n} = \boldsymbol{P}^m\boldsymbol{P}^n$ gives

$$P^{m+n}_{ij} = \sum_k P^m_{ik}P^n_{kj}, \quad \text{for all} \quad i, j.$$

By time-homogeneity, this gives

$$
\begin{aligned}
P(X_{n+m} = j|X_0 = i) &= \sum_k P(X_m = k|X_0 = i)P(X_n = j|X_0 = k) \\
&= \sum_k P(X_m = k|X_0 = i)P(X_{m+n} = j|X_m = k).
\end{aligned}
$$

The probabilistic interpretation is that transitioning from i to j in $m + n$ steps is equivalent to transitioning from i to some state k in m steps and then moving from that state to j in the remaining n steps. This is known as the *Chapman–Kolmogorov relationship*.

Distribution of X_n

In general, a Markov chain $X_0, X_1, \ldots$ is not a sequence of identically distributed random variables. For $n \geq 1$, the marginal distribution of X_n depends on the n-step transition matrix $\boldsymbol{P}^n$, as well as the initial distribution $\boldsymbol{\alpha}$. To obtain the probability mass function of X_n, condition on the initial state X_0. For all j,

$$P(X_n = j) = \sum_i P(X_n = j|X_0 = i)P(X_0 = i) = \sum_i P^n_{ij}\alpha_i. \tag{2.3}$$

The sum in the last expression of Equation (2.3) can be interpreted in terms of matrix operations on vectors. It is the dot product of the initial probability vector $\boldsymbol{\alpha}$ with the jth column of $\boldsymbol{P}^n$. That is, it is the jth component of the vector–matrix product $\boldsymbol{\alpha P}^n$. (Remember: $\boldsymbol{\alpha}$ is a *row* vector.)

Distribution of X_n

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix $\boldsymbol{P}$ and initial distribution $\boldsymbol{\alpha}$. For all $n \geq 0$, the distribution of X_n is $\boldsymbol{\alpha P}^n$. That is,

$$P(X_n = j) = (\alpha P^n)_j, \quad \text{for all } j.$$

Example 2.15 Consider the weather chain introduced in Example 2.3. For tomorrow, the meteorologist predicts a 50% chance of snow and a 50% chance of rain. Find the probability that it will snow 2 days later.

Solution As the ordered states of the chain are rain, snow, and clear, the initial distribution is $\boldsymbol{\alpha} = (0.5, 0.5, 0)$. We have

$$\boldsymbol{P} = \begin{array}{c} \\ r \\ s \\ c \end{array}\!\!\begin{array}{c} \begin{array}{ccc} r & s & c \end{array} \\ \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{pmatrix} \end{array} \quad \text{and} \quad \boldsymbol{P}^2 = \begin{array}{c} \\ r \\ s \\ c \end{array}\!\!\begin{array}{c} \begin{array}{ccc} r & s & c \end{array} \\ \begin{pmatrix} 0.12 & 0.72 & 0.16 \\ 0.11 & 0.76 & 0.13 \\ 0.11 & 0.72 & 0.17 \end{pmatrix} \end{array}$$

This gives

$$\boldsymbol{\alpha P}^2 = (0.5, 0.5, 0)\begin{pmatrix} 0.12 & 0.72 & 0.16 \\ 0.11 & 0.76 & 0.13 \\ 0.11 & 0.72 & 0.17 \end{pmatrix} = (0.115, 0.74, 0.145).$$

The desired probability of snow is $P(X_2 = s) = (\alpha P^2)_s = 0.74$. ■

Present, Future, and Most Recent Past

The Markov property says that past and future are independent given the present. It is also true that past and future are independent, given *the most recent* past.

Markov Property

Let $X_0, X_1, \dots$ be a Markov chain. Then, for all $m < n$,

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-m-1} = i_{n-m-1}, X_{n-m} = i) \\ = P(X_{n+1} = j | X_{n-m} = i) \\ = P(X_{m+1} = j | X_0 = i) = P_{ij}^{m+1}, \end{aligned} \tag{2.4}$$

for all $i, j, i_0, \dots, i_{n-m-1}$, and $n \geq 0$.

Proof. With $m = 0$, Equation (2.4) reduces to the defining Markov relationship as stated in Equation (2.1).

Let $m = 1$. By conditioning on X_n,

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i) \\ &= \sum_k P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i, X_n = k) \\ &\quad \times P(X_n = k | X_0 = i_0, \dots, X_{n-1} = i) \\ &= \sum_k P(X_{n+1} = j | X_n = k) P(X_n = k | X_{n-1} = i) \\ &= \sum_k P(X_1 = j | X_0 = k) P(X_1 = k | X_0 = i) \\ &= \sum_k P_{kj} P_{ik} = P_{ij}^2. \end{aligned}$$

The third equality is by time-homogeneity. For general m, induction gives the result by conditioning on X_{n-m+1}. ■

Joint Distribution

The marginal distributions of a Markov chain are determined by the initial distribution $\boldsymbol{\alpha}$ and the transition matrix $\boldsymbol{P}$. However, a much stronger result is true. In fact, $\boldsymbol{\alpha}$ and $\boldsymbol{P}$ determine all the joint distributions of a Markov chain, that is, the joint distribution of any finite subset of $X_0, X_1, X_2, \dots$. In that sense, the initial distribution and transition matrix give a complete probabilistic description of a Markov chain.

To illustrate, consider an arbitrary joint probability, such as

$$P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l), \text{ for some states } i, j, k, l.$$

For the underlying event, the chain moves to i in five steps, then to j in one step, then to k in three steps, and then to l in eight steps. With initial distribution $\boldsymbol{\alpha}$, intuition suggests that

$$P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l) = (\alpha P^5)_i P_{ij} P^3_{jk} P^8_{kl}.$$

Indeed, conditional probability, the Markov property, and time-homogeneity give

$$\begin{aligned}
&P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l)\\
&\quad = P(X_{17} = l | X_5 = i, X_6 = j, X_9 = k) P(X_9 = k | X_5 = i, X_6 = j)\\
&\qquad \times P(X_6 = j | X_5 = i) P(X_5 = i)\\
&\quad = P(X_{17} = l | X_9 = k) P(X_9 = k | X_6 = j) P(X_6 = j | X_5 = i) P(X_5 = i)\\
&\quad = P(X_8 = l | X_0 = k) P(X_3 = k | X_0 = j) P(X_1 = j | X_0 = i) P(X_5 = i)\\
&\quad = P^8_{kl} P^3_{jk} P_{ij} (\alpha P^5)_i.
\end{aligned}$$

The joint probability is obtained from just the initial distribution $\boldsymbol{\alpha}$ and the transition matrix $\boldsymbol{P}$. For completeness, here is the general formula.

Joint Distribution

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix $\boldsymbol{P}$ and initial distribution $\boldsymbol{\alpha}$. For all $0 \le n_1 < n_2 < \cdots < n_{k-1} < n_k$ and states $i_1, i_2, \ldots, i_{k-1}, i_k$,

$$\begin{aligned}
&P(X_{n_1} = i_1, X_{n_2} = i_2, \ldots, X_{n_{k-1}} = i_{k-1}, X_{n_k} = i_k)\\
&\qquad = (\alpha P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} \cdots (P^{n_k - n_{k-1}})_{i_{k-1} i_k}. \qquad (2.5)
\end{aligned}$$

Example 2.16 Danny's daily lunch choices are modeled by a Markov chain with transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ \text{Burrito} \\ \text{Falafel} \\ \text{Pizza} \\ \text{Sushi} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Burrito} & \text{Falafel} & \text{Pizza} & \text{Sushi} \end{array} \\ \begin{pmatrix} 0.0 & 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.6 \\ 0.0 & 0.2 & 0.6 & 0.2 \end{pmatrix} \end{array}.$$

On Sunday, Danny chooses lunch uniformly at random. Find the probability that he chooses sushi on the following Wednesday and Friday, and pizza on Saturday.

Solution Let b, f, p, s denote Danny's lunch choices, respectively. Let X_0 denote Danny's lunch choice on Sunday. The desired probability is

$$P(X_3 = s, X_5 = s, X_6 = p) = (\alpha P^3)_s P^2_{ss} P_{sp},$$

where $\boldsymbol{\alpha} = (1/4, 1/4, 1/4, 1/4)$. We have

$$\boldsymbol{P}^2 = \begin{array}{c} \\ b \\ f \\ p \\ s \end{array}\begin{array}{c} \begin{array}{cccc} b & f & p & s \end{array} \\ \begin{pmatrix} 0.45 & 0.00 & 0.25 & 0.30 \\ 0.20 & 0.25 & 0.25 & 0.30 \\ 0.00 & 0.32 & 0.56 & 0.12 \\ 0.34 & 0.04 & 0.22 & 0.40 \end{pmatrix} \end{array}, \quad \boldsymbol{P}^3 = \begin{array}{c} \\ b \\ f \\ p \\ s \end{array}\begin{array}{c} \begin{array}{cccc} b & f & p & s \end{array} \\ \begin{pmatrix} 0.100 & 0.285 & 0.405 & 0.210 \\ 0.225 & 0.160 & 0.405 & 0.210 \\ 0.384 & 0.024 & 0.232 & 0.360 \\ 0.108 & 0.250 & 0.430 & 0.212 \end{pmatrix} \end{array}.$$

The desired probability is

$$(\alpha P^3)_s P^2_{ss} P_{sp} = (0.248)(0.40)(0.60) = 0.05952. \qquad \blacksquare$$

2.4 LONG-TERM BEHAVIOR—THE NUMERICAL EVIDENCE

In any stochastic—or deterministic—process, the long-term behavior of the system is often of interest.

The Canadian Forest Fire Weather Index is widely used as a means to estimate the risk of wildfire. The Ontario Ministry of Natural Resources uses the index to classify each day's risk of forest fire as either nil, low, moderate, high, or extreme.

Martell (1999) gathered daily fire risk data over 26 years at 15 weather stations across Ontario to construct a five-state Markov chain model for the daily changes in the index. The transition matrix from one location for the early summer subseason is

$$\boldsymbol{P} = \begin{array}{l} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array}\begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{pmatrix} 0.575 & 0.118 & 0.172 & 0.109 & 0.026 \\ 0.453 & 0.243 & 0.148 & 0.123 & 0.033 \\ 0.104 & 0.343 & 0.367 & 0.167 & 0.019 \\ 0.015 & 0.066 & 0.318 & 0.505 & 0.096 \\ 0.000 & 0.060 & 0.149 & 0.567 & 0.224 \end{pmatrix} \end{array}.$$

Of interest to forest managers is the long-term probability distribution of the daily index. Regardless of the risk on any particular day, what is the long-term likelihood of risk for a typical day in the early summer?

Consider the n-step transition matrix for several increasing values of n.

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.404 & 0.164 & 0.218 & 0.176 & 0.038 \\ 0.388 & 0.173 & 0.212 & 0.185 & 0.042 \\ 0.256 & 0.234 & 0.259 & 0.210 & 0.041 \\ 0.079 & 0.166 & 0.304 & 0.372 & 0.079 \\ 0.051 & 0.117 & 0.277 & 0.446 & 0.109 \end{pmatrix},$$

$$\boldsymbol{P}^3 = \begin{pmatrix} 0.332 & 0.176 & 0.235 & 0.211 & 0.046 \\ 0.326 & 0.175 & 0.235 & 0.216 & 0.047 \\ 0.283 & 0.192 & 0.247 & 0.229 & 0.049 \\ 0.158 & 0.183 & 0.280 & 0.312 & 0.067 \\ 0.118 & 0.165 & 0.286 & 0.353 & 0.078 \end{pmatrix},$$

$$P^5 = \begin{pmatrix} 0.282 & 0.180 & 0.248 & 0.239 & 0.051 \\ 0.279 & 0.180 & 0.248 & 0.241 & 0.052 \\ 0.273 & 0.181 & 0.250 & 0.244 & 0.052 \\ 0.235 & 0.183 & 0.259 & 0.266 & 0.057 \\ 0.217 & 0.183 & 0.264 & 0.277 & 0.060 \end{pmatrix},$$

$$P^{10} = \begin{pmatrix} 0.264 & 0.181 & 0.252 & 0.249 & 0.053 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.263 & 0.181 & 0.252 & 0.25 & 0.054 \\ 0.262 & 0.181 & 0.252 & 0.251 & 0.054 \end{pmatrix},$$

$$P^{17} = \begin{pmatrix} 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \end{pmatrix},$$

$$P^{18} = \begin{pmatrix} 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \\ 0.264 & 0.181 & 0.252 & 0.249 & 0.054 \end{pmatrix}.$$

Numerical evidence suggests that matrix powers are converging to a limit. Furthermore, the rows of that limiting matrix are all the same. The fact that the rows of $\boldsymbol{P}^{17}$ are the same means that the probability of a particular fire index after 17 days does not depend on today's level of risk. After 17 days, the effect of the initial state has worn off, and no longer affects the distribution of the fire index.

Furthermore, $\boldsymbol{P}^{18} = \boldsymbol{P}^{17}$ (at least to three decimal places). In fact, $\boldsymbol{P}^n = \boldsymbol{P}^{17}$, for $n \geq 17$. The latter is intuitive since if the probability of hitting state j in 17 steps is independent of the initial state, then the probability of hitting j in 17 or more steps is also independent of the initial state. See also Exercise 2.16.

The long-term fire index distribution taken from the common row of $\boldsymbol{P}^{17}$ is

Nil	Low	Moderate	High	Extreme
0.264	0.181	0.252	0.249	0.054

`R` : Matrix Powers

The function `matrixpower(mat,n)` computes the nth power of a square matrix `mat`, for $n = 0, 1, 2, \ldots$ The function is found in the `R` script file **utilities.R**, which includes several useful utility functions for working with Markov chains.

R : Fire Weather Index

```
> P
            Nil      Low Moderate     High  Extreme
Nil       0.575    0.118    0.172    0.109    0.026
Low       0.453    0.243    0.148    0.123    0.033
Moderate  0.104    0.343    0.367    0.167    0.019
High      0.015    0.066    0.318    0.505    0.096
Extreme   0.000    0.060    0.149    0.567    0.224
> matrixpower(P,17)
               Nil      Low Moderate     High   Extreme
Nil       0.263688 0.181273 0.251976 0.249484 0.0535768
Low       0.263688 0.181273 0.251976 0.249485 0.0535769
Moderate  0.263685 0.181273 0.251977 0.249486 0.0535772
High      0.263671 0.181273 0.251981 0.249495 0.0535789
Extreme   0.263663 0.181274 0.251982 0.249499 0.0535798
# round entries to three decimal places
> round(matrixpower(P,17),3)
            Nil      Low Moderate     High  Extreme
Nil       0.264    0.181    0.252    0.249    0.054
Low       0.264    0.181    0.252    0.249    0.054
Moderate  0.264    0.181    0.252    0.249    0.054
High      0.264    0.181    0.252    0.249    0.054
Extreme   0.264    0.181    0.252    0.249    0.054
```

Example 2.17 Changes in the distribution of wetlands in Yinchuan Plain, China are studied in Zhang et al. (2011). Wetlands are considered among the most important ecosystems on earth. A Markov model is developed to track yearly changes in wetland type. Based on imaging and satellite data from 1991, 1999, and 2006, researchers measured annual distributions of wetland type throughout the region and estimated the Markov transition matrix

$$\boldsymbol{P} = \begin{array}{l} \\ \text{River} \\ \text{Lake} \\ \text{Pond} \\ \text{Paddy} \\ \text{Non} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{River} & \text{Lake} & \text{Pond} & \text{Paddy} & \text{Non} \end{array} \\ \begin{pmatrix} 0.342 & 0.005 & 0.001 & 0.020 & 0.632 \\ 0.001 & 0.252 & 0.107 & 0.005 & 0.635 \\ 0.000 & 0.043 & 0.508 & 0.015 & 0.434 \\ 0.001 & 0.002 & 0.004 & 0.665 & 0.328 \\ 0.007 & 0.007 & 0.007 & 0.025 & 0.954 \end{pmatrix} \end{array}.$$

The state *Non* refers to nonwetland regions. Based on their model, the scientists predict that “The wetland distribution will essentially be in a steady state in Yinchuan Plain in approximately 100 years.”

With technology one checks that $\boldsymbol{P}^{100} = \boldsymbol{P}^{101}$ has identical rows. The common row gives the predicted long-term, *steady-state* wetland distribution.

River	Lake	Pond	Paddy	Non
0.010	0.010	0.015	0.068	0.897

R : Limiting Distribution for Wetlands Type

```
> P
       River  Lake   Pond   Paddy    Non
River 0.342  0.005  0.001  0.020  0.632
Lake  0.001  0.252  0.107  0.005  0.635
Pond  0.003  0.043  0.507  0.014  0.433
Paddy 0.001  0.002  0.004  0.665  0.328
Non   0.007  0.007  0.007  0.025  0.954
> matrixpower(P,100)
       River  Lake   Pond   Paddy    Non
River   0.01  0.01   0.015  0.068  0.897
Lake    0.01  0.01   0.015  0.068  0.897
Pond    0.01  0.01   0.015  0.068  0.897
Paddy   0.01  0.01   0.015  0.068  0.897
Non     0.01  0.01   0.015  0.068  0.897
```

■

Random Walk on Cycle

Assume that the hopping frog of Example 2.8 now finds itself on a 25-lily pad cycle graph. If the frog starts hopping from vertex 1, where is it likely to be after many hops?

After a small number of hops the frog will tend to be close to its starting position at the top of the pond. But after a large number of hops the frog's position will tend to be randomly distributed about the cycle, that is, it will tend to be uniformly distributed on all the vertices. See Figure 2.7 and Table 2.1, where probabilities are shown for the frog's position after n steps, for several values of n.

Vertices 24, 25, 1, 2, and 3 are closest to the frog's starting position. We consider these vertices *near the top* of the cycle. Vertices 12, 13, 14, and 15 are the furthest away. We consider these *near the bottom* of the cycle. After just 12 hops, the frog is still relatively close to the starting vertex. The chance that the frog is near the top of the cycle is 0.61. After 25 steps, the probability of being near the top is eight times greater than the probability of being near the bottom—0.32 compared with 0.04. Even after 100 steps it is still almost twice as likely that the frog will be near the starting vertex as compared to being at the opposite side of the cycle.

After 400 steps, however, the frog's position is *mixed up* throughout the cycle and is very close to being uniformly distributed on all the vertices. The dependency on the frog's initial position has worn off and all vertices are essentially equally likely. Numerical evidence suggests that the long-term distribution of the frog's position is uniform on the vertices and independent of starting state.

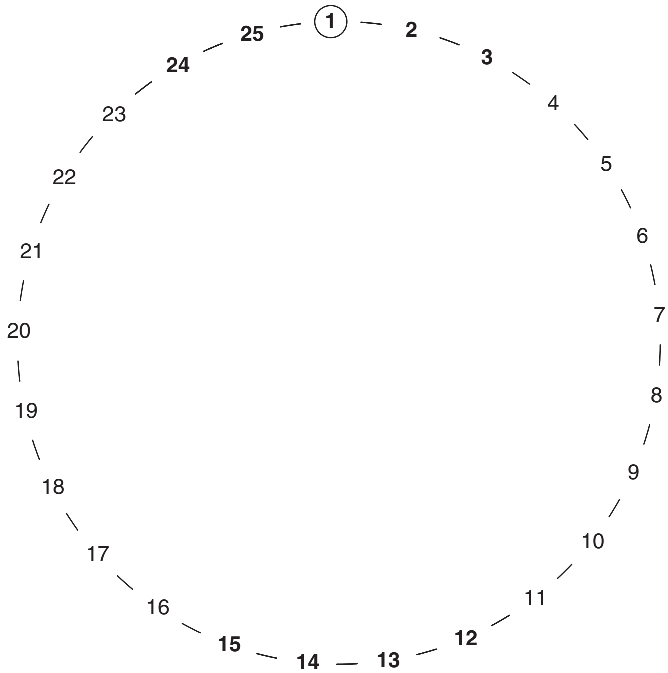

Figure 2.7 Frog starts hopping from vertex 1 of the 25-cycle graph.

TABLE 2.1 Probabilities, After *n* Steps, of the Frog's Position Near the Top and Bottom of the 25-Cycle Graph

n	24	25	1	2	3	12	13	14	15
1	0	0.5	0	0.5	0	0	0	0	0
2	0.25	0	0.5	0	0.25	0	0	0	0
6	0.23	0	0.31	0	0.23	0	0	0	0
12	0.19	0	0.23	0	0.19	0	0.00	0.00	0
25	0.00	0.16	0.00	0.16	0.0	0.01	0.01	0.01	0.01
60	0.10	0.00	0.10	0.00	0.10	0.02	0.03	0.03	0.02
100	0.08	0.01	0.08	0.01	0.08	0.03	0.04	0.04	0.03
400	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
401	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
402	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

Not all Markov chains exhibit the kind of long-term limiting behavior seen in random walk on the 25-cycle graph or in the fire index chain. Consider random walk on the cycle graph with six vertices. Here are several powers of the transition matrix.

$$
\boldsymbol{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
0 & 0.5 & 0 & 0 & 0 & 0.5 \\
0.5 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0.5 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0.5 \\
0.5 & 0 & 0 & 0 & 0.5 & 0
\end{pmatrix}
\end{array},
$$

$$
\boldsymbol{P}^2 = \begin{pmatrix}
0.5 & 0 & 0.25 & 0 & 0.25 & 0 \\
0 & 0.5 & 0 & 0.25 & 0 & 0.25 \\
0.25 & 0 & 0.5 & 0 & 0.25 & 0 \\
0 & 0.25 & 0 & 0.5 & 0 & 0.25 \\
0.25 & 0 & 0.25 & 0 & 0.5 & 0 \\
0 & 0.25 & 0 & 0.25 & 0 & 0.5
\end{pmatrix},
$$

$$
\boldsymbol{P}^3 = \begin{pmatrix}
0 & 0.375 & 0 & 0.25 & 0 & 0.375 \\
0.375 & 0 & 0.375 & 0 & 0.25 & 0 \\
0 & 0.375 & 0 & 0.375 & 0 & 0.25 \\
0.25 & 0 & 0.375 & 0 & 0.375 & 0 \\
0 & 0.25 & 0 & 0.375 & 0 & 0.375 \\
0.375 & 0 & 0.25 & 0 & 0.375 & 0
\end{pmatrix},
$$

$$
\boldsymbol{P}^8 = \begin{pmatrix}
0.336 & 0 & 0.332 & 0 & 0.332 & 0 \\
0 & 0.336 & 0 & 0.332 & 0 & 0.332 \\
0.332 & 0 & 0.336 & 0 & 0.332 & 0 \\
0 & 0.332 & 0 & 0.336 & 0 & 0.332 \\
0.332 & 0 & 0.332 & 0 & 0.336 & 0 \\
0. & 0.332 & 0 & 0.332 & 0 & 0.336
\end{pmatrix},
$$

$$
\boldsymbol{P}^{11} = \begin{pmatrix}
0 & 0.333 & 0 & 0.333 & 0 & 0.333 \\
0.336 & 0 & 0.336 & 0 & 0.333 & 0 \\
0 & 0.333 & 0 & 0.333 & 0 & 0.333 \\
0.333 & 0 & 0.333 & 0 & 0.333 & 0 \\
0 & 0.333 & 0 & 0.333 & 0 & 0.333 \\
0.333 & 0 & 0.333 & 0 & 0.333 & 0
\end{pmatrix},
$$

$$
\boldsymbol{P}^{12} = \begin{pmatrix}
0.333 & 0 & 0.333 & 0 & 0.333 & 0 \\
0 & 0.333 & 0 & 0.333 & 0 & 0.333 \\
0.333 & 0 & 0.333 & 0 & 0.333 & 0 \\
0 & 0.333 & 0 & 0.333 & 0 & 0.333 \\
0.333 & 0 & 0.333 & 0 & 0.333 & 0 \\
0. & 0.333 & 0 & 0.333 & 0 & 0.333
\end{pmatrix}.
$$

The higher-order transition matrices flip-flop for odd and even powers. If the walk starts at an even vertex, it will always be on an even vertex after an even number of steps, and on an odd vertex after an odd number of steps. The parity of the position of the walk matches the parity of the starting vertex after an even number of steps. The parity switches after an odd number of steps.

High powers of the transition matrix do not converge to a limiting matrix. The long-term behavior of the walk depends on the starting state and depends on how many steps are taken.

2.5 SIMULATION

Simulation is a powerful tool for studying Markov chains. For many chains that arise in applications, state spaces are huge and matrix methods may not be practical, or even possible, to implement.

For instance, the card-shuffling chain introduced in Example 2.10 has a state space of $k!$ elements. The transition matrix for a standard deck of cards is $52! \times 52!$, which has about 6.5×10^{135} entries.

Even for a moderately sized 50×50 matrix, as in the cancer study of Example 2.12, numerical matrix computations can be difficult to obtain. The researchers of that study found it easier to derive their results by simulation.

A Markov chain can be simulated from an initial distribution and transition matrix. To simulate a Markov sequence $X_0, X_1, \ldots,$ simulate each random variable sequentially conditional on the outcome of the previous variable. That is, first simulate X_0 according to the initial distribution. If $X_0 = i$, then simulate X_1 from the ith row of the transition matrix. If $X_1 = j$, then simulate X_2 from the jth row of the transition matrix, and so on.

Algorithm for Simulating a Markov Chain

Input: (i) initial distribution $\boldsymbol{\alpha}$, (ii) transition matrix $\boldsymbol{P}$, (iii) number of steps n.

Output: $X_0, X_1, \ldots, X_n$

Algorithm:

Generate X_0 according to $\boldsymbol{\alpha}$
FOR $i = 1, \ldots, n$
  Assume that $X_{i-1} = j$
  Set $\boldsymbol{p} = j$th row of $\boldsymbol{P}$
  Generate X_i according to $\boldsymbol{p}$
END FOR

To simulate a finite Markov chain, the algorithm is implemented in `R` by the function `markov(init,mat,n)`, which is contained in the file **utilities.R**. The arguments of the function are `init`, the initial distribution, `mat`, the transition matrix, and `n`, the number of steps to simulate. A call to `markov(init,mat,n)` generates the $(n + 1)$-element vector $(X_0, \ldots, X_n)$. The `markov` function allows for an optional fourth argument `states`, which is the state space given as a vector. If the state space has k elements, the function assigns the default value to `states` of $(1, \ldots, k)$.

Example 2.18 (Lung cancer study) Medical researchers can use simulation to study the progression of lung cancer in the body, as described in Example 2.12. The 50×50 transition matrix is stored in an Excel spreadsheet and can be downloaded into `R` from the file **cancerstudy.R**. The initial distribution is a vector of all 0s with a 1 at position 23, corresponding to the lung. See the documentation in the script file for the 50-site numbering system. Common sites are 24 and 25 (lymph nodes) and 22 (liver). Following are several simulations of the process for eight steps.

R : Simulating Lung Cancer Growth

```
> mat <- read.csv("lungcancer.csv",header=TRUE)
> init <- c(rep(0,22),1,rep(0,27))
    # all 0s with 1 at site 23 (lung)
> n <- 8
> markov(init,mat,n)
[1] 23 17 24 23 44  6  1 24 28
> markov(init,mat,n)
[1] 23 25 25 19 25  1 24 22 22
> markov(init,mat,n)
[1] 23 18 44  7 22 23 30 24 33
> markov(init,mat,n)
[1] 23 22 24 24 30 24 24 23 24
```

Newton et al. (2012) use simulation to estimate the *mean first passage time*—the number of steps, on average, it takes for cancer to pass from the lung to each other location in the body, "something a static autopsy data set cannot give us directly." The authors conclude that their study gives "important baseline quantitative insight into the structure of lung cancer progression networks." ■

Example 2.19 University administrators have developed a Markov model to simulate graduation rates at their school. Students might drop out, repeat a year, or move on to the next year. Students have a 3% chance of repeating the year. First-years and sophomores have a 6% chance of dropping out. For juniors and seniors, the drop-out rate is 4%. The transition matrix for the model is

$$
\boldsymbol{P} = \begin{array}{c} \\ \text{Drop} \\ \text{Fr} \\ \text{So} \\ \text{Jr} \\ \text{Sr} \\ \text{Grad} \end{array}
\begin{array}{c}
\begin{array}{cccccc} \text{Drop} & \text{Fr} & \text{So} & \text{Jr} & \text{Sr} & \text{Grad} \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0.06 & 0.03 & 0.91 & 0 & 0 & 0 \\
0.06 & 0 & 0.03 & 0.91 & 0 & 0 \\
0.04 & 0 & 0 & 0.03 & 0.93 & 0 \\
0.04 & 0 & 0 & 0 & 0.03 & 0.93 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Eventually, students will either drop out or graduate. See Figure 2.8 for the transition graph. To simulate the long-term probability that a new student graduates, the chain is run for 10 steps with initial distribution $\boldsymbol{\alpha} = (0, 1, 0, 0, 0, 0)$, taking X_{10} as a *long-term* sample. (With high probability, a student will either drop out or graduate by 10 years.) The simulation is repeated 10,000 times, each time keeping track of whether a student graduates or drops out. The estimated long-term probability of graduating is 0.8037.

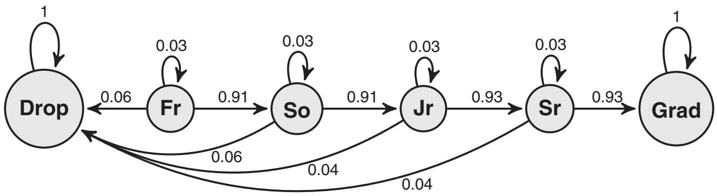

Figure 2.8 Transition graph of the graduation Markov chain.

R : Simulating Graduation, Drop-out Rate

```
# graduation.R
> init <- c(0,1,0,0,0,0) #Student starts as first-year
> P <- matrix(c(1,0,0,0,0,0,0.06,0.03,0.91,0,
  0,0,0.06,0,0.03,0.91,0,0,0.04,0,0,0.03,0.93,0,
  0.04,0,0,0,0.03,0.93,0,0,0,0,0,1),nrow=6,byrow=T)
> states <-  c("Drop","Fr","So","Jr","Se","Grad")
> rownames(P) <- states
> colnames(P) <- states
> P
     Drop   Fr   So   Jr   Se Grad
Drop 1.00 0.00 0.00 0.00 0.00 0.00
Fr   0.06 0.03 0.91 0.00 0.00 0.00
So   0.06 0.00 0.03 0.91 0.00 0.00
Jr   0.04 0.00 0.00 0.03 0.93 0.00
Se   0.04 0.00 0.00 0.00 0.03 0.93
Grad 0.00 0.00 0.00 0.00 0.00 1.00
> markov(init,P,10,states)
 [1] "Fr"   "So"   "Ju"   "Se"   "Grad" "Grad"
 [7] "Grad" "Grad" "Grad" "Grad"
> sim <- replicate(10000,markov(init,P,10,states)[11])
> table(sim)/10000
  Drop   Grad
0.1963 0.8037
```

The graduation transition matrix is small enough so that it is possible to use technology to take high matrix powers. We find that

$$\boldsymbol{P}^{20} = \boldsymbol{P}^{21} = \begin{array}{c} \\ \text{Drop} \\ \text{Fr} \\ \text{So} \\ \text{Jr} \\ \text{Sr} \\ \text{Grad} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Drop} & \text{Fr} & \text{So} & \text{Jr} & \text{Sr} & \text{Grad} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.1910 & 0 & 0 & 0 & 0 & 0.8090 \\ 0.1376 & 0 & 0 & 0 & 0 & 0.8624 \\ 0.0808 & 0 & 0 & 0 & 0 & 0.9192 \\ 0.0412 & 0 & 0 & 0 & 0 & 0.9588 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The interpretation of this limiting matrix as representing long-term probabilities shows that the probability that a first-year student eventually graduates is 0.809. The last column of the matrix gives the probability of eventually graduating for each class year.

Note that although matrix powers $\boldsymbol{P}^n$ converge to a limiting matrix, as $n \to \infty$, unlike the forest fire and frog-hopping examples, the rows of this matrix are not identical. In this case, the long-term probability of hitting a particular state depends on the initial state. ■

2.6 MATHEMATICAL INDUCTION*

Mathematical induction is a technique for proving theorems, or properties, which hold for the natural numbers $1, 2, 3, \ldots$ An example of such a theorem is that the sum of the first n positive integers is

$$1 + 2 + \cdots + n = \binom{n+1}{2} = \frac{n(n+1)}{2}.$$

An example for Markov chains is given in Section 2.3. Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix $\boldsymbol{P}$ and initial distribution $\boldsymbol{\alpha}$. Then, for all $n \geq 1$, the distribution of X_n is $\boldsymbol{\alpha P}^n$. That is,

$$P(X_n = j) = (\alpha P^n)_j = \sum_i \alpha_i P^n_{ij} \text{for all states } j.$$

Both of these results can be proven using induction.

The *principle of mathematical induction* states that (i) if a statement is true for $n = 1$, and (ii) whenever the statement is true for a natural number $n = k$, it is also true for $n = k + 1$, then the statement will be true for all natural numbers.

Proving theorems by mathematical induction is a two-step process.

First, the *base case* is established by proving the result true for $n = 1$.

Second, one assumes the result true for a given n, and then shows that it is true for $n + 1$. Assuming that the result true for a given n is called the *induction hypothesis*.

To illustrate the proof technique, consider the claim that the sum of the first n integers is equal to $n(n + 1)/2$.

For the base case, when $n = 1$, the sum of the first n integers is trivially equal to 1. And

$$n(n + 1)/2 = 1(2)/2 = 1.$$

This establishes the base case.

Assume that the property true for a given n. We need to show that the sum of the first $n+1$ integers is $(n+1)(n+2)/2$. The sum of the first $n+1$ integers is

$$\sum_{k=1}^{n+1} k = \left(\sum_{k=1}^{n} k\right) + (n+1).$$

By the induction hypothesis, the latter is equal to

$$\frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} = \frac{n^2 + 3n + 2}{2} = \frac{(n+2)(n+1)}{2}.$$

This establishes the sum formula.

For another example of an induction proof, take the Markov chain result that the distribution of X_n is $\boldsymbol{\alpha P}^n$. The base case $n = 1$ is shown in Section 2.3. The distribution of X_1 is $\boldsymbol{\alpha P}^1 = \boldsymbol{\alpha P}$. Assume the result true for a given n. For the distribution of X_{n+1}, condition on X_n. This gives

$$\begin{aligned} P(X_{n+1} = j) &= \sum_i P(X_{n+1} = j | X_n = i) P(X_n = i) \\ &= \sum_i P_{ij} P(X_n = i) \\ &= \sum_i P_{ij} (\alpha P^n)_i \\ &= (\alpha P^{n+1})_j, \end{aligned}$$

where the next-to-last equality is by the induction hypothesis.

If the reader would like more practice applying induction, see Exercise 2.22.

Figure 2.9 Andrei Andreyevich Markov (1856–1922). *Source:* Wikimedia Commons, https://commons.wikimedia.org/wiki/File:AAMarkov.jpg. Public domain.

EXERCISES

2.1 A Markov chain has transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0.1 & 0.3 & 0.6 \\ 0 & 0.4 & 0.6 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \end{array}$$

with initial distribution $\boldsymbol{\alpha} = (0.2, 0.3, 0.5)$. Find the following:

(a) $P(X_7 = 3 | X_6 = 2)$

(b) $P(X_9 = 2 | X_1 = 2, X_5 = 1, X_7 = 3)$

(c) $P(X_0 = 3 | X_1 = 1)$

(d) $E(X_2)$

2.2 Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}$$

and initial distribution $\boldsymbol{\alpha} = (1/2, 0, 1/2)$. Find the following:

(a) $P(X_2 = 1 | X_1 = 3)$

(b) $P(X_1 = 3, X_2 = 1)$

(c) $P(X_1 = 3 | X_2 = 1)$

(d) $P(X_9 = 1 | X_1 = 3, X_4 = 1, X_7 = 2)$

2.3 See Example 2.6. Consider the Wright–Fisher model with a population of $k = 3$ genes. If the population initially has one A allele, find the probability that there are no A alleles in three generations.

2.4 For the general two-state chain with transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}$$

and initial distribution $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, find the following:

(a) the two-step transition matrix

(b) the distribution of X_1

2.5 Consider a random walk on $\{0, \dots, k\}$, which moves left and right with respective probabilities q and p. If the walk is at 0 it transitions to 1 on the next step. If the walk is at k it transitions to $k-1$ on the next step. This is called *random walk with reflecting boundaries*. Assume that $k = 3$, $q = 1/4$, $p = 3/4$, and the initial distribution is uniform. For the following, use technology if needed.

(a) Exhibit the transition matrix.

(b) Find $P(X_7 = 1 | X_0 = 3, X_2 = 2, X_4 = 2)$.

(c) Find $P(X_3 = 1, X_5 = 3)$.

2.6 A tetrahedron die has four faces labeled 1, 2, 3, and 4. In repeated independent rolls of the die $R_0, R_1, \dots$, let $X_n = \max\{R_0, \dots, R_n\}$ be the maximum value after $n + 1$ rolls, for $n \geq 0$.

(a) Give an intuitive argument for why $X_0, X_1, \dots$, is a Markov chain, and exhibit the transition matrix.

(b) Find $P(X_3 \geq 3)$.

2.7 Let $X_0, X_1, \dots$ be a Markov chain with transition matrix $\boldsymbol{P}$. Let $Y_n = X_{3n}$, for $n = 0, 1, 2, \dots$. Show that $Y_0, Y_1, \dots$ is a Markov chain and exhibit its transition matrix.

2.8 Give the Markov transition matrix for random walk on the weighted graph in Figure 2.10.

2.9 Give the transition matrix for the transition graph in Figure 2.11.

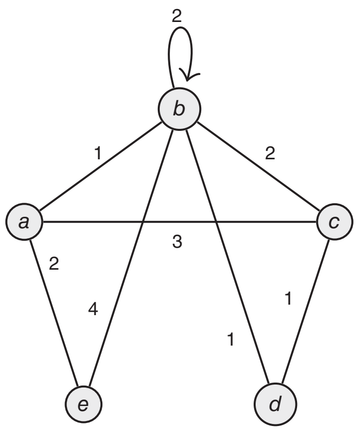

Figure 2.10

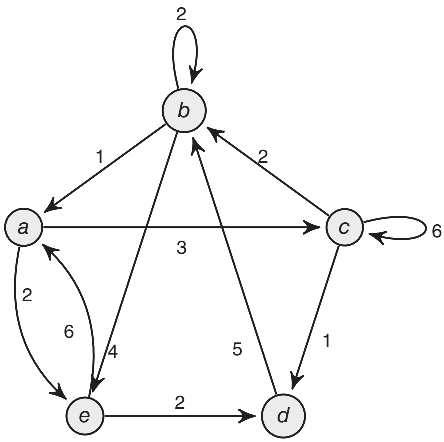

Figure 2.11

2.10 Consider a Markov chain with transition matrix

$$
\boldsymbol{P} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array}
\begin{array}{c}
\begin{array}{cccc} a & b & c & d \end{array} \\
\begin{pmatrix} 0 & 3/5 & 1/5 & 1/5 \\ 3/4 & 0 & 1/4 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 1/4 & 1/2 \end{pmatrix}
\end{array}.
$$

(a) Exhibit the directed, weighted transition graph for the chain.

(b) The transition graph for this chain can be given as a weighted graph without directed edges. Exhibit the graph.

2.11 You start with five dice. Roll all the dice and put aside those dice that come up 6. Then, roll the remaining dice, putting aside those dice that come up 6. And so on. Let X_n be the number of dice that are sixes after n rolls.

(a) Describe the transition matrix $\boldsymbol{P}$ for this Markov chain.

(b) Find the probability of getting all sixes by the third play.

(c) What do you expect $\boldsymbol{P}^{100}$ to look like? Use technology to confirm your answer.

2.12 Two urns contain k balls each. Initially, the balls in the left urn are all red and the balls in the right urn are all blue. At each step, pick a ball at random from each urn and exchange them. Let X_n be the number of blue balls in the left urn. (Note that necessarily $X_0 = 0$ and $X_1 = 1$.) Argue that the process is a Markov chain. Find the transition matrix. This model is called the Bernoulli–Laplace model of diffusion and was introduced by Daniel Bernoulli in 1769 as a model for the flow of two incompressible liquids between two containers.

2.13 See the move-to-front process in Example 2.10. Here is another way to organize the bookshelf. When a book is returned it is put back on the library shelf one position forward from where it was originally. If the book at the front of the shelf is returned it is put back at the front of the shelf. Thus, if the order of books is (a, b, c, d, e) and book d is picked, the new order is (a, b, d, c, e). This reorganization method is called the *transposition*, or *move-ahead-1*, scheme. Give the transition matrix for the transposition scheme for a shelf with three books.

2.14 There are k songs on Mary's music player. The player is set to *shuffle* mode, which plays songs uniformly at random, sampling with replacement. Thus, repeats are possible. Let X_n denote the number of *unique* songs that have been heard after the nth play.

(a) Show that $X_0, X_1, \ldots$ is a Markov chain and give the transition matrix.

(b) If Mary has four songs on her music player, find the probability that all songs are heard after six plays.

2.15 Assume that $X_0, X_1, \ldots$ is a two-state Markov chain on $\mathcal{S} = \{0, 1\}$ with transition matrix

$$\boldsymbol{P} = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}.$$

The present state of the chain only depends on the previous state. One can model a bivariate process that looks back two time periods by the following construction. Let $Z_n = (X_{n-1}, X_n)$, for $n \geq 1$. The sequence $Z_1, Z_2, \ldots$ is a Markov chain with state space $\mathcal{S} \times \mathcal{S} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Give the transition matrix of the new chain.

2.16 Assume that $\boldsymbol{P}$ is a stochastic matrix with equal rows. Show that $\boldsymbol{P}^n = \boldsymbol{P}$, for all $n \geq 1$.

2.17 Let $\boldsymbol{P}$ be a stochastic matrix. Show that $\lambda = 1$ is an eigenvalue of $\boldsymbol{P}$. What is the associated eigenvector?

2.18 A stochastic matrix is called *doubly stochastic* if its columns sum to 1. Let $X_0, X_1, \ldots$ be a Markov chain on $\{1, \ldots, k\}$ with doubly stochastic transition matrix and initial distribution that is uniform on $\{1, \ldots, k\}$. Show that the distribution of X_n is uniform on $\{1, \ldots, k\}$, for all $n \geq 0$.

2.19 Let $\boldsymbol{P}$ be the transition matrix of a Markov chain on k states. Let $\mathbf{I}$ denote the $k \times k$ identity matrix. Consider the matrix

$$\boldsymbol{Q} = (1-p)\mathbf{I} + p\boldsymbol{P}, \text{ for } 0 < p < 1.$$

Show that $\boldsymbol{Q}$ is a stochastic matrix. Give a probabilistic interpretation for the dynamics of a Markov chain governed by the $\boldsymbol{Q}$ matrix in terms of the original Markov chain.

2.20 Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ p & 1-p & 0 \end{pmatrix} \end{array},$$

for $0 < p < 1$. Let g be a function defined by

$$g(x) = \begin{cases} 0, & \text{if } x = 1, \\ 1, & \text{if } x = 2, 3. \end{cases}$$

Let $Y_n = g(X_n)$, for $n \geq 0$. Show that $Y_0, Y_1, \ldots$ is not a Markov chain.

2.21 Let P and Q be two transition matrices on the same state space. We define two processes, both started in some initial state i.
In process #1, a coin is flipped. If it lands heads, then the process unfolds according to the P matrix. If it lands tails, the process unfolds according to the Q matrix.
In process #2, at each step a coin is flipped. If it lands heads, the next state is chosen according to the P matrix. If it lands tails, the next state is chosen according to the Q matrix.
Thus, in #1, one coin is initially flipped, which governs the entire evolution of the process. And in #2, a coin is flipped at each step to decide the next step of the process.
Decide whether either of these processes is a Markov chain. If not, explain why, if yes, exhibit the transition matrix.

2.22 Prove the following using mathematical induction.
(a) $1 + 3 + 5 + \cdots + (2n - 1) = n^2$.
(b) $1^2 + 2^2 + \cdots + n^2 = n(n + 1)(2n + 1)/6$.
(c) For all real $x > -1$, $(1 + x)^n \geq 1 + nx$.

2.23 `R` : Simulate the first 20 letters (vowel/consonant) of the Pushkin poem Markov chain of Example 2.2.

2.24 `R` : Simulate 50 steps of the random walk on the graph in Figure 2.1. Repeat the simulation 10 times. How many of your simulations end at vertex c? Compare with the exact long-term probability the walk visits c.

2.25 `R` : The behavior of dolphins in the presence of tour boats in Patagonia, Argentina is studied in Dans et al. (2012). A Markov chain model is developed, with state space consisting of five primary dolphin activities (socializing, traveling, milling, feeding, and resting). The following transition matrix is obtained.

$$
\begin{array}{c}
\\ \\ \boldsymbol{P} = \\ \\ \\
\end{array}
\begin{array}{c}
 \\ s \\ t \\ m \\ f \\ r
\end{array}
\begin{array}{c}
\begin{array}{ccccc} s & t & m & f & r \end{array} \\
\begin{pmatrix}
0.84 & 0.11 & 0.01 & 0.04 & 0.00 \\
0.03 & 0.80 & 0.04 & 0.10 & 0.03 \\
0.01 & 0.15 & 0.70 & 0.07 & 0.07 \\
0.03 & 0.19 & 0.02 & 0.75 & 0.01 \\
0.03 & 0.09 & 0.05 & 0.00 & 0.83
\end{pmatrix}
\end{array}.
$$

Use technology to estimate the long-term distribution of dolphin activity.

2.26 `R` : In computer security applications, a *honeypot* is a trap set on a network to detect and counteract computer hackers. Honeypot data are studied in Kimou et al. (2010) using Markov chains. The authors obtain honeypot data from a central database and observe attacks against four computer ports—80, 135, 139, and 445—over 1 year. The ports are the states of a Markov chain along with a state corresponding to no port is attacked. Weekly data are monitored, and the port most often attacked during the week is recorded. The estimated Markov transition matrix for weekly attacks is

$$
\boldsymbol{P} =
\begin{array}{c}
 \\ 80 \\ 135 \\ 139 \\ 445 \\ \text{No}
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 80 & 135 & 139 & 445 & \text{No attack} \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 \\
0 & 8/13 & 3/13 & 1/13 & 1/13 \\
1/16 & 3/16 & 3/8 & 1/4 & 1/8 \\
0 & 1/11 & 4/11 & 5/11 & 1/11 \\
0 & 1/8 & 1/2 & 1/8 & 1/4
\end{pmatrix}
\end{array},
$$

with initial distribution $\boldsymbol{\alpha} = (0, 0, 0, 0, 1)$.

(a) Which are the least and most likely attacked ports after 2 weeks?

(b) Find the long-term distribution of attacked ports.

2.27 `R` : See **gamblersruin.R.** Simulate gambler's ruin for a gambler with initial stake \$2, playing a fair game.

(a) Estimate the probability that the gambler is ruined before he wins \$5.

(b) Construct the transition matrix for the associated Markov chain. Estimate the desired probability in (a) by taking high matrix powers.

(c) Compare your results with the exact probability.