# 3 Markov Chains: Introduction

## 3.1 Definitions

*A Markov process* $\{X_t\}$ is a stochastic process with the property that, given the value of $X_t$, the values of $X_s$ for $s > t$ are not influenced by the values of $X_u$ for $u < t$. In words, the probability of any particular future behavior of the process, when its current state is known exactly, is not altered by additional knowledge concerning its past behavior. A *discrete-time Markov chain* is a Markov process whose state space is a finite or countable set, and whose (time) index set is $T = (0, 1, 2, \ldots)$. In formal terms, the Markov property is that

$$\begin{aligned}\Pr\{X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} \\ = \Pr\{X_{n+1} = j | X_n = i\}\end{aligned} \tag{3.1}$$

for all time points $n$ and all states $i_0, \ldots, i_{n-1}, i, j$.

It is frequently convenient to label the state space of the Markov chain by the non-negative integers $\{0, 1, 2, \ldots\}$, which we will do unless the contrary is explicitly stated, and it is customary to speak of $X_n$ as being in state $i$ if $X_n = i$.

The probability of $X_{n+1}$ being in state $j$ given that $X_n$ is in state $i$ is called the *one-step transition probability* and is denoted by $P_{ij}^{n,n+1}$. That is,

$$P_{ij}^{n,n+1} = \Pr\{X_{n+1} = j | X_n = i\}. \tag{3.2}$$

The notation emphasizes that in general the transition probabilities are functions not only of the initial and final states but also of the time of transition as well. When the one-step transition probabilities are independent of the time variable $n$, we say that the Markov chain has *stationary transition probabilities*. Since the vast majority of Markov chains that we shall encounter have stationary transition probabilities, we limit our discussion to this case. Then, $P_{ij}^{n,n+1} = P_{ij}$ is independent of $n$, and $P_{ij}$ is the conditional probability that the state value undergoes a transition from $i$ to $j$ in one trial. It is customary to arrange these numbers $P_{ij}$ in a *matrix*, in the infinite square array

$$\mathbf{P} = \left\| \begin{matrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{matrix} \right\|$$

**An Introduction to Stochastic Modeling**

and refer to $\mathbf{P} = \|P_{ij}\|$ as the Markov matrix or *transition probability matrix* of the process.

The $i$th row of $\mathbf{P}$, for $i = 0, 1, \ldots,$ is the probability distribution of the values of $X_{n+1}$ under the condition that $X_n = i$. If the number of states is finite, then $\mathbf{P}$ is a finite square matrix whose order (the number of rows) is equal to the number of states. Clearly, the quantities $P_{ij}$ satisfy the conditions

$$P_{ij} \geq 0 \quad \text{for } i, j = 0, 1, 2, \ldots, \tag{3.3}$$

$$\sum_{j=0}^{\infty} P_{ij} = 1 \quad \text{for } i = 0, 1, 2, \ldots. \tag{3.4}$$

The condition (3.4) merely expresses the fact that some transition occurs at each trial. (For convenience, one says that a transition has occurred even if the state remains unchanged.)

A Markov process is completely defined once its transition probability matrix and initial state $X_0$ (or, more generally, the probability distribution of $X_0$) are specified. We shall now prove this fact.

Let $\Pr\{X_0 = i\} = p_i$. It is enough to show how to compute the quantities

$$\Pr\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\}, \tag{3.5}$$

since any probability involving $X_{j_1}, \ldots, X_{j_k}$, for $j_1 < \cdots < j_k$, can be obtained, according to the axiom of total probability, by summing terms of the form (3.5).

By the definition of conditional probabilities, we obtain

$$\begin{aligned} &\Pr\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\} \\ &\quad = \Pr\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \\ &\qquad \times \Pr\{X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\}. \end{aligned} \tag{3.6}$$

Now, by the definition of a Markov process,

$$\begin{aligned} &\Pr\{X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \\ &\quad = \Pr\{X_n = i_n | X_{n-1} = i_{n-1}\} = P_{i_{n-1}, i_n}. \end{aligned} \tag{3.7}$$

Substituting (3.7) into (3.6) gives

$$\begin{aligned} &\Pr\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} \\ &\quad = \Pr\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} P_{i_{n-1}, i_n}. \end{aligned}$$

Then, upon repeating the argument $n - 1$ additional times, (3.5) becomes

$$\begin{aligned} &\Pr\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} \\ &\quad = p_{i_0} P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}} P_{i_{n-1}, i_n}. \end{aligned} \tag{3.8}$$

This shows that all finite-dimensional probabilities are specified once the transition probabilities and initial distribution are given, and in this sense, the process is defined by these quantities.

Related computations show that (3.1) is equivalent to the Markov property in the form

$$\begin{aligned}\Pr\{X_{n+1} = j_1, \ldots, X_{n+m} = j_m | X_0 = i_0, \ldots, X_n = i_n\} \\ = \Pr\{X_{n+1} = j_1, \ldots, X_{n+m} = j_m | X_n = i_n\}\end{aligned} \tag{3.9}$$

for all time points $n, m$ and all states $i_0, \ldots, i_n, j_1, \ldots, j_m$. In other words, once (3.9) is established for the value $m = 1$, it holds for all $m \geq 1$ as well.

## Exercises

**3.1.1** A Markov chain $X_0, X_1, \ldots$ on states 0, 1, 2 has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.1 & 0.2 & 0.7 \\ 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \end{array} \right\| \end{array}$$

and initial distribution $p_0 = \Pr\{X_0 = 0\} = 0.3, p_1 = \Pr\{X_0 = 1\} = 0.4$, and $p_2 = \Pr\{X_0 = 2\} = 0.3$. Determine $\Pr\{X_0 = 0, X_1 = 1, X_2 = 2\}$.

**3.1.2** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{array} \right\| \end{array}.$$

Determine the conditional probabilities

$$\Pr\{X_2 = 1, X_3 = 1 | X_1 = 0\} \quad \text{and} \quad \Pr\{X_1 = 1, X_2 = 1 | X_0 = 0\}.$$

**3.1.3** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.4 & 0.1 & 0.5 \end{array} \right\| \end{array}.$$

If it is known that the process starts in state $X_0 = 1$, determine the probability $\Pr\{X_0 = 1, X_1 = 0, X_2 = 2\}$.

**3.1.4** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.1 & 0.1 & 0.8 \\ 0.2 & 0.2 & 0.6 \\ 0.3 & 0.3 & 0.4 \end{array} \right\| \end{array}.$$

Determine the conditional probabilities

$$\Pr\{X_1 = 1, X_2 = 1 | X_0 = 0\} \quad \text{and} \quad \Pr\{X_2 = 1, X_3 = 1 | X_1 = 0\}.$$

**3.1.5** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{array} \right\| \end{array}$$

and initial distribution $p_0 = 0.5$ and $p_1 = 0.5$. Determine the probabilities

$$\Pr\{X_0 = 1, X_1 = 1, X_2 = 0\} \quad \text{and} \quad \Pr\{X_1 = 1, X_2 = 1, X_3 = 0\}.$$

## Problems

**3.1.1** A simplified model for the spread of a disease goes this way: The total population size is $N = 5$, of which some are diseased and the remainder are healthy. During any single period of time, two people are selected at random from the population and assumed to interact. The selection is such that an encounter between any pair of individuals in the population is just as likely as between any other pair. If one of these persons is diseased and the other not, with probability $\alpha = 0.1$ the disease is transmitted to the healthy person. Otherwise, no disease transmission takes place. Let $X_n$ denote the number of diseased persons in the population at the end of the $n$th period. Specify the transition probability matrix.

**3.1.2** Consider the problem of sending a binary message, 0 or 1, through a signal channel consisting of several stages, where transmission through each stage is subject to a fixed probability of error $\alpha$. Suppose that $X_0 = 0$ is the signal that is sent and let $X_n$ be the signal that is received at the $n$th stage. Assume that $\{X_n\}$ is a Markov chain with transition probabilities $P_{00} = P_{11} = 1 - \alpha$ and $P_{01} = P_{10} = \alpha$, where $0 < \alpha < 1$.

**(a)** Determine $\Pr\{X_0 = 0, X_1 = 0, X_2 = 0\}$, the probability that no error occurs up to stage $n = 2$.

**(b)** Determine the probability that a correct signal is received at stage 2.

**Hint:** This is $\Pr\{X_0 = 0, X_1 = 0, X_2 = 0\} + \Pr\{X_0 = 0, X_1 = 1, X_2 = 0\}$.

**3.1.3** Consider a sequence of items from a production process, with each item being graded as good or defective. Suppose that a good item is followed by another good item with probability $\alpha$ and is followed by a defective item with probability $1-\alpha$. Similarly, a defective item is followed by another defective item with probability $\beta$ and is followed by a good item with probability $1-\beta$. If the first item is good, what is the probability that the first defective item to appear is the fifth item?

**3.1.4** The random variables $\xi_1, \xi_2, \ldots$ are independent and with the common probability mass function

| $k =$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\Pr\{\xi = k\} =$ | 0.1 | 0.3 | 0.2 | 0.4 |

Set $X_0 = 0$, and let $X_n = \max\{\xi_1, \ldots, \xi_n\}$ be the largest $\xi$ observed to date. Determine the transition probability matrix for the Markov chain $\{X_n\}$.

## 3.2 Transition Probability Matrices of a Markov Chain

A Markov chain is completely defined by its one-step transition probability matrix and the specification of a probability distribution on the state of the process at time 0. The analysis of a Markov chain concerns mainly the calculation of the probabilities of the possible realizations of the process.

Central in these calculations are the $n$-step transition probability matrices $\mathbf{P}^{(n)} = \|P_{ij}^{(n)}\|$. Here, $P_{ij}^{(n)}$ denotes the probability that the process goes from state $i$ to state $j$ in $n$ transitions. Formally,

$$P_{ij}^{(n)} = \Pr\{X_{m+n} = j | X_m = i\}. \tag{3.10}$$

Observe that we are dealing only with temporally homogeneous processes having stationary transition probabilities, since otherwise the left side of (3.10) would also depend on $m$.

The Markov property allows us to express (3.10) in terms of $\|P_{ij}\|$ as stated in the following theorem.

**Theorem 3.1.** *The n-step transition probabilities of a Markov chain satisfy*

$$P_{ij}^{(n)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)}, \tag{3.11}$$

*where we define*

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

From the theory of matrices, we recognize the relation (3.11) as the formula for matrix multiplication so that $\mathbf{P}^{(n)} = \mathbf{P} \times \mathbf{P}^{(n-1)}$. By iterating this formula, we obtain

$$\mathbf{P}^{(n)} = \underbrace{\mathbf{P} \times \mathbf{P} \times \cdots \times \mathbf{P}}_{n\,\text{factors}} = \mathbf{P}^n; \tag{3.12}$$

in other words, the $n$-step transition probabilities $P_{ij}^{(n)}$ are the entries in the matrix $\mathbf{P}^n$, the $n$th power of $\mathbf{P}$.

**Proof.** The proof proceeds via a *first step analysis*, a breaking down, or analysis, of the possible transitions on the first step, followed by an application of the Markov property. The event of going from state $i$ to state $j$ in $n$ transitions can be realized in the mutually exclusive ways of going to some intermediate state $k(k = 0, 1, \ldots)$ in the first transition, and then going from state $k$ to state $j$ in the remaining $(n-1)$ transitions. Because of the Markov property, the probability of the second transition is $P_{kj}^{(n-1)}$ and that of the first is clearly $P_{ik}$. If we use the law of total probability, then (3.11) follows. The steps are

$$\begin{aligned} P_{ij}^{(n)} &= \Pr\{X_n = j | X_0 = i\} = \sum_{k=0}^{\infty} \Pr\{X_n = j, X_1 = k | X_0 = i\} \\ &= \sum_{k=0}^{\infty} \Pr\{X_1 = k | X_0 = i\} \Pr\{X_n = j | X_0 = i, X_1 = k\} \\ &= \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)}. \end{aligned}$$

If the probability of the process initially being in state $j$ is $p_j$, i.e., the distribution law of $X_0$ is $\Pr\{X_0 = j\} = p_j$, then the probability of the process being in state $k$ at time $n$ is

$$p_k^{(n)} = \sum_{j=0}^{\infty} p_j P_{jk}^{(n)} = \Pr\{X_n = k\}. \tag{3.13}$$

■

## Exercises

**3.2.1** A Markov chain $\{X_n\}$ on the states $0, 1, 2$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.1 & 0.2 & 0.7 \\ 0.2 & 0.2 & 0.6 \\ 0.6 & 0.1 & 0.3 \end{array} \right\| \end{array}.$$

**(a)** Compute the two-step transition matrix $P^2$.
**(b)** What is $\Pr\{X_3 = 1|X_1 = 0\}$?
**(c)** What is $\Pr\{X_3 = 1|X_0 = 0\}$?

**3.2.2** A particle moves among the states 0, 1, 2 according to a Markov process whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{array} \right\| \end{array}.$$

Let $X_n$ denote the position of the particle at the $n$th move. Calculate $\Pr\{X_n = 0|X_0 = 0\}$ for $n = 0, 1, 2, 3, 4$.

**3.2.3** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{array} \right\| \end{array}.$$

Determine the conditional probabilities

$$\Pr\{X_3 = 1|X_0 = 0\} \quad \text{and} \quad \Pr\{X_4 = 1|X_0 = 0\}.$$

**3.2.4** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.4 & 0.1 & 0.5 \end{array} \right\| \end{array}.$$

If it is known that the process starts in state $X_0 = 1$, determine the probability $\Pr\{X_2 = 2\}$.

**3.2.5** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.1 & 0.1 & 0.8 \\ 0.2 & 0.2 & 0.6 \\ 0.3 & 0.3 & 0.4 \end{array} \right\| \end{array}.$$

Determine the conditional probabilities

$$\Pr\{X_3 = 1 | X_1 = 0\} \quad \text{and} \quad \Pr\{X_2 = 1 | X_0 = 0\}.$$

**3.2.6** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{array} \right\| \end{array}$$

and initial distribution $p_0 = 0.5$ and $p_1 = 0.5$. Determine the probabilities $\Pr\{X_2 = 0\}$ and $\Pr\{X_3 = 0\}$.

## Problems

**3.2.1** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.4 & 0.3 & 0.2 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.2 & 0.1 & 0.4 & 0.3 \end{array} \right\| \end{array}.$$

Suppose that the initial distribution is $p_i = \frac{1}{4}$ for $i = 0, 1, 2, 3$. Show that $\Pr\{X_n = k\} = \frac{1}{4}, k = 0, 1, 2, 3,$ for all $n$. Can you deduce a general result from this example?

**3.2.2** Consider the problem of sending a binary message, 0 or 1, through a signal channel consisting of several stages, where transmission through each stage is subject to a fixed probability of error $\alpha$. Let $X_0$ be the signal that is sent, and let $X_n$ be the signal that is received at the $n$th stage. Suppose $X_n$ is a Markov chain with transition probabilities $P_{00} = P_{11} = 1 - \alpha$ and $P_{01} = P_{10} = \alpha$, $(0 < \alpha < 1)$. Determine $\Pr\{X_5 = 0 | X_0 = 0\}$, the probability of correct transmission through five stages.

**3.2.3** Let $X_n$ denote the quality of the $n$th item produced by a production system with $X_n = 0$ meaning "good" and $X_n = 1$ meaning "defective." Suppose that $X_n$ evolves as a Markov chain whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left\| \begin{array}{cc} 0.99 & 0.01 \\ 0.12 & 0.88 \end{array} \right\| \end{array}.$$

What is the probability that the fourth item is defective given that the first item is defective?

**3.2.4** Suppose $X_n$ is a two-state Markov chain whose transition probability matrix is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & \quad 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{Vmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{Vmatrix} \end{matrix}.$$

Then, $Z_n = (X_{n-1}, X_n)$ is a Markov chain having the four states $(0,0), (0,1), (1,0)$, and $(1,1)$. Determine the transition probability matrix.

**3.2.5** A Markov chain has the transition probability matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & \;\; 1 & \;\; 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{Vmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0 & 0 & 1 \end{Vmatrix} \end{matrix}.$$

The Markov chain starts at time zero in state $X_0 = 0$. Let

$$T = \min\{n \geq 0; X_n = 2\}$$

be the first time that the process reaches state 2. Eventually, the process will reach and be absorbed into state 2. If in some experiment we observed such a process and noted that absorption had not yet taken place, we might be interested in the conditional probability that the process is in state 0 (or 1), given that absorption had not yet taken place. Determine $\Pr\{X_3 = 0 | X_0, T > 3\}$.

**Hint:** The event $\{T > 3\}$ is exactly the same as the event $\{X_3 \neq 2\} = \{X_3 = 0\} \cup \{X_3 = 1\}$.

## 3.3 Some Markov Chain Models

Markov chains can be used to model and quantify a large number of natural physical, biological, and economic phenomena that can be described by them. This is enhanced by the amenability of Markov chains to quantitative manipulation. In this section, we give several examples of Markov chain models that arise in various parts of science. General methods for computing certain functionals on Markov chains are derived in the following section.

### *3.3.1 An Inventory Model*

Consider a situation in which a commodity is stocked in order to satisfy a continuing demand. We assume that the replenishment of stock takes place at the end of periods labeled $n = 0, 1, 2, \ldots$, and we assume that the total aggregate demand for the commodity during period $n$ is a random variable $\xi_n$ whose distribution function is independent of the time period,

$$\Pr\{\xi_n = k\} = a_k \quad \text{for } k = 0, 1, 2, \ldots, \tag{3.14}$$

where $a_k \geq 0$ and $\sum_{k=0}^{\infty} a_k = 1$. The stock level is examined at the end of each period. A replenishment policy is prescribed by specifying two nonnegative critical numbers $s$ and $S > s$ whose interpretation is, if the end-of-period stock quantity is not greater than $s$, then an amount sufficient to increase the quantity of stock on hand up to the level $S$ is immediately procured. If, however, the available stock is in excess of $s$, then no replenishment of stock is undertaken. Let $X_n$ denote the quantity on hand at the end of period $n$ just prior to restocking. The states of the process $\{X_n\}$ consist of the possible values of stock size

$$S, S-1, \ldots, +1, 0, -1, -2, \ldots,$$

where a negative value is interpreted as an unfilled demand that will be satisfied immediately upon restocking.

The process $\{X_n\}$ is depicted in Figure 3.1.

According to the rules of the inventory policy, the stock levels at two consecutive periods are connected by the relation

$$X_{n+1} = \begin{cases} X_n - \xi_{n+1} & \text{if } s < X_n \leq S, \\ S - \xi_{n+1} & \text{if } X_n \leq s, \end{cases} \tag{3.15}$$

where $\xi_n$ is the quantity demanded in the $n$th period, stipulated to follow the probability law (3.14). If we assume that the successive demands $\xi_1, \xi_2, \ldots$ are independent random variables, then the stock values $X_0, X_1,\ X_2, \ldots$ constitute a Markov chain whose transition probability matrix can be calculated in accordance with relation (3.15). Explicitly,

$$\begin{aligned} P_{ij} &= \Pr\{X_{n+1} = j | X_n = i\} \\ &= \begin{cases} \Pr\{\xi_{n+1} = i - j\} & \text{if } s < i \leq S, \\ \Pr\{\xi_{n+1} = S - j\} & \text{if } i \leq s. \end{cases} \end{aligned}$$

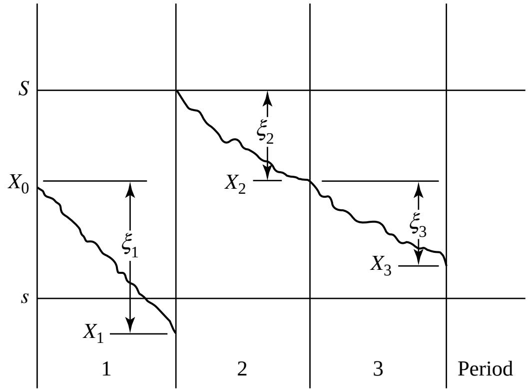



**Figure 3.1** The inventory process.

Consider a spare parts inventory model as a numerical example in which either 0, 1, or 2 repair parts are demanded in any period, with

$$\Pr\{\xi_n = 0\} = 0.5, \quad \Pr\{\xi_n = 1\} = 0.4, \quad \Pr\{\xi_n = 2\} = 0.1,$$

and suppose $s = 0$, while $S = 2$. The possible values for $X_n$ are $S = 2, 1, 0$, and $-1$. To illustrate the transition probability calculations, we will consider first the determination of $P_{10} = \Pr\{X_{n+1} = 0 | X_n = 1\}$. When $X_n = 1$, then no replenishment takes place and the next state $X_{n+1} = 0$ results when the demand $\xi_{n+1} = 1$, and this occurs with probability $P_{10} = 0.4$. To illustrate another case, if $X_n = 0$, then instantaneous replenishment to $S = 2$ ensues, and a next period level of $X_{n+1} = 0$ results from the demand quantity $\xi_{n+1} = 2$. The corresponding probability of this outcome yields $P_{00} = 0.1$. Continuing in this manner, we obtain the transition probability matrix

$$\mathbf{P} = \begin{array}{r} \\ -0 \\ 0 \\ +1 \\ +2 \end{array} \begin{array}{c} \begin{array}{cccc} -1 & 0 & +1 & +2 \end{array} \\ \left\| \begin{array}{cccc} 0 & 0.1 & 0.4 & 0.5 \\ 0 & 0.1 & 0.4 & 0.5 \\ 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \end{array} \right\| \end{array}.$$

Important quantities of interest in inventory models of this type are the long-term fraction of periods in which demand is not met ($X_n < 0$) and long-term average inventory level. Using the notation $p_j^{(n)} = \Pr\{X_n = j\}$, we give these quantities, respectively, as $\lim_{n\to\infty} \Sigma_{j<0} p_j^{(n)}$ and $\lim_{n\to\infty} \Sigma_{j>0} j p_j^{(n)}$. This illustrates the importance of determining conditions under which the probabilities $p_j^{(n)}$ stabilize and approach limiting probabilities $\pi_j$ as $n \to \infty$ and of determining methods for calculating the limiting probabilities $\pi_j$ when they exist. These topics are the subject of Chapter 4.

### *3.3.2 The Ehrenfest Urn Model*

A classical mathematical description of diffusion through a membrane is the famous Ehrenfest urn model. Imagine two containers containing a total of $2a$ balls (molecules). Suppose the first container, labeled A, holds $k$ balls and the second container, B, holds the remaining $2a - k$ balls. A ball is selected at random (all selections are equally likely) from the totality of the $2a$ balls and moved to the other container. (A molecule diffuses at random through the membrane.) Each selection generates a transition of the process. Clearly, the balls fluctuate between the two containers with an average drift from the urn with the excess numbers to the one with the smaller concentration.

Let $Y_n$ be the number of balls in urn A at the $n$th stage, and define $X_n = Y_n - a$. Then, $\{X_n\}$ is a Markov chain on the states $i = -a, \ -a+1, \ldots, -1, 0, +1, \ldots, a$ with transition probabilities

$$P_{ij} = \begin{cases} \dfrac{a-i}{2a} & \text{if } j = i+1, \\ \dfrac{a+i}{2a} & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

An important quantity in the Ehrenfest urn model is the long-term, or equilibrium, distribution of the number of balls in each urn.

### 3.3.3 Markov Chains in Genetics

The following idealized genetics model was introduced by S. Wright to investigate the fluctuation of gene frequency under the influence of mutation and selection. We begin by describing the so-called simple haploid model of random reproduction, disregarding mutation pressures and selective forces. We assume that we are dealing with a fixed population size of $2N$ genes composed of type-**a** and type-**A** individuals. The makeup of the next generation is determined by $2N$ independent Bernoulli trials as follows: If the parent population consists of $j$ **a**-genes and $2N-j$ **A**-genes, then each trial results in **a** or **A** with probabilities

$$p_j = \frac{j}{2N}, \quad q_j = 1 - \frac{j}{2N},$$

respectively. Repeated selections are done with replacement. By this procedure, we generate a Markov chain $\{X_n\}$, where $X_n$ is the number of **a**-genes in the $n$th generation among a constant population size of $2N$ individuals. The state space contains the $2N+1$ values $\{0, 1, 2, \ldots, 2N\}$. The transition probability matrix is computed according to the binomial distribution as

$$\Pr\{X_{n+1} = k | X_n = j\} = P_{jk} = \binom{2N}{k} p_j^k q_j^{2N-k} \qquad (3.16)$$
$$(j, k = 0, 1, \ldots, 2N).$$

For some discussion of the biological justification of these postulates, we refer the reader to Fisher.*

Notice that states 0 and $2N$ are completely absorbing in the sense that once $X_n = 0$ (or $2N$), then $X_{n+k} = 0$ (or $2N$, respectively) for all $k \geq 0$. One of the questions of interest is to determine the probability, under the condition $X_0 = i$, that the population will attain fixation, i.e., that it will become a pure population composed only of **a**-genes or **A**-genes. It is also pertinent to determine the rate of approach to fixation. We will examine such questions in our general analysis of absorption probabilities.

A more complete model takes account of mutation pressures. We assume that prior to the formation of the new generation, each gene has the possibility to mutate, i.e., to change into a gene of the other kind. Specifically, we assume that for each gene the mutation $\mathbf{a} \to \mathbf{A}$ occurs with probability $\alpha$, and $\mathbf{A} \to \mathbf{a}$ occurs with probability $\beta$. Again we assume that the composition of the next generation is determined by $2N$ independent binomial trials. The relevant values of $p_j$ and $q_j$ when the parent

* R. A. Fisher, *The Genetical Theory of Natural Selection*, Oxford (Clarendon) Press, London and New York, 1962.

population consists of $j$ **a**-genes are now taken to be

$$p_j = \frac{j}{2N}(1-\alpha) + \left(1 - \frac{j}{2N}\right)\beta$$

and (3.17)

$$q_j = \frac{j}{2N}\alpha + \left(1 - \frac{j}{2N}\right)(1-\beta).$$

The rationale is as follows: We assume that the mutation pressures operate first, after which a new gene is chosen by random selection from the population. Now, the probability of selecting an **a**-gene after the mutation forces have acted is just $1/(2N)$ times the number of **a**-genes present; hence, the average probability (averaged with respect to the possible mutations) is simply $1/(2N)$ times the average number of **a**-genes after mutation. But this average number is clearly $j(1-\alpha) + (2N-j)\beta$, which leads at once to (3.17).

The transition probabilities of the associated Markov chain are calculated by (3.16) using the values of $p_j$ and $q_j$ given in (3.17).

If $\alpha\beta > 0$, then fixation will not occur in any state. Instead, as $n \to \infty$, the distribution function of $X_n$ will approach a steady-state distribution of a random variable $\xi$, where $\Pr\{\xi = k\} = \pi_k (k = 0, 1, 2, \ldots, 2N) \left(\sum_{k=0}^{n} \pi_k = 1, \pi_k > 0\right)$. The distribution function of $\xi$ is called the steady-state gene frequency distribution.

We return to the simple random mating model and discuss the concept of a selection force operating in favor of, say, **a**-genes. Suppose we wish to impose a selective advantage for **a**-genes over **A**-genes so that the relative number of offspring have expectations proportional to $1+s$ and 1, respectively, where $s$ is small and positive. We replace $p_j = j/(2N)$ and $q_j = 1 - j/(2N)$ by

$$p_j = \frac{(1+s)j}{2N+sj}, \quad q_j = 1 - p_j,$$

and build the next generation by binomial sampling as before. If the parent population consisted of $j$ **a**-genes, then in the next generation the expected population sizes of **a**-genes and **A**-genes, respectively, are

$$2N\frac{(1+s)j}{2N+sj}, \quad 2N\frac{(2N-j)}{2N+sj}.$$

The ratio of expected population size of **a**-genes to **A**-genes at the $(n+1)$st generation is

$$\frac{1+s}{1} \times \frac{j}{2N-j} = \left(\frac{1+s}{1}\right)\left(\frac{\text{number of } \mathbf{a}\text{-genes in the } n\text{th generation}}{\text{number of } \mathbf{A}\text{-genes in the } n\text{th generation}}\right),$$

which explains the meaning of selection.

### 3.3.4 A Discrete Queueing Markov Chain

Customers arrive for service and take their place in a waiting line. During each period of time, a single customer is served, provided that at least one customer is present. If no customer awaits service, then during this period no service is performed. (We can imagine, e.g., a taxi stand at which a cab arrives at fixed time intervals to give service. If no one is present, the cab immediately departs.) During a service period, new customers may arrive. We suppose that the actual number of customers that arrive during the $n$th period is a random variable $\xi_n$ whose distribution is independent of the period and is given by

$$\Pr\{k \text{ customers arrive in a service period}\} = \Pr\{\xi_n = k\} = a_k,$$

for $k = 0, 1, \ldots,$ where $a_k \geq 0$ and $\sum_{k=0}^{\infty} a_k = 1$.

We also assume that $\xi_1, \xi_2, \ldots$ are independent random variables. The state of the system at the start of each period is defined to be the number of customers waiting in line for service. If the present state is $i$, then after the lapse of one period the state is

$$j = \begin{cases} i - 1 + \xi & \text{if } i \geq 1, \\ \xi & \text{if } i = 0, \end{cases} \tag{3.18}$$

where $\xi$ is the number of new customers having arrived in this period while a single customer was served. In terms of the random variables of the process, we can express (3.18) formally as

$$X_{n+1} = (X_n - 1)^+ + \xi_n,$$

where $Y^+ = \max\{Y, 0\}$. In view of (3.18), the transition probability matrix may be calculated easily, and we obtain

$$\mathbf{P} = \left\| \begin{matrix} a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{matrix} \right\|.$$

It is intuitively clear that if the expected number of new customers, $\sum_{k=0}^{\infty} k a_k$, who arrive during a service period exceeds one, then with the passage of time the length of the waiting line increases without limit. On the other hand, if $\sum_{k=0}^{\infty} k a_k < 1$, then the length of the waiting line approaches a statistical equilibrium that is described by a limiting distribution

$$\lim_{n\to\infty} \Pr\{X_n = k | X_0 = j\} = \pi_k > 0, \quad \text{for } k = 0, 1, \ldots,$$

where $\sum_{k=0}^{x} \pi_k = 1$. Important quantities to be determined by this model include the long run fraction of time that the service facility is idle, given by $\pi_0$, and the long run mean time that a customer spends in the system, given by $\sum_{k=0}^{\infty} (1 + k)\pi_k$.

## Exercises

**3.3.1** Consider a spare parts inventory model in which either 0, 1, or 2 repair parts are demanded in any period, with

$$\Pr\{\xi_n = 0\} = 0.4, \quad \Pr\{\xi_n = 1\} = 0.3, \quad \Pr\{\xi_n = 2\} = 0.3,$$

and suppose $s = 0$ and $S = 3$. Determine the transition probability matrix for the Markov chain $\{X_n\}$, where $X_n$ is defined to be the quantity on hand at the end-of-period $n$.

**3.3.2** Consider two urns A and B containing a total of $N$ balls. An experiment is performed in which a ball is selected at random (all selections equally likely) at time $t(t = 1, 2, \ldots)$ from among the totality of $N$ balls. Then, an urn is selected at random (A is chosen with probability $p$ and B is chosen with probability $q$) and the ball previously drawn is placed in this urn. The state of the system at each trial is represented by the number of balls in A. Determine the transition matrix for this Markov chain.

**3.3.3** Consider the inventory model of Section 3.3.1. Suppose that $S = 3$. Set up the corresponding transition probability matrix for the end-of-period inventory level $X_n$.

**3.3.4** Consider the inventory model of Section 3.3.1. Suppose that $S = 3$ and that the probability distribution for demand is $\Pr\{\xi = 0\} = 0.1$, $\Pr\{\xi = 1\} = 0.4$, $\Pr\{\xi = 2\} = 0.3$, and $\Pr\{\xi = 3\} = 0.2$. Set up the corresponding transition probability matrix for the end-of-period inventory level $X_n$.

**3.3.5** An urn initially contains a single red ball and a single green ball. A ball is drawn at random, removed, and replaced by a ball of the opposite color, and this process repeats so that there are always exactly two balls in the urn. Let $X_n$ be the number of red balls in the urn after $n$ draws, with $X_0 = 1$. Specify the transition probabilities for the Markov chain $\{X_n\}$.

## Problems

**3.3.1** An urn contains six tags, of which three are red and three are green. Two tags are selected from the urn. If one tag is red and the other is green, then the selected tags are discarded and two blue tags are returned to the urn. Otherwise, the selected tags are resumed to the urn. This process repeats until the urn contains only blue tags. Let $X_n$ denote the number of red tags in the urn after the $n$th draw, with $X_0 = 3$. (This is an elementary model of a chemical reaction in which red and green atoms combine to form a blue molecule.) Give the transition probability matrix.

**3.3.2** Three fair coins are tossed, and we let $X_1$ denote the number of heads that appear. Those coins that were heads on the first trial (there were $X_1$ of them) we pick up and toss again, and now we let $X_2$ be the total number of tails, including those left from the first toss. We toss again all coins showing tails, and let $X_3$ be the resulting total number of heads, including those left from the previous toss. We continue the process. The pattern is, count heads, toss heads, count tails, toss tails, count heads, toss heads, etc., and $X_0 = 3$. Then, $\{X_n\}$ is a Markov chain. What is the transition probability matrix?

**3.3.3** Consider the inventory model of Section 3.3.1. Suppose that unfulfilled demand is not back ordered but is lost.

**(a)** Set up the corresponding transition probability matrix for the end-of-period inventory level $X_n$.

**(b)** Express the long run fraction of lost demand in terms of the demand distribution and limiting probabilities for the end-of-period inventory.

**3.3.4** Consider the queueing model of Section 3.4. Now, suppose that at most a single customer arrives during a single period, but that the service time of a customer is a random variable $Z$ with the geometric probability distribution

$$\Pr\{Z = k\} = \alpha(1-\alpha)^{k-1} \quad \text{for } k = 1, 2, \ldots.$$

Specify the transition probabilities for the Markov chain whose state is the number of customers waiting for service or being served at the start of each period. Assume that the probability that a customer arrives in a period is $\beta$ and that no customer arrives with probability $1-\beta$.

**3.3.5** You are going to successively flip a quarter until the pattern *HHT* appears, that is, until you observe two successive heads followed by a tails. In order to calculate some properties of this game, you set up a Markov chain with the following states: $0, H, HH$, and $HHT$, where 0 represents the starting point, $H$ represents a single observed head on the last flip, $HH$ represents two successive heads on the last two flips, and $HHT$ is the sequence that you are looking for. Observe that if you have just tossed a tails, followed by a heads, a next toss of a tails effectively starts you over again in your quest for the *HHT* sequence. Set up the transition probability matrix.

**3.3.6** Two teams, A and B, are to play a best of seven series of games. Suppose that the outcomes of successive games are independent, and each is won by A with probability $p$ and won by B with probability $1-p$. Let the state of the system be represented by the pair $(a, b)$, where $a$ is the number of games won by A, and $b$ is the number of games won by B. Specify the transition probability matrix. Note that $a + b \le 7$ and that the entries end whenever $a = 4$ or $b = 4$.

**3.3.7** A component in a system is placed into service, where it operates until its failure, whereupon it is replaced *at the end of the period* with a new component having statistically identical properties, and the process repeats. The probability that a component lasts for $k$ periods is $\alpha_k$, for $k = 1, 2, \ldots$. Let $X_n$ be the remaining life of the component in service *at the end-of-period n*. Then, $X_n = 0$ means that $X_{n+1}$ will be the total operating life of the next component. Give the transition probabilities for the Markov chain $\{X_n\}$.

**3.3.8** Two urns A and B contain a total of $N$ balls. Assume that at time $t$, there were exactly $k$ balls in A. At time $t+1$, an urn is selected at random in proportion to its contents (i.e., A is chosen with probability $k/N$ and B is chosen with probability $(N-k)/N$). Then, a ball is selected from A with probability $p$ or from B with probability $q$ and placed in the previously chosen urn. Determine the transition matrix for this Markov chain.

**3.3.9** Suppose that two urns A and B contain a total of $N$ balls. Assume that at time $t$, there are exactly $k$ balls in A. At time $t+1$, a ball and an urn are chosen with probability depending on the contents of the urn (i.e., a ball is chosen from A with probability $k/N$ or from B with probability $(N-k)/N$). Then, the ball is placed into one of the urns, where urn A is chosen with probability $k/N$ or urn B is chosen with probability $(N-k)/N$. Determine the transition matrix of the Markov chain with states represented by the contents of A.

**3.3.10** Consider a discrete-time, periodic review inventory model and let $\xi_n$ be the total demand in period $n$, and let $X_n$ be the inventory quantity on hand at the end-of-period $n$. An $(s,S)$ inventory policy is used: If the end-of-period stock is not greater than $s$, then a quantity is instantly procured to bring the level up to $S$. If the end-of-period stock exceeds $s$, then no replenishment takes place.

**(a)** Suppose that $s=1, S=4$, and $X_0=S=4$. If the period demands turn out to be $\xi_1=2, \xi_2=3, \xi_3=4, \xi_4=0, \xi_5=2, \xi_6=1, \xi_7=2$, and $\xi_8=2$, what are the end-of-period stock levels $X_n$ for periods $n=1,2,\ldots,8$?

**(b)** Suppose $\xi_1, \xi_2, \ldots$ are independent random variables where $\Pr\{\xi_n=0\}= 0.1, \Pr\{\xi_n=1\}=0.3, \Pr\{\xi_n=2\}=0.3, \Pr\{\xi_n=3\}=0.2$, and $\Pr\{\xi_n=4\}= 0.1$. Then, $X_0, X_1, \ldots$ is a Markov chain. Determine $P_{41}$ and $P_{04}$.

## 3.4 First Step Analysis

A surprising number of functionals on a Markov chain can be evaluated by a technique that we call *first step analysis*. This method proceeds by analyzing, or breaking down, the possibilities that can arise at the end of the first transition, and then invoking the law of total probability coupled with the Markov property to establish a characterizing relationship among the unknown variables. We first applied this technique in Theorem 3.1. In this section, we develop a series of applications of the technique.

### 3.4.1 Simple First Step Analyses

Consider the Markov chain $\{X_n\}$ whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{array} \right\| \end{array},$$

where $\alpha > 0, \beta > 0, \gamma > 0$, and $\alpha+\beta+\gamma=1$. If the Markov chain begins in state 1, it remains there for a random duration and then proceeds either to state 0 or to state 2, where it is trapped or absorbed. That is, once in state 0, the process remains there for ever after, as it also does in state 2. Two questions arise: In which state, 0 or 2, is the process ultimately trapped, and how long, on the average, does it take to reach one of these states? Both questions are easily answered by instituting a first step analysis.

We begin by more precisely defining the questions. Let

$$T = \min\{n \geq 0; X_n = 0 \quad \text{or} \quad X_n = 2\}$$

be the time of absorption of the process. In terms of this random absorption time, the two questions ask us to find

$$u = \Pr\{X_T = 0 | X_0 = 1\}$$

and

$$v = E[T | X_0 = 1].$$

We proceed to institute a first step analysis, considering separately the three contingencies $X_1 = 0, X_1 = 1$, and $X_1 = 2$, with respective probabilities $\alpha$, $\beta$, and $\gamma$. Consider $u = \Pr\{X_T = 0|X_0 = 1\}$. If $X_1 = 0$, which occurs with probability $\alpha$, then $T = 1$ and $X_T = 0$. If $X_1 = 2$, which occurs with probability $\gamma$, then again $T = 1$, but $X_T = 2$. Finally, if $X_1 = 1$, which occurs with probability $\beta$, then the process returns to state 1 and the problem repeats from the same state as before. In symbols, we claim that

$$\begin{aligned}
&\Pr\{X_T = 0|X_1 = 0\} = 1,\\
&\Pr\{X_T = 0|X_1 = 2\} = 0,\\
&\Pr\{X_T = 0|X_1 = 1\} = u,
\end{aligned}$$

which inserted into the law of total probability gives

$$\begin{aligned}
u &= \Pr\{X_T = 0|X_0 = 1\}\\
&= \sum_{k=0}^{2} \Pr\{X_T = 0|X_0 = 1, X_1 = k\} \Pr\{X_1 = k|X_0 = 1\}\\
&= \sum_{k=0}^{2} \Pr\{X_T = 0|X_1 = k\} \Pr\{X_1 = k|X_0 = 1\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(by the Markov property)}\\
&= 1(\alpha) + u(\beta) + 0(\gamma).
\end{aligned}$$

Thus, we obtain the equation

$$u = \alpha + \beta u, \tag{3.19}$$

which gives

$$u = \frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha+\gamma}.$$

Observe that this quantity is the conditional probability of a transition to 0, given that a transition to 0 or 2 occurred. That is, the answer makes sense.

We turn to determining the mean time to absorption, again analyzing the possibilities arising on the first step. The absorption time $T$ is always at least 1. If either $X_1 = 0$ or $X_1 = 2$, then no further steps are required. If, on the other hand, $X_1 = 1$, then the process is back at its starting point, and on the average, $v = E[T|X_0 = 1]$ *additional* steps are required for absorption. Weighting these contingencies by their respective probabilities, we obtain for $v = E[T|X_0 = 1]$,

$$\begin{aligned} v &= 1 + \alpha(0) + \beta(v) + \gamma(0) \\ &= 1 + \beta v, \end{aligned} \tag{3.20}$$

which gives

$$v = \frac{1}{1-\beta}.$$

In the example just studied, the reader is invited to verify that $T$ has the geometric distribution in which

$$\Pr\{T > k|X_0 = 1\} = \beta^k \quad \text{for } k = 0, 1, \ldots,$$

and, therefore,

$$E[T|X_0 = 1] = \sum_{k=0}^{\infty} \Pr\{T > k|X_0 = 1\} = \frac{1}{1-\beta}.$$

That is, a direct calculation verifies the result of the first step analysis. Unfortunately, in more general Markov chains, a direct calculation is rarely possible, and first step analysis provides the only solution technique.

A significant extension occurs when we move up to the four-state Markov chain whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Absorption now occurs in states 0 and 3, and states 1 and 2 are “transient.” The probability of ultimate absorption in state 0, say, now depends on the transient state in which the process began. Accordingly, we must extend our notation to include the starting state. Let

$$T = \min\{n \geq 0; X_n = 0 \quad \text{or} \quad X_n = 3\},$$
$$u_i = \Pr\{X_T = 0|X_0 = i\} \quad \text{for } i = 1, 2,$$

and

$$v_i = E[T|X_0 = i] \quad \text{for } i = 1, 2.$$

We may extend the definitions for $u_i$ and $v_i$ in a consistent and commonsense manner by prescribing $u_0 = 1, u_3 = 0$, and $v_0 = v_3 = 0$.

The first step analysis now requires us to consider the two possible starting states $X_0 = 1$ and $X_0 = 2$ separately. Considering $X_0 = 1$ and applying a first step analysis to $u_1 = \Pr\{X_T = 0|X_0 = 1\}$, we obtain

$$u_1 = P_{10} + P_{11}u_1 + P_{12}u_2. \tag{3.21}$$

The three terms on the right correspond to the contingencies $X_1 = 0, X_1 = 1$, and $X_1 = 2$, respectively, with the conditional probabilities

$$\Pr\{X_T = 0|X_1 = 0\} = 1,$$
$$\Pr\{X_T = 0|X_1 = 1\} = u_1,$$

and

$$\Pr\{X_T = 0|X_1 = 2\} = u_2.$$

The law of total probability then applies to give (3.21), just as it was used in obtaining (3.19). A similar equation is obtained for $u_2$:

$$u_2 = P_{20} + P_{21}u_1 + P_{22}u_2. \tag{3.22}$$

The two equations in $u_1$ and $u_2$ are now solved simultaneously. To give a numerical example, we will suppose

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}. \tag{3.23}$$

The first step analysis equations (3.21) and (3.22) for $u_1$ and $u_2$ are

$$u_1 = 0.4 + 0.3u_1 + 0.2u_2,$$
$$u_2 = 0.1 + 0.3u_1 + 0.3u_2,$$

or

$$0.7u_1 - 0.2u_2 = 0.4,$$
$$-0.3u_1 + 0.7u_2 = 0.1.$$

The solution is $u_1 = \frac{30}{43}$ and $u_2 = \frac{19}{43}$. Note that one cannot, in general, solve for $u_1$ without bringing in $u_2$, and vice versa. The result $u_2 = \frac{19}{43}$ tells us that once begun in state $X_0 = 2$, the Markov chain $\{X_n\}$ described by (3.23) will ultimately end up in state 0 with probability $u_2 = \frac{19}{43}$, and alternatively, will be absorbed in state 3 with probability $1 - u_2 = \frac{24}{43}$.

The mean time to absorption also depends on the starting state. The first step analysis equations for $v_i = E[T|X_0 = i]$ are

$$\begin{aligned} v_1 &= 1 + P_{11}v_1 + P_{12}v_2, \\ v_2 &= 1 + P_{21}v_1 + P_{22}v_2. \end{aligned} \tag{3.24}$$

The right side of (3.24) asserts that at least one step is always taken. If the first move is to either $X_1 = 1$ or $X_1 = 2$, then additional steps are needed, and on the average, these are $v_1$ and $v_2$, respectively. Weighting the contingencies $X_1 = 1$ and $X_1 = 2$ by their respective probabilities and summing according to the law of total probability results in (3.24).

For the transition matrix given in (3.23), the equations are

$$\begin{aligned} v_1 &= 1 + 0.3v_1 + 0.2v_2, \\ v_2 &= 1 + 0.3v_1 + 0.3v_2, \end{aligned}$$

and their solutions are $v_1 = \frac{90}{43}$ and $v_2 = \frac{100}{43}$. Again, $v_1$ cannot be obtained without also considering $v_2$, and vice versa. For a process that begins in state $X_0 = 2$, on the average $v_2 = \frac{100}{43} = 2.33$ steps will transpire prior to absorption.

To study the method in a more general context, let $\{X_n\}$ be a finite-state Markov chain whose states are labeled $0, 1, \ldots, N$. Suppose that states $0, 1, \ldots, r-1$ are *transient** in that $P_{ij}^{(n)} \to 0$ as $n \to \infty$ for $0 \le i, j < r$, while states $r, \ldots, N$ are *absorbing* ($P_{ii} = 1$ for $r \le i \le N$). The transition matrix has the form

$$\mathbf{P} = \left\| \begin{matrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{matrix} \right\| \tag{3.25}$$

where $\mathbf{0}$ is an $(N - r + 1) \times r$ matrix all of whose entries are zero, $\mathbf{I}$ is an $(N - r + 1) \times (N - r + 1)$ identity matrix, and $Q_{ij} = P_{ij}$ for $0 \le i, j < r$.

Started at one of the transient states $X_0 = i$, where $0 \le i < r$, such a process will remain in the transient states for some random duration, but ultimately the process gets trapped in one of the absorbing states $i = r, \ldots, N$. Functionals of importance are the mean duration until absorption and the probability distribution over the states in which absorption takes place.

Let us consider the second question first and fix a state $k$ among the absorbing states ($r \le k \le N$). The probability of ultimate absorption in state $k$, as opposed to some other absorbing state, depends on the initial state $X_0 = i$. Let $U_{ik} = u_i$ denote this probability, where we suppress the target state $k$ in the notation for typographical convenience.

* The definition of a transient state is different for an infinite-state Markov chain. See Chapter 4, Section 4.3.

We begin a first step analysis by enumerating the possibilities in the first transition. Starting from state $i$, with probability $P_{ik}$ the process immediately goes to state $k$, thereafter to remain, and this is the first possibility considered. Alternatively, the process could move on its first step to an absorbing state $j \neq k$, where $r \leq j \leq N$, in which case ultimate absorption in state $k$ is precluded. Finally, the process could move to a transient state $j < r$. Because of the Markov property, once in state $j$, then the probability of ultimate absorption in state $k$ is $u_j = U_{jk}$ by definition. Weighting the enumerated possibilities by their respective probabilities via the law of total probability, we obtain the relation

$$\begin{aligned} u_i &= \Pr\{\text{Absorption in } k | X_0 = i\} \\ &= \sum_{j=0}^{N} \Pr\{\text{Absorption in } k | X_0 = i, X_1 = j\} P_{ij} \\ &= P_{ik} + \sum_{\substack{j=r \\ j \neq k}}^{N} P_{ij} \times 0 + \sum_{j=0}^{r-1} P_{ij} u_j. \end{aligned}$$

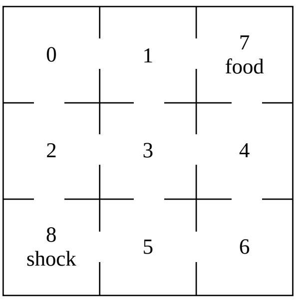
To summarize, for a fixed absorbing state $k$, the quantities

$$u_i = U_{ik} = \Pr\{\text{Absorption in } k | X_0 = i\} \quad \text{for } 0 \leq i < r$$

satisfy the inhomogeneous system of linear equations

$$U_{ik} = P_{ik} + \sum_{j=0}^{r-1} P_{ij} U_{jk}, \quad i = 0, 1, \ldots, r-1. \tag{3.26}$$

**Example** *A Maze* A white rat is put into the maze shown:




In the absence of learning, one might hypothesize that the rat would move through the maze at random; i.e., if there are $k$ ways to leave a compartment, then the rat would choose each of these with probability $1/k$. Assume that the rat makes one change to some adjacent compartment at each unit of time and let $X_n$ denote the compartment occupied at stage $n$. We suppose that compartment 7 contains food and compartment

8 contains an electrical shocking mechanism, and we ask the probablity that the rat, moving at random, encounters the food before being shocked. The appropriate transition probability matrix is

$$
\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\left\| \begin{array}{ccccccccc}
 & \frac{1}{2} & \frac{1}{2} & & & & & & \\
\frac{1}{3} & & & \frac{1}{3} & & & & \frac{1}{3} & \\
\frac{1}{3} & & & \frac{1}{3} & & & & & \frac{1}{3} \\
 & \frac{1}{4} & \frac{1}{4} & & \frac{1}{4} & \frac{1}{4} & & & \\
 & & & \frac{1}{3} & & & \frac{1}{3} & \frac{1}{3} & \\
 & & & \frac{1}{3} & & & \frac{1}{3} & & \frac{1}{3} \\
 & & & & \frac{1}{2} & \frac{1}{2} & & & \\
 & & & & & & & 1 & \\
 & & & & & & & & 1
\end{array} \right\|
\end{array}.
$$

Let $u_i = u_i(7)$ denote the probability of absorption in the food compartment 7, given that the rat is dropped initially in compartment $i$. Then, equation (3.26) becomes, in this particular instance,

$$
\begin{aligned}
u_0 &= \tfrac{1}{2}u_1 + \tfrac{1}{2}u_2, \\
u_1 &= \tfrac{1}{3} + \tfrac{1}{3}u_0 + \tfrac{1}{3}u_3, \\
u_2 &= \tfrac{1}{3}u_0 + \tfrac{1}{3}u_3, \\
u_3 &= \tfrac{1}{4}u_1 + \tfrac{1}{4}u_2 + \tfrac{1}{4}u_4 + \tfrac{1}{4}u_5, \\
u_4 &= \tfrac{1}{3} + \tfrac{1}{3}u_3 + \tfrac{1}{3}u_6, \\
u_5 &= \tfrac{1}{3}u_3 + \tfrac{1}{3}u_6, \\
u_6 &= \tfrac{1}{2}u_4 + \tfrac{1}{2}u_5.
\end{aligned}
$$

Turning to the solution, we see that the symmetry of the maze implies that $u_0 = u_6, u_2 = u_5$, and $u_1 = u_4$. We also must have $u_3 = \frac{1}{2}$. With these simplifications, the equations for $u_0, u_1$, and $u_2$ become

$$u_0 = \frac{1}{2}u_1 + \frac{1}{2}u_2,$$
$$u_1 = \frac{1}{2} + \frac{1}{3}u_0,$$
$$u_2 = \frac{1}{6} + \frac{1}{3}u_0,$$

and the natural substitutions give $u_0 = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{3}u_0\right) + \frac{1}{2}\left(\frac{1}{6} + \frac{1}{3}u_0\right)$, or $u_0 = \frac{1}{2}, u_1 = \frac{2}{3}$, and $u_2 = \frac{1}{3}$.

One might compare these theoretical values under random moves with actual observations as an indication of whether or not learning is taking place.

### 3.4.2 The General Absorbing Markov Chain

Let $\{X_n\}$ be a Markov chain whose transition probability matrix takes the form (3.25). We turn to a more general form of the first question by introducing the random absorption time $T$. Formally, we define

$$T = \min\{n \geq 0; X_n \geq r\}.$$

Let us suppose that associated with each transient state $i$ is a rate $g(i)$ and that we wish to determine the mean total rate that is accumulated up to absorption. Let $w_i$ be this mean total amount, where the subscript $i$ denotes the starting position $X_0 = i$. To be precise, let

$$w_i = E\left[\sum_{n=0}^{T-1} g(X_n) | X_0 = i\right].$$

The choice $g(i) = 1$ for all $i$ yields $\sum_{n=0}^{T-1} g(X_n) = \sum_{n=0}^{T-1} 1 = T$, and then $w_i$ is identical to $v_i \equiv E[T|X_0 = i]$, the mean time until absorption. For a transient state $k$, the choice

$$g(i) = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k, \end{cases}$$

gives $w_i = W_{ik}$, the mean number of visits to state $k (0 \leq k < r)$ prior to absorption.

We again proceed via a first step analysis. The sum $\sum_{n=0}^{T-1} g(X_n)$ always includes the first term $g(X_0) = g(i)$. In addition, if a transition is made from $i$ to a transient state $j$, then the sum includes future terms as well. By invoking the Markov property, we deduce that this future sum proceeding from state $j$ has an expected value equal to $w_j$.

Weighting this by the transition probability $P_{ij}$ and then summing all contributions in accordance with the law of total probability, we obtain the joint relations

$$w_i = g(i) + \sum_{j=0}^{r-1} P_{ij} w_j \quad \text{for } i = 0, \ldots, r-1. \tag{3.27}$$

The special case in which $g(i) = 1$ for all $i$ determines $v_i = E[T|X_0 = i]$ as solving

$$v_i = 1 + \sum_{j=0}^{r-1} P_{ij} v_j \quad \text{for } i = 0, 1, \ldots, r-1. \tag{3.28}$$

The case in which

$$g(i) = \delta_{ik} = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k, \end{cases}$$

determines $W_{ik}$, the mean number of visits to state $k$ prior to absorption starting from state $i$, as solving

$$W_{ik} = \delta_{ik} + \sum_{j=0}^{r-1} P_{ij} W_{jk} \quad \text{for } i = 0, 1, \ldots, r-1. \tag{3.29}$$

**Example** *A Model of Fecundity* Changes in sociological patterns such as increase in age at marriage, more remarriages after widowhood, and increased divorce rates have profound effects on overall population growth rates. Here, we attempt to model the lifespan of a female in a population in order to provide a framework for analyzing the effect of social changes on average fecundity.

The general model we propose has a large number of states delimiting the age and status of a typical female in the population. For example, we begin with the 12 age groups 0–4 years, 5–9 years, ..., 50–54 years, 55 years, and over. In addition, each of these age groups might be further subdivided according to marital status: single, married, separated, divorced, or widowed, and might also be subdivided according to the number of children. Each female would begin in the (0–4, single) category and end in a distinguished state $\Delta$ corresponding to death or emigration from the population. However, the duration spent in the various other states might differ among different females. Of interest is the mean duration spent in the categories of maximum fertility, or more generally, a mean sum of durations weighted by appropriate fecundity rates.

When there are a large number of states in the model, as just sketched, the relevant calculations require a computer. We turn to a simpler model which, while less realistic, will serve to illustrate the concepts and approach. We introduce the states

$E_0$: Prepuberty, $E_3$: Divorced,
$E_1$: Single, $E_4$: Widowed,
$E_2$: Married, $E_5$: $\Delta$,

and we are interested in the mean duration spent in state $E_2$: Married, since this corresponds to the state of maximum fecundity. To illustrate the computations, we will suppose the transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ E_0 \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{array} \begin{array}{c} \begin{array}{cccccc} E_0 & E_1 & E_2 & E_3 & E_4 & E_5 \end{array} \\ \left\| \begin{array}{cccccc} 0 & 0.9 & 0 & 0 & 0 & 0.1 \\ 0 & 0.5 & 0.4 & 0 & 0 & 0.1 \\ 0 & 0 & 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0 & 0.4 & 0.5 & 0 & 0.1 \\ 0 & 0 & 0.4 & 0 & 0.5 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 1.0 \end{array} \right\| \end{array}.$$

In practice, such a matrix would be estimated from demographic data.

Every person begins in state $E_0$ and ends in state $E_5$, but a variety of intervening states may be visited. We wish to determine the mean duration spent in state $E_2$: Married. The powerful approach of *first step analysis* begins by considering the slightly more general problem in which the initial state is varied. Let $w_i = W_{i2}$ be the mean duration in state $E_2$ given the initial state $X_0 = E_i$ for $i = 0, 1, \ldots, 5$. We are interested in $w_0$, the mean duration corresponding to the initial state $E_0$.

First step analysis breaks down, or analyzes, the possibilities arising in the first transition, and using the Markov property, an equation that relates $w_0, \ldots, w_5$ results.

We begin by considering $w_0$. From state $E_0$, a transition to one of the states $E_1$ or $E_5$ occurs, and the mean duration spent in $E_2$ starting from $E_0$ must be the appropriately weighted average of $w_1$ and $w_5$. That is,

$$w_0 = 0.9w_1 + 0.1w_5.$$

Proceeding in a similar manner, we obtain

$$w_1 = 0.5w_1 + 0.4w_2 + 0.1w_5.$$

The situation changes when the process begins in state $E_2$ because in counting the mean duration spent in $E_2$, we must count this initial visit plus any subsequent visits that may occur. Thus, for $E_2$, we have

$$w_2 = 1 + 0.6w_2 + 0.2w_3 + 0.1w_4 + 0.1w_5.$$

The other states give us

$$\begin{aligned} w_3 &= 0.4w_2 + 0.5w_3 + 0.1w_5, \\ w_4 &= 0.4w_2 + 0.5w_4 + 0.1w_5, \\ w_5 &= w_5. \end{aligned}$$

Since state $E_5$ corresponds to death, it is clear that we must have $w_5 = 0$. With this prescription, the reduced equations become, after elementary simplification,

$$\begin{aligned} -1.0w_0 + 0.9w_1 \qquad\qquad\qquad\qquad &= \ 0, \\ -0.5w_1 + 0.4w_2 \qquad\qquad\qquad &= \ 0, \\ -0.4w_2 + 0.2w_3 + 0.1w_4 &= -1, \\ 0.4w_2 - 0.5w_3 \qquad\quad &= \ 0, \\ 0.4w_2 \qquad\qquad - 0.5w_4 &= \ 0. \end{aligned}$$

The unique solution is

$$w_0 = 4.5, \quad w_1 = 5.00, \quad w_2 = 6.25, \quad w_3 = w_4 = 5.00.$$

Each female, on the average, spends $w_0 = W_{02} = 4.5$ periods in the childbearing state $E_2$ during her lifetime.

## Exercises

**3.4.1** Find the mean time to reach state 3 starting from state 0 for the Markov chain whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 0.4 & 0.3 & 0.2 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

**3.4.2** Consider the Markov chain whose transition probablity matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 1 & 0 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0 & 1 \end{array} \right\| \end{array}.$$

**(a)** Starting in state 1, determine the probability that the Markov chain ends in state 0.
**(b)** Determine the mean time to absorption.

**3.4.3** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

**(a)** Starting in state 1, determine the probability that the Markov chain ends in state 0.

**(b)** Determine the mean time to absorption.

**3.4.4** A coin is tossed repeatedly until two successive heads appear. Find the mean number of tosses required.

**Hint:** Let $X_n$ be the cumulative number of successive heads. The state space is $0, 1, 2$, and the transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Determine the mean time to reach state 2 starting from state 0 by invoking a first step analysis.

**3.4.5** A coin is tossed repeatedly until either two successive heads appear or two successive tails appear. Suppose the first coin toss results in a head. Find the probability that the game ends with two successive tails.

**3.4.6** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.4 & 0.1 & 0.4 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

**(a)** Starting in state 1, determine the probability that the Markov chain ends in state 0.

**(b)** Determine the mean time to absorption.

**3.4.7** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Starting in state 1, determine the mean time that the process spends in state 1 prior to absorption and the mean time that the process spends in state 2 prior to absorption. Verify that the sum of these is the mean time to absorption.

**3.4.8** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.5 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Starting in state 1, determine the mean time that the process spends in state 1 prior to absorption and the mean time that the process spends in state 2 prior to absorption. Verify that the sum of these is the mean time to absorption.

**3.4.9** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Starting in state 1, determine the probability that the process is absorbed into state 0. Compare this with the $(1,0)$th entry in the matrix powers $\mathbf{P}^2, \mathbf{P}^4, \mathbf{P}^8$, and $\mathbf{P}^{16}$.

## Problems

**3.4.1** Which will take fewer flips, on average: successively flipping a quarter until the pattern *HHT* appears, i.e., until you observe two successive heads followed by a tails; or successively flipping a quarter until the pattern *HTH* appears? Can you explain why these are different?

**3.4.2** A zero-seeking device operates as follows: If it is in state $m$ at time $n$, then at time $n+1$, its position is uniformly distributed over the states $0, 1, \ldots, m-1$. Find the expected time until the device first hits zero starting from state $m$.

**Note:** This is a highly simplified model for an algorithm that seeks a maximum over a finite set of points.

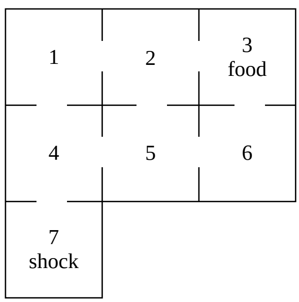
**3.4.3** A zero-seeking device operates as follows: If it is in state $j$ at time $n$, then at time $n+1$, its position is 0 with probability $1/j$, and its position is $k$ (where $k$ is one of the states $1, 2, \ldots, j-1$) with probability $2k/j^2$. Find the expected time until the device first hits zero starting from state $m$.

**3.4.4** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{array} \right\| \end{array}.$$

Starting in state $X_0 = 1$, determine the probability that the process never visits state 2. Justify your answer.

**3.4.5** A white rat is put into compartment 4 of the maze shown here:




It moves through the compartments at random; i.e., if there are $k$ ways to leave a compartment, it chooses each of these with probability $1/k$. What is the probability that it finds the food in compartment 3 before feeling the electric shock in compartment 7?

**3.4.6** Consider the Markov chain whose transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \left\| \begin{array}{ccccc} q & p & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ q & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{array} \right\| \end{array},$$

where $p+q=1$. Determine the mean time to reach state 4 starting from state 0. That is, find $E[T|X_0=0]$, where $T=\min\{n\geq 0; X_n=4\}$.

**Hint:** Let $v_i=E[T|X_0=i]$ for $i=0,1,\ldots,4$. Establish equations for $v_0,v_1,\ldots,v_4$ by using a first step analysis and the boundary condition $v_4=0$. Then, solve for $v_0$.

**3.4.7** Let $X_n$ be a Markov chain with transition probabilities $P_{ij}$. We are given a "discount factor" $\beta$ with $0<\beta<1$ and a cost function $c(i)$, and we wish to determine the total expected discounted cost starting from state $i$, defined by

$$h_i=E\left[\sum_{n=0}^{\infty}\beta^n c(X_n)|X_0=i\right].$$

Using a first step analysis show that $h_i$ satisfies the system of linear equations

$$h_i=c(i)+\beta\sum_j P_{ij}h_j \quad \text{for all states } i.$$

**3.4.8** An urn contains five red and three green balls. The balls are chosen at random, one by one, from the urn. If a red ball is chosen, it is removed. Any green ball that is chosen is returned to the urn. The selection process continues until all of the red balls have been removed from the urn. What is the mean duration of the game?

**3.4.9** An urn contains five red and three yellow balls. The balls are chosen at random, one by one, from the urn. Each ball removed is replaced in the urn by a yellow ball. The selection process continues until all of the red balls have been removed from the urn. What is the mean duration of the game?

**3.4.10** You have five fair coins. You toss them all so that they randomly fall heads or tails. Those that fall tails in the first toss you pick up and toss again. You toss again those that show tails after the second toss, and so on, until all show heads. Let $X$ be the number of coins involved in the *last* toss. Find $\Pr\{X=1\}$.

**3.4.11** An urn contains two red and two green balls. The balls are chosen at random, one by one, and removed from the urn. The selection process continues until all of the green balls have been removed from the urn. What is the probability that a single red ball is in the urn at the time that the last green ball is chosen?

**3.4.12** A Markov chain $X_0,X_1,X_2,\ldots$ has the transition probability matrix

$$\mathbf{P}=\begin{array}{c|ccc|} & 0 & 1 & 2\\ \hline 0 & 0.3 & 0.2 & 0.5\\ 1 & 0.5 & 0.1 & 0.4\\ 2 & 0 & 0 & 1 \end{array}$$

and is known to start in state $X_0=0$. Eventually, the process will end up in state 2. What is the probability that when the process moves into state 2, it does so from state 1?

**Hint:** Let $T = \min\{n \geq 0; X_n = 2\}$, and let

$$z_i = \Pr\{X_{T-1} = 1 | X_0 = i\} \quad \text{for } i = 0, 1.$$

Establish and solve the first step equations

$$\begin{aligned} z_0 &= \qquad 0.3z_0 + 0.2z_1, \\ z_1 &= 0.4 + 0.5z_0 + 0.1z_1. \end{aligned}$$

**3.4.13** A Markov chain $X_0, X_1, X_2, \ldots$ has the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0 & 0 & 1 \end{array} \right\| \end{array}$$

and is known to start in state $X_0 = 0$. Eventually, the process will end up in state 2. What is the probability that the time $T = \min\{n \geq 0; X_n = 2\}$ is an odd number?

**3.4.14** A single die is rolled repeatedly. The game stops the first time that the sum of two successive rolls is either 5 or 7. What is the probability that the game stops at a sum of 5?

**3.4.15** A simplified model for the spread of a rumor goes this way: There are $N = 5$ people in a group of friends, of which some have heard the rumor and the others have not. During any single period of time, two people are selected at random from the group and assumed to interact. The selection is such that an encounter between any pair of friends is just as likely as between any other pair. If one of these persons has heard the rumor and the other has not, then with probability $\alpha = 0.1$ the rumor is transmitted. Let $X_n$ denote the number of friends who have heard the rumor at the end of the $n$th period.

Assuming that the process begins at time 0 with a single person knowing the rumor, what is the mean time that it takes for everyone to hear it?

**3.4.16** An urn contains five tags, of which three are red and two are green. A tag is randomly selected from the urn and replaced with a tag of the opposite color. This continues until only tags of a single color remain in the urn. Let $X_n$ denote the number of red tags in the urn after the $n$th draw, with $X_0 = 3$. What is the probability that the game ends with the urn containing only red tags?

**3.4.17** The *damage* $X_n$ of a system subjected to wear is a Markov chain with the transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 0.7 & 0.3 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 1 \end{array} \right\| \end{array}.$$

The system starts in state 0 and fails when it first reaches state 2. Let $T = \min\{n \geq 0; X_n = 2\}$ be the time of failure. Use a first step analysis to evaluate $\phi(s) = E\left[s^T\right]$ for a fixed number $0 < s < 1$. (This is called the *generating function* of $T$. See Section 3.9.)

**3.4.18** *Time-dependent transition probabilities*. A well-disciplined man, who smokes exactly one half of a cigar each day, buys a box containing $N$ cigars. He cuts a cigar in half, smokes half, and returns the other half to the box. In general, on a day in which his cigar box contains $w$ whole cigars and $h$ half cigars, he will pick one of the $w+h$ smokes at random, each whole and half cigar being equally likely, and if it is a half cigar, he smokes it. If it is a whole cigar, he cuts it in half, smokes one piece, and returns the other to the box. What is the expected value of $T$, the day on which the last whole cigar is selected from the box?

**Hint:** Let $X_n$ be the number of whole cigars in the box after the $n$th smoke. Then, $X_n$ is a Markov chain whose transition probabilities vary with $n$. Define $v_n(w) = E[T|X_n = w]$. Use a first step analysis to develop a recursion for $v_n(w)$ and show that the solution is

$$v_n(w) = \frac{2Nw + n + 2w}{w+1} - \sum_{k=1}^{w} \frac{1}{k},$$

whence

$$E[T] = v_0(N) = 2N - \sum_{k=1}^{N} \frac{1}{k}.$$

**3.4.19** *Computer Challenge*. Let $N$ be a positive integer and let $Z_1, \ldots, Z_N$ be independent random variables, each having the geometric distribution

$$\Pr\{Z = k\} = \left(\frac{1}{2}\right)^k, \quad \text{for } k = 1, 2, \ldots.$$

Since these are discrete random variables, the maximum among them may be unique, or there may be ties for the maximum. Let $p_N$ be the probability that the maximum is unique. How does $p_N$ behave when $N$ is large? (Alternative formulation: You toss $N$ dimes. Those that are heads you set aside; those that are tails you toss again. You repeat this until all of the coins are heads. Then, $p_N$ is the probability that the last toss was of a single coin.)

## 3.5 Some Special Markov Chains

We introduce several particular Markov chains that arise in a variety of applications.

### 3.5.1 The Two-State Markov Chain

Let

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left\| \begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right\| \end{array}, \quad \text{where } 0 < a, b < 1, \tag{3.30}$$

be the transition matrix of a two-state Markov chain.

When $a = 1 - b$ so that the rows of $\mathbf{P}$ are the same, then the states $X_1, X_2, \ldots$ are independent identically distributed random variables with $\Pr\{X_n = 0\} = b$ and $\Pr\{X_n = 1\} = a$. When $a \neq 1 - b$, the probability distribution for $X_n$ varies depending on the outcome $X_{n-1}$ at the previous stage.

For the two-state Markov chain, it is readily verified by induction that the $n$-step transition matrix is given by

$$\mathbf{P}^n = \frac{1}{a+b}\left\| \begin{array}{cc} b & a \\ b & a \end{array} \right\| + \frac{(1-a-b)^n}{a+b}\left\| \begin{array}{cc} a & -a \\ -b & b \end{array} \right\|. \tag{3.31}$$

To verify this general formula, introduce the abbreviations

$$\mathbf{A} = \left\| \begin{array}{cc} b & a \\ b & a \end{array} \right\| \quad \text{and} \quad \mathbf{B} = \left\| \begin{array}{cc} a & -a \\ -b & b \end{array} \right\|$$

so that (3.31) can be written

$$\mathbf{P}^n = (a+b)^{-1}\left[\mathbf{A} + (1-a-b)^n \mathbf{B}\right].$$

Next, check the multiplications

$$\mathbf{AP} = \left\| \begin{array}{cc} b & a \\ b & a \end{array} \right\| \times \left\| \begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right\| = \left\| \begin{array}{cc} b & a \\ b & a \end{array} \right\| = \mathbf{A}$$

and

$$\begin{aligned} \mathbf{BP} &= \left\| \begin{array}{cc} a & -a \\ -b & b \end{array} \right\| \times \left\| \begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right\| \\ &= \left\| \begin{array}{cc} a - a^2 - ab & a^2 - a + ab \\ -b + ab + b^2 & -ab + b - b^2 \end{array} \right\| = (1-a-b)\mathbf{B}. \end{aligned}$$

Now, (3.31) is easily seen to be true when $n = 1$, since then

$$\mathbf{P}^1 = \frac{1}{a+b}\begin{Vmatrix} b & a \\ b & a \end{Vmatrix} + \frac{(1-a-b)}{a+b}\begin{Vmatrix} a & -a \\ -b & b \end{Vmatrix}$$

$$= \frac{1}{a+b}\begin{Vmatrix} b+a-a^2-ab & a-a+a^2+ab \\ b-b+ab+b^2 & a+b-ab-b^2 \end{Vmatrix}$$

$$= \begin{Vmatrix} 1-a & a \\ b & 1-b \end{Vmatrix} = \mathbf{P}.$$

To complete an induction proof, assume that the formula is true for $n$. Then,

$$\begin{aligned} \mathbf{P}^n\mathbf{P} &= (a+b)^{-1}\left[\mathbf{A} + (1-a-b)^n\mathbf{B}\right]\mathbf{P} \\ &= (a+b)^{-1}\left[\mathbf{AP} + (1-a-b)^n\mathbf{BP}\right] \\ &= (a+b)^{-1}\left[\mathbf{A} + (1-a-b)^{n+1}\mathbf{B}\right] = \mathbf{P}^{n+1}. \end{aligned}$$

We have verified that the formula holds for $n+1$. It, therefore, is established for all $n$.

Note that $|1-a-b| < 1$ when $0 < a, b < 1$, and thus $|1-a-b|^n \to 0$ as $n \to \infty$ and

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{Vmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{Vmatrix}. \tag{3.32}$$

This tells us that such a system, in the long run, will be in state 0 with probability $b/(a+b)$ and in state 1 with probability $a/(a+b)$, irrespective of the initial state in which the system started.

For a numerical example, suppose that the items produced by a certain worker are graded as defective or not and that due to trends in raw material quality, whether or not a particular item is defective depends in part on whether or not the previous item was defective. Let $X_n$ denote the quality of the $n$th item with $X_n = 0$ meaning "good" and $X_n = 1$ meaning "defective." Suppose that $\{X_n\}$ evolves as a Markov chain whose transition matrix is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & \;\;1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{Vmatrix} 0.99 & 0.01 \\ 0.12 & 0.88 \end{Vmatrix} \end{matrix}.$$

Defective items would tend to appear in bunches in the output of such a system.

In the long run, the probability that an item produced by this system is defective is given by $a/(a+b) = 0.01/(0.01+0.12) = 0.077$.

### 3.5.2 Markov Chains Defined by Independent Random Variables

Let $\xi$ denote a discrete-valued random variable whose possible values are the nonnegative integers and where $\Pr\{\xi = i\} = a_i \geq 0$, for $i = 0, 1, \ldots,$ and $\sum_{i=0}^{\infty} a_i = 1$. Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ represent independent observations of $\xi$.

We shall now describe three different Markov chains connected with the sequence $\xi_1, \xi_2, \ldots$. In each case, the state space of the process is the set of nonnegative integers.

**Example** *Independent Random Variables* Consider the process $X_n, n = 0, 1, 2, \ldots,$ defined by $X_n = \xi_n$ ($X_0 = \xi_0$ prescribed). Its Markov matrix has the form

$$\mathbf{P} = \left\|\begin{matrix} a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ \vdots & \vdots & \vdots & \end{matrix}\right\|. \tag{3.33}$$

That all rows are identical plainly expresses the fact that the random variable $X_{n+1}$ is independent of $X_n$.

**Example** *Successive Maxima* The partial maxima of $\xi_1, \xi_2, \ldots$ define a second important Markov chain. Let

$$\theta_n = \max\{\xi_1, \ldots, \xi_n\}, \quad \text{for } n = 1, 2, \ldots,$$

with $\theta_0 = 0$. The process defined by $X_n = \theta_n$ is readily seen to be a Markov chain, and the relation $X_{n+1} = \max\{X_n, \xi_{n+1}\}$ allows the transition probabilities to be computed to be

$$\mathbf{P} = \left\|\begin{matrix} A_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & A_1 & a_2 & a_3 & \cdots \\ 0 & 0 & A_2 & a_3 & \cdots \\ 0 & 0 & 0 & A_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{matrix}\right\|, \tag{3.34}$$

where $A_k = a_0 + \cdots + a_k$ for $k = 0, 1, \ldots$.

Suppose $\xi_1, \xi_2, \ldots$ represent successive bids on a certain asset that is offered for sale. Then, $X_n = \max\{\xi_1, \ldots, \xi_n\}$ is the maximum that is bid up to stage $n$. Suppose that the bid that is accepted is the first bid that equals or exceeds a prescribed level $M$. The time of sale is the random variable $T = \min\{n \geq 1; X_n \geq M\}$. A first step analysis shows that the mean $\mu = E[T]$ satisfies

$$\mu = 1 + \mu \Pr\{\xi_1 < M\}, \tag{3.35}$$

or $\mu = 1/\Pr\{\xi_1 \geq M\} = 1/(a_M + a_{M+1} + \cdots)$. The first step analysis invoked in establishing (3.35) considers the two possibilities $\{\xi_1 < M\}$ and $\{\xi_1 \geq M\}$. With this breakdown, the law of total probability justifies the sum

$$E[T] = E[T|\xi_1 \geq M]\Pr\{\xi_1 \geq M\} + E[T|\xi_1 < M]\Pr\{\xi_1 < M\}. \tag{3.36}$$

Clearly, $E[T|\xi_1 \geq M] = 1$, since no further bids are examined in this case. On the other hand, when $\xi_1 < M$, we have the first bid, which was not accepted, plus some future bids. The future bids $\xi_2, \xi_3, \ldots$ have the same probabilistic properties as in the original problem, and they are examined until the first acceptable bid appears. This reasoning leads to $E[T|\xi_1 < M] = 1 + \mu$. Substitution into (3.36) then yields (3.35) as follows:

$$\begin{aligned} E[T] &= 1 \times \Pr\{\xi_1 \geq M\} + (1+\mu)\Pr\{\xi_1 < M\} \\ &= 1 + \mu\Pr\{\xi_1 < M\}. \end{aligned}$$

To restate the argument somewhat differently, one always examines the first bid $\xi_1$. If $\xi_1 < M$, then further bids are examined in a future that is probabilistically similar to the original problem. That is, when $\xi_1 < M$, then on the average $\mu$ bids in addition to $\xi_1$ must be examined before an acceptable bid appears. Equation (3.35) results.

**Example** *Partial Sums* Another important Markov chain arises from consideration of the successive partial sums $\eta_n$ of the $\xi_i$, i.e.,

$$\eta_n = \xi_1 + \cdots + \xi_n, \quad n = 1, 2, \ldots,$$

and by definition, $\eta_0 = 0$. The process $X_n = \eta_n$ is readily seen to be a Markov chain via

$$\begin{aligned} &\Pr\{X_{n+1} = j|X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i\} \\ &\quad = \Pr\{\xi_{n+1} = j - i|\xi_1 = i_1, \xi_2 = i_2 - i_1, \ldots, \xi_n = i - i_{n-1}\} \\ &\quad = \Pr\{\xi_{n+1} = j - i\} \quad \text{(independence of } \xi_1, \xi_2, \ldots) \\ &\quad = \Pr\{X_{n+1} = j|X_n = i\}. \end{aligned}$$

The transition probability matrix is determined by

$$\begin{aligned} \Pr\{X_{n+1} = j|X_n = i\} &= \Pr\{\xi_1 + \cdots + \xi_{n+1} = j|\xi_1 + \cdots + \xi_n = i\} \\ &= \Pr\{\xi_{n+1} = j - i\} \\ &= \begin{cases} a_{j-i} & \text{for } j \geq i, \\ 0 & \text{for } j < i, \end{cases} \end{aligned}$$

where we have used the independence of the $\xi_i$.

Schematically, we have

$$\mathbf{P} = \begin{Vmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix}. \tag{3.37}$$

If the possible values of the random variable $\xi$ are permitted to be the positive and negative integers, then the possible values of $\eta_n$ for each $n$ will be contained among the totality of all integers. Instead of labeling the states conventionally by means of the nonnegative integers, it is more convenient to identify the state space with the totality of integers, since the transition probability matrix will then appear in a more symmetric form. The state space consists then of the values $\ldots -2, -1, 0, 1, 2, \ldots$.

The transition probability matrix becomes

$$\mathbf{P} = \begin{Vmatrix} & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & a_{-1} & a_0 & a_1 & a_2 & a_3 & \cdots \\ \cdots & a_{-2} & a_{-1} & a_0 & a_1 & a_2 & \cdots \\ \cdots & a_{-3} & a_{-2} & a_{-1} & a_0 & a_1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \end{Vmatrix},$$

where $\Pr\{\xi = k\} = a_k$ for $k = 0, \pm 1, \pm 2, \ldots$, and $a_k \geq 0$, $\sum_{k=-\infty}^{+\infty} a_k = 1$.

### 3.5.3 One-Dimensional Random Walks

When we discuss random walks, it is an aid to intuition to speak about the state of the system as the position of a moving "particle."

A one-dimensional random walk is a Markov chain whose state space is a finite or infinite subset $a, a+1, \ldots, b$ of the integers, in which the particle, if it is in state $i$, can in a single transition either stay in $i$ or move to one of the neighboring states $i-1, i+1$. If the state space is taken as the nonnegative integers, the transition matrix of a random walk has the form

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ i \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0 & 1 & 2 & & i-1 & & i & i+1 & \end{array} \\ \begin{Vmatrix} r_0 & p_0 & 0 & \cdots & 0 & \cdots & & & & \\ q_1 & r_1 & p_1 & \cdots & 0 & \cdots & & & & \\ 0 & q_2 & r_2 & \cdots & 0 & \cdots & & & & \\ & & \ddots & & & & & & & \\ & & 0 & & q_i & & r_i & p_i & 0 & \\ & & & \ddots & & & & & & \ddots \end{Vmatrix} \end{array}, \tag{3.38}$$

where $p_i > 0, q_i > 0, r_i \geq 0$, and $q_i + r_i + p_i = 1, i = 1, 2, \ldots (i \geq 1), p_0 \geq 0, r_0 \geq 0$, $r_0 + p_0 = 1$. Specifically, if $X_n = i$, then for $i \geq 1$,

$$\begin{aligned}
&\Pr\{X_{n+1} = i+1 | X_n = i\} = p_i, \\
&\Pr\{X_{n+1} = i-1 | X_n = i\} = q_j, \\
&\Pr\{X_{n+1} = i | X_n = i\} = r_i,
\end{aligned}$$

with the obvious modifications holding for $i = 0$.

The designation "random walk" seems apt, since a realization of the process describes the path of a person (suitably intoxicated) moving randomly one step forward or backward.

The fortune of a player engaged in a series of contests is often depicted by a random walk process. Specifically, suppose an individual (player A) with fortune $k$ plays a game against an infinitely rich adversary and has probability $p_k$ of winning one unit and probability $q_k = 1 - p_k (k \geq 1)$ of losing one unit in the next contest (the choice of the contest at each stage may depend on his fortune), and $r_0 = 1$. The process $X_n$, where $X_n$ represents his fortune after $n$ contests, is clearly a random walk. Note that once the state 0 is reached (i.e., player A is wiped out), the process remains in that state. The event of reaching state $k = 0$ is commonly known as the "gambler's ruin."

If the adversary, player B, also starts with a limited fortune $l$ and player A has an initial fortune $k(k + l = N)$, then we may again consider the Markov chain process $X_n$ representing player A's fortune. However, the states of the process are now restricted to the values $0, 1, 2, \ldots, N$. At any trial, $N - X_n$ is interpreted as player B's fortune. If we allow the possibility of neither player winning in a contest, the transition probability matrix takes the form

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \\ \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & & N \end{array} \\ \left\| \begin{array}{cccccc} 1 & 0 & 0 & 0 & \cdots & \\ q_1 & r_1 & p_1 & 0 & \cdots & \\ 0 & q_2 & r_2 & p_2 & \cdots & \\ & \ddots & & & & \\ & & & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{array} \right\| \end{array}. \tag{3.39}$$

Again $p_i(q_i), i = 1, 2, \ldots, N - 1$, denotes the probability of player A's fortune increasing (decreasing) by 1 at the subsequent trial when his present fortune is $i$, and $r_i$ may be interpreted as the probability of a draw. Note that, in accordance with the Markov chain given in (3.39), when player A's fortune (the state of the process) reaches 0 or $N$, it remains in this same state forever. We say player A is ruined when the state of the process reaches 0, and player B is ruined when the state of the process reaches $N$.

The probability of gambler's ruin (for player A) is derived in the next section by solving a first step analysis. Some more complex functionals on random walk processes are also derived in the next section.

The random walk corresponding to $p_k = p, q_k = 1 - p = q$ for all $k \geq 1$ and $r_0 = 1$ describes the situation of identical contests. There is a definite advantage to player A in each individual trial if $p > q$, and conversely, an advantage to player B if $p < q$. A "fair" contest corresponds to $p = q = \frac{1}{2}$. Suppose the total of both players' fortunes is $N$. Then, the corresponding walk, where $X_n$ is player A's fortune at stage $n$, has the transition probability matrix

$$\mathbf{P} = \begin{array}{c|ccccccc|} & 0 & 1 & 2 & 3 & & N-1 & N \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & q & 0 & p & 0 & \cdots & 0 & 0 \\ 2 & 0 & q & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ N-1 & 0 & 0 & 0 & 0 & \cdots & 0 & p \\ N & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array}. \tag{3.40}$$

Let $u_i = U_{i0}$ be the probability of gambler's ruin starting with the initial fortune $i$. Then, $u_i$ is the probability that the random walk reaches state 0 before reaching state $N$, starting from $X_0 = i$. The first step analysis of Section 3.4, as used in deriving equation (3.26), shows that these ruin probabilities satisfy

$$u_i = pu_{i+1} + qu_{i-1} \quad \text{for } i = 1, \ldots, N-1 \tag{3.41}$$

together with the obvious boundary conditions

$$u_0 = 1 \quad \text{and} \quad u_N = 0.$$

These equations are solved in the next section following a straight-forward but arduous method. There it is shown that the gambler's ruin probabilities corresponding to the transition probability matrix given in (3.40) are

$$\begin{aligned} u_i &= \Pr\{X_n \text{ reaches state 0 before state } N | X_0 = i\} \\ &= \begin{cases} \dfrac{N-i}{N} & \text{when } p = q = \dfrac{1}{2}, \\ \dfrac{(q/p)^i - (q/p)^N}{1 - (q/p)^N} & \text{when } p \neq q. \end{cases} \end{aligned} \tag{3.42}$$

The ruin probabilities $u_i$ given by (3.42) have the following interpretation. In a game in which player A begins with an initial fortune of $i$ units and player B begins with $N - i$ units, the probability that player A loses all his money before player B goes broke is given by $u_i$, where $p$ is the probability that player A wins in a single contest. If player B is infinitely rich ($N \to \infty$), then passing to the limit in (3.42) and using $(q/p)^N \to \infty$ as $N \to \infty$ if $p < q$, while $(q/p)^N \to 0$ if $p > q$, we see that the ruin probabilities become

$$u_i = \begin{cases} 1 & \text{if } p \le q, \\ \left(\dfrac{q}{p}\right)^i & \text{if } p > q. \end{cases} \tag{3.43}$$

(In passing to the limit, the case $p = q = \frac{1}{2}$ must be treated separately.) We see that ruin is certain ($u_i = 1$) against an infinitely rich adversary when the game is unfavorable ($p < q$), and even when the game is fair ($p = q$). In a favorable game ($p > q$), starting with initial fortune $i$, then ruin occurs (player A goes broke) with probability $(q/p)^i$. This ruin probability decreases as the initial fortune $i$ increases. In a favorable game against an infinitely rich opponent, with probability $1 - (q/p)^i$ player A's fortune increases, in the long run, without limit.

More complex gambler's-ruin-type problems find practical relevance in certain models describing the fluctuation of insurance company assets over time.

Random walks are not only useful in simulating situations of gambling but frequently serve as reasonable discrete approximations to physical processes describing the motion of diffusing particles. If a particle is subjected to collisions and random impulses, then its position fluctuates randomly, although the particle describes a continuous path. If the future position (i.e., its probability distribution) of the particle depends only on the present position, then the process $X_t$, where $X_t$ is the position at time $t$, is Markov. A discrete approximation to such a continuous motion corresponds to a random walk. A classical discrete version of Brownian motion (VIII) is provided by the symmetric random walk. By a symmetric random walk on the integers (say all the integers) we mean a Markov chain with state space the totality of all integers and whose transition probability matrix has the elements

$$P_{ij} = \begin{cases} p & \text{if } j = i+1, \\ p & \text{if } j = i-1, \\ r & \text{if } j = i, \\ 0 & \text{otherwise,} \end{cases} \qquad i, j = 0, 1, 2, \ldots,$$

where $p > 0, r \ge 0$, and $2p + r = 1$. Conventionally, "simple random walk" refers only to the case $r = 0, p = \frac{1}{2}$.

The classical simple random walk in $n$ dimensions admits the following formulation. The state space is identified with the set of all integral lattice points in $E^n$ (Euclidean $n$ space); that is, a state is an $n$-tuple $k = (k_1, k_2, \ldots, k_n)$ of integers. The transition probability matrix is defined by

$$\mathbf{P_{kl}} = \begin{cases} \dfrac{1}{2n} & \text{if } \sum_{i=0}^{n} |l_i - k_i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Analogous to the one-dimensional case, the simple random walk in $E^n$ represents a discrete version of $n$-dimensional Brownian motion.

### 3.5.4 Success Runs

Consider a Markov chain on the nonnegative integers with transition probability matrix of the form

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \end{array} \\ \left\| \begin{array}{cccccc} p_0 & q_0 & 0 & 0 & 0 & \cdots \\ p_1 & r_1 & q_1 & 0 & 0 & \cdots \\ p_2 & 0 & r_2 & q_2 & 0 & \cdots \\ p_3 & 0 & 0 & r_3 & q_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{array} \right\| \end{array}, \tag{3.44}$$

where $q_i > 0$, $p_i > 0$, and $p_i + q_i + r_i = 1$ for $i = 0, 1, 2, \ldots$. The zero state plays a distinguished role in that it can be reached in one transition from any other state, while state $i + 1$ can be reached only from state $i$.

This example arises surprisingly often in applications and at the same time is very easy to compute with. We will frequently illustrate concepts and results in terms of it.

A special case of this transition matrix arises when one is dealing with success runs resulting from repeated trials, each of which admits two possible outcomes, success $S$ or failure $F$. More explicitly, consider a sequence of trials with two possible outcomes, $S$ or $F$. Moreover, suppose that in each trial, the probability of $S$ is $\alpha$ and the probability of $F$ is $\beta = 1 - \alpha$. We say a success run of length $r$ happened at trial $n$ if the outcomes in the preceding $r + 1$ trials, including the present trial as the last, were respectively $F, S, S, \ldots, S$. Let us now label the present state of the process by the length of the success run currently under way. In particular, if the last trial resulted in a failure, then the state is zero. Similarly, when the preceding $r + 1$ trials in order have the outcomes $F, S, S, \ldots, S$, the state variable would carry the label $r$. The process is clearly Markov (since the individual trials were independent of each other), and its transition matrix has the form (3.44), where

$$p_n = \beta, \quad r_n = 0, \quad \text{and} \quad q_n = \alpha \quad \text{for } n = 0, 1, 2, \ldots.$$

A second example is furnished by the *current age* in a *renewal process*. Consider a light bulb whose lifetime, measured in discrete units, is a random variable $\xi$, where

$$\Pr\{\xi = k\} = a_k > 0 \quad \text{for } k = 1, 2, \ldots, \sum_{k=1}^{\infty} a_k = 1.$$

Let each bulb be replaced by a new one when it burns out. Suppose the first bulb lasts until time $\xi_1$, the second bulb until time $\xi_1 + \xi_2$, and the $n$th bulb until time $\xi_1 + \cdots + \xi_n$, where the individual lifetimes $\xi_1, \xi_2, \ldots$ are independent random variables each having the same distribution as $\xi$. Let $X_n$ be the age of the bulb in service at time $n$. This current age process is depicted in Figure 3.2.

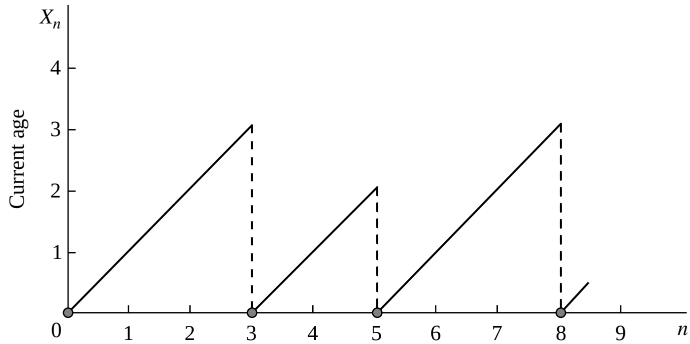



**Figure 3.2** The current age $X_n$ in a renewal process. Here, $\xi_1 = 3, \xi_2 = 2$, and $\xi_3 = 3$.

By convention, we set $X_n = 0$ at the time of a failure.

The current age is a success run Markov process for which

$$p_k = \frac{a_{k+1}}{a_{k+1} + a_{k+2} + \cdots}, \quad r_k = 0, q_k = 1 - p_k, \quad \text{for } k = 0, 1, \ldots. \tag{3.45}$$

We reason as follows: The age process reverts to zero upon failure of the item in service. Given that the age of the item in current service is $k$, then failure occurs in the next time period with *conditional probability* $p_k = a_{k+1}/(a_{k+1} + a_{k+2} + \cdots)$. Given that the item has survived $k$ periods, it survives at least to the next period with the remaining probability $q_k = 1 - p_k$.

Renewal processes are extensively discussed in Chapter 7.

## Exercises

**3.5.1** The probability of the thrower winning in the dice game called "craps" is $p = 0.4929$. Suppose Player A is the thrower and begins the game with \$5, and Player B, his opponent, begins with \$10. What is the probability that Player A goes bankrupt before Player B? Assume that the bet is \$1 per round.

**Hint:** Use equation (3.42).

**3.5.2** Determine the gambler's ruin probability for Player A when both players begin with \$50, bet \$1 on each play, and where the win probability for Player A in each game is

**(a)** $p = 0.49292929$

**(b)** $p = 0.5029237$

(See Chapter 2, Section 2.2.)

What are the gambler's ruin probabilities when each player begins with \$500?

**3.5.3** Determine $\mathbf{P}^n$ for $n = 2, 3, 4, 5$ for the Markov chain whose transition probability matrix is

$$\mathbf{P} = \begin{Vmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{Vmatrix}.$$

**3.5.4** A coin is tossed repeatedly until three heads in a row appear. Let $X_n$ record the current number of successive heads that have appeared. That is, $X_n = 0$ if the $n$th toss resulted in tails; $X_n = 1$ if the $n$th toss was heads and the $(n-1)$st toss was tails; and so on. Model $X_n$ as a success runs Markov chain by specifying the probabilities $p_i$ and $q_i$.

**3.5.5** Suppose that the items produced by a certain process are each graded as defective or good and that whether or not a particular item is defective or good depends on the quality of the previous item. To be specific, suppose that a defective item is followed by another defective item with probability 0.80, whereas a good item is followed by another good item with probability 0.95. Suppose that the initial (zeroth) item is good. Using equation (3.31), determine the probability that the eighth item is good, and verify this by computing the eighth matrix power of the transition probability matrix.

**3.5.6** A baseball trading card that you have for sale may be quite valuable. Suppose that the successive bids $\xi_1, \xi_2, \ldots$ that you receive are independent random variables with the geometric distribution

$$\Pr\{\xi = k\} = 0.01(0.99)^k \quad \text{for } k = 0, 1, \ldots.$$

If you decide to accept any bid over \$100, how many bids, on the average, will you receive before an acceptable bid appears?

**Hint:** Review the discussion surrounding equation (3.35).

**3.5.7** Consider the random walk Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \end{array}.$$

Starting in state 1, determine the probability that the process is absorbed into state 0. Do this first using the basic first step approach of equations (3.21) and (3.22) and second using the particular results for a random walk given in equation (3.42).

**3.5.8** As a special case, consider a discrete-time queueing model in which at most a single customer arrives in any period and at most a single customer completes service. Suppose that in any single period, a single customer arrives with probability $\alpha$, and no customers arrive with probability $1 - \alpha$. Provided that there are customers in the system, in a single period a single customer completes service with probability $\beta$, and no customers leave with probability $1-\beta$. Then $X_n$, the number of customers in the system at the end-of-period $n$, is a random walk in the sense of Section 3.5.3. Referring to equation (3.38), specify the transition probabilities $p_i, q_i$, and $r_i$ for $i = 0, 1, \ldots$.

**3.5.9** In a simplified model of a certain television game show, suppose that the contestant, having won $k$ dollars, will at the next play have $k+1$ dollars with probability $q$ and be put out of the game and leave with nothing with probability $p = 1-q$. Suppose that the contestant begins with one dollar. Model her winnings after $n$ plays as a success runs Markov chain by specifying the transition probabilities $p_i, q_i$, and $r_i$ in equation (3.44).

## Problems

**3.5.1** As a special case of the successive maxima Markov chain whose transition probabilities are given in equation (3.34), consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} a_0 & a_1 & a_2 & a_3 \\ 0 & a_0+a_1 & a_2 & a_3 \\ 0 & 0 & a_0+a_1+a_2 & a_3 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Starting in state 0, show that the mean time until absorption is $v_0 = 1/a_3$.

**3.5.2** A component of a computer has an active life, measured in discrete units, that is a random variable $T$, where $\Pr\{T = k\} = a_k$ for $k = 1, 2, \ldots$. Suppose one starts with a fresh component, and each component is replaced by a new component upon failure. Let $X_n$ be the age of the component in service at time $n$. Then, $\{X_n\}$ is a success runs Markov chain.

**(a)** Specify the probabilities $p_i$ and $q_i$.

**(b)** A "planned replacement" policy calls for replacing the component upon its failure or upon its reaching age $N$, whichever occurs first. Specify the success runs probabilities $p_i$ and $q_i$ under the planned replacement policy.

**3.5.3** *A Batch Processing Model.* Customers arrive at a facility and wait there until $K$ customers have accumulated. Upon the arrival of the $K$th customer, all are instantaneously served, and the process repeats. Let $\xi_0, \xi_1, \ldots$ denote the arrivals in successive periods, assumed to be independent random variables whose distribution is given by

$$\Pr\{\xi_k = 0\} = \alpha, \quad \Pr\{\xi_k = 1\} = 1-\alpha,$$

where $0 < \alpha < 1$. Let $X_n$ denote the number of customers in the system at time $n$. Then, $\{X_n\}$ is a Markov chain on the states $0, 1, \ldots, K-1$. With $K = 3$, give the transition probability matrix for $\{X_n\}$. Be explicit about any assumptions you make.

**3.5.4** Martha has a fair die with the usual six sides. She throws the die and records the number. She throws the die again and adds the second number to the first. She repeats this until the cumulative sum of all the tosses first exceeds 10. What is the probability that she stops at a cumulative sum of 13?

**3.5.5** Let $\{X_n\}$ be a random walk for which zero is an absorbing state and such that from a positive state, the process is equally likely to go up or down one unit. The transition probability matrix is given by (3.38) with $r_0 = 1$ and $p_i = q_i = \frac{1}{2}$ for $i \geq 1$. (a) Show that $\{X_n\}$ is a nonnegative martingale. (b) Use the maximal inequality in Chapter 2, (2.53) to limit the probability that the process ever gets as high as $N > 0$.

## 3.6 Functionals of Random Walks and Success Runs

Consider first the random walk on $N+1$ states whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & N \end{array} \\ \left\| \begin{array}{cccccc} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{array} \right\| \end{array}.$$

"Gambler's ruin" is the event that the process reaches state 0 before reaching state $N$. This event can be stated more formally if we introduce the concept of *hitting time*. Let $T$ be the (random) time that the process first reaches, or hits, state 0 or $N$. In symbols,

$$T = \min\{n \geq 0; X_n = 0 \text{ or } X_n = N\}.$$

The random time $T$ is shown in Figure 3.3 in a typical case.

In terms of $T$, the event written as $X_T = 0$ is the event of gambler's ruin, and the probability of this event starting from the initial state $k$ is

$$u_k = \Pr\{X_T = 0 | X_0 = k\}.$$

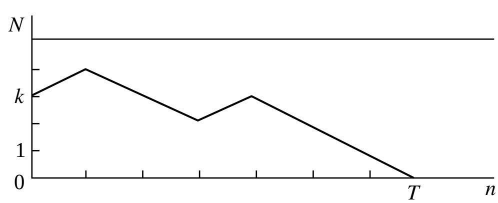



**Figure 3.3** The hitting time to 0 or $N$. As depicted here, state 0 was reached first.

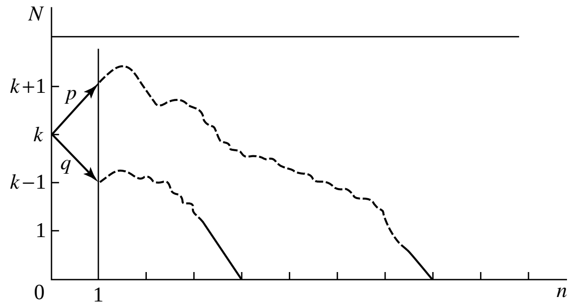



**Figure 3.4** First step analysis for the gambler's ruin problem.

Figure 3.4 shows the first step analysis that leads to the equation

$$u_k = pu_{k+1} + qu_{k-1}, \quad \text{for } k = 1, \ldots, N-1, \tag{3.46}$$

with the obvious boundary conditions

$$u_0 = 1, \quad u_N = 0.$$

Equation (3.46) yields to straightforward but tedious manipulations. Because the approach has considerable generality and arises frequently, it is well worth pursuing in this simplest case.

We begin the solution by introducing the differences $x_k = u_k - u_{k-1}$ for $k = 1, \ldots, N$. Using $p + q = 1$ to write $u_k = (p+q)u_k = pu_k + qu_k$, equation (3.46) becomes

$$\begin{aligned}
k = 1; &\qquad 0 = p(u_2 - u_1) - q(u_1 - u_0) = px_2 - qx_1;\\
k = 2; &\qquad 0 = p(u_3 - u_2) - q(u_2 - u_1) = px_3 - qx_2;\\
k = 3; &\qquad 0 = p(u_4 - u_3) - q(u_3 - u_2) = px_4 - qx_3;\\
&\vdots\\
k = N-1; &\qquad 0 = p(u_N - u_{N-1}) - q(u_{N-1} - u_{N-2}) = px_N - qx_{N-1};
\end{aligned}$$

or

$$\begin{aligned}
x_2 &= (q/p)x_1,\\
x_3 &= (q/p)x_2 = (q/p)^2 x_1,\\
x_4 &= (q/p)x_3 = (q/p)^3 x_1\\
&\vdots\\
x_k &= (q/p)x_{k-1} = (q/p)^{k-1} x_1,\\
&\vdots\\
x_N &= (q/p)x_{N-1} = (q/p)^{N-1} x_1.
\end{aligned}$$

We now recover $u_0, u_1, \ldots, u_N$ by invoking the conditions $u_0 = 1, u_N = 0$ and summing the $x_k$'s:

$$
\begin{aligned}
&x_1 = u_1 - u_0 = u_1 - 1, && \\
&x_2 = u_2 - u_1, & x_1 + x_2 &= u_2 - 1, \\
&x_3 = u_3 - u_2, & x_1 + x_2 + x_3 &= u_3 - 1, \\
&\quad\vdots & &\vdots \\
&x_k = u_k - u_{k-1}, & x_1 + \cdots + x_k &= u_k - 1, \\
&\quad\vdots & &\vdots \\
&x_N = u_N - u_{N-1} = -u_{N-1}, & x_1 + \cdots + x_N &= u_N - 1 = -1.
\end{aligned}
$$

The equation for general $k$ gives

$$
\begin{aligned}
u_k &= 1 + x_1 + x_2 + \cdots + x_k \\
&= 1 + x_1 + (q/p)x_1 + \cdots + (q/p)^{k-1}x_1 \\
&= 1 + [1 + (q/p) + \cdots + (q/p)^{k-1}]x_1,
\end{aligned}
\tag{3.47}
$$

which expresses $u_k$ in terms of the as yet undetermined $x_1$. But $u_N = 0$ gives

$$0 = 1 + [1 + (q/p) + \cdots + (q/p)^{N-1}]x_1,$$

or

$$x_1 = -\frac{1}{1 + (q/p) + \cdots + (q/p)^{N-1}},$$

which substituted into (3.47) gives

$$u_k = 1 - \frac{1 + (q/p) + \cdots + (q/p)^{k-1}}{1 + (q/p) + \cdots + (q/p)^{N-1}}.$$

The geometric series sums to

$$1 + (q/p) + \cdots + (q/p)^{k-1} = \begin{cases} k & \text{if } p = q = \dfrac{1}{2}, \\ \dfrac{1-(q/p)^k}{1-(q/p)} & \text{if } p \neq q, \end{cases}$$

whence

$$u_k = \begin{cases} 1 - (k/N) = (N-k)/N & \text{when } p = q = \dfrac{1}{2}, \\ 1 - \dfrac{1-(q/p)^k}{1-(q/p)^N} = \dfrac{(q/p)^k - (q/p)^N}{1-(q/p)^N} & \text{when } p \neq q. \end{cases} \tag{3.48}$$

A similar approach works to evaluate the mean duration

$$v_i = E[T|X_0 = i]. \tag{3.49}$$

The time $T$ is composed of a first step plus the remaining steps. With probability $p$, the first step is to state $i+1$, and then, the remainder, on the average, is $v_{i+1}$ additional steps. With probability $q$, the first step is to $i-1$, and then, on the average, there are $v_{i-1}$ further steps. Thus, for the mean duration, a first step analysis leads to the equation

$$v_i = 1 + pv_{i+1} + qv_{i-1} \quad \text{for } i = 1, \ldots, N-1. \tag{3.50}$$

Of course, the game ends in states 0 and $N$, and thus,

$$v_0 = 0, \quad v_N = 0.$$

We will solve equation (3.50) when $p = q = \frac{1}{2}$. The solution for other values of $p$ proceeds in a similar manner, and the solution for a general random walk is given later in this section.

Again, we introduce the differences $x_k = v_k - v_{k-1}$ for $k = 1, \ldots, N$, writing (3.50) in the form

$$\begin{aligned}
k &= 1; & -1 &= \frac{1}{2}(v_2 - v_1) - \frac{1}{2}(v_1 - v_0) = \frac{1}{2}x_2 - \frac{1}{2}x_1;\\
k &= 2; & -1 &= \frac{1}{2}(v_3 - v_2) - \frac{1}{2}(v_2 - v_1) = \frac{1}{2}x_3 - \frac{1}{2}x_2;\\
k &= 3; & -1 &= \frac{1}{2}(v_4 - v_3) - \frac{1}{2}(v_3 - v_2) = \frac{1}{2}x_4 - \frac{1}{2}x_3;\\
&\vdots\\
k &= N-1; & -1 &= \frac{1}{2}(v_N - v_{N-1}) - \frac{1}{2}(v_{N-1} - v_{N-2}) = \frac{1}{2}x_N - \frac{1}{2}x_{N-1}.
\end{aligned}$$

The right side forms a collapsing sum. Upon adding, we obtain

$$\begin{aligned}
k &= 1; & -1 &= \frac{1}{2}x_2 - \frac{1}{2}x_1;\\
k &= 2; & -2 &= \frac{1}{2}x_3 - \frac{1}{2}x_1;\\
k &= 3; & -3 &= \frac{1}{2}x_4 - \frac{1}{2}x_1;\\
&\vdots\\
k &= N-1; & -(N-1) &= \frac{1}{2}x_N - \frac{1}{2}x_1.
\end{aligned}$$

The general line gives $x_k = x_1 - 2(k-1)$ for $k = 2, 3, \ldots, N$. We return to the $v_k$'s by means of

$$\begin{aligned}
&x_1 = v_1 - v_0 = v_1; \\
&x_2 = v_2 - v_1; && x_1 + x_2 = v_2; \\
&x_3 = v_3 - v_2; && x_1 + x_2 + x_3 = v_3; \\
&\quad\vdots \\
&x_k = v_k - v_{k-1}; && x_1 + \cdots + x_k = v_k;
\end{aligned}$$

or

$$v_k = kv_1 - 2[1 + 2 + \cdots + (k-1)] = kv_1 - k(k-1), \tag{3.51}$$

which gives $v_k$ in terms of the as yet unknown $v_1$. We impose the boundary condition $v_N = 0$ to obtain $0 = Nv_1 - N(N-1)$ or $v_1 = (N-1)$. Substituting this into (3.51), we obtain

$$v_k = k(N-k), \quad k = 0, 1, \ldots, N, \tag{3.52}$$

for the mean duration of the game. Note that the mean duration is greatest for initial fortunes $k$ that are midway between the boundaries 0 and $N$, as we would expect.

### *3.6.1 The General Random Walk*

We give the results of similar derivations on the random walk whose transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & N \end{array} \\ \left\| \begin{array}{cccccc} 1 & 0 & 0 & 0 & \cdots & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{array} \right\| \end{array},$$

where $q_k > 0$ and $p_k > 0$ for $k = 1, \ldots, N-1$. Let $T = \min\{n \geq 0; X_n = 0 \text{ or } X_n = N\}$ be the hitting time to states 0 and $N$.

**Example** As a sample calculation of these functionals, we consider the special case in which the transition probabilities are the same from row to row. That is, we study the random walk whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & N \end{array} \\ \left\| \begin{array}{cccccc} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & r & p & 0 & \cdots & 0 \\ 0 & q & r & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & & 1 \end{array} \right\| \end{array},$$

with $p > 0,\ q > 0$, and $p + q + r = 1$. Let us abbreviate by setting $\theta = (q/p)$, and then $\rho_k$, as defined in (3.63), simplifies according to

$$\rho_k = \frac{q_1 q_2 \cdots q_k}{p_1 p_2 \cdots p_k} = \left(\frac{q}{p}\right)^k = \theta^k \quad \text{for } k = 1, \ldots, N-1.$$

The probability of gambler's ruin, as defined in (3.61) and evaluated in (3.62), becomes

$$\begin{aligned} u_k &= \Pr\{X_T = 0 | X_0 = k\} \\ &= \frac{\theta^k + \cdots + \theta^{N-1}}{1 + \theta + \cdots + \theta^{N-1}} \\ &= \begin{cases} \dfrac{\theta^k - \theta^N}{1 - \theta^N} & \text{if } \theta \equiv (q/p) \neq 1, \\ \dfrac{N-k}{N} & \text{if } \theta \equiv (q/p) = 1. \end{cases} \end{aligned}$$

This, of course, agrees with the answer given in (3.48).

We turn to evaluating the mean time

$$v_k = E[T | X_0 = k] \quad \text{for } k = 1, \ldots, N = 1$$

by first substituting $\rho_i = \theta^i$ into (3.67) to obtain

$$\begin{aligned} \Phi_i &= \left(\frac{1}{q} + \frac{1}{q\theta} + \cdots + \frac{1}{q\theta^{i-1}}\right)\theta^i \\ &= \frac{1}{q}(\theta^i + \theta^{i-1} + \cdots + \theta) \\ &= \frac{1}{p}(1 + \theta + \cdots + \theta^{i-1}) \end{aligned}$$

$$= \begin{cases} \dfrac{i}{p} & \text{when } p = q(\theta = 1), \\ \dfrac{1}{p}\left(\dfrac{1-\theta^i}{1-\theta}\right) & \text{when } p \neq q(\theta \neq 1). \end{cases}$$

Now observe that

$$\begin{aligned} 1 + \rho_1 + \cdots + \rho_{i-1} &= 1 + \theta + \cdots + \theta^{i-1} \\ &= p\Phi_i \end{aligned}$$

so that (3.66) reduces to

$$v_k = \frac{\Phi_k}{\Phi_N}(\Phi_1 + \cdots + \Phi_{N-1}) - (\Phi_1 + \cdots + \Phi_{k-1}). \tag{3.53}$$

In order to continue, we need to simplify the terms of the form $\Phi_1 + \cdots + \Phi_{j-1}$. We consider the two cases $\theta \equiv (q/p) = 1$ and $\theta \equiv (q/p) \neq 1$ separately.

When $p = q$, or equivalently, $\theta = 1$, then $\Phi_{\mathrm{i}} = i/p$, whence

$$\Phi_1 + \cdots + \Phi_{j-1} = \frac{1 + \cdots + (j-1)}{p} = \frac{j(j-1)}{2p},$$

which inserted into (3.53) gives

$$\begin{aligned} v_i &\equiv E[T|X_0 = i] \\ &= \frac{i}{N}\left[\frac{N(N-1)}{2p}\right] - \frac{i(i-1)}{2p} \\ &= \frac{i(N-i)}{2p} \quad \text{if } p = q. \end{aligned} \tag{3.54}$$

When $p = \frac{1}{2}$, then $v_i = i(N-i)$ in agreement with (3.52).

When $p \neq q$, so that $\theta \equiv q/p \neq 1$, then

$$\Phi_i = \frac{1}{p}\left(\frac{1-\theta^i}{1-\theta}\right),$$

whence

$$\begin{aligned} \Phi_1 + \cdots + \Phi_{j-1} &= \frac{1}{p(1-\theta)}\left[(j-1) - \left(\theta + \theta^2 + \cdots + \theta^{j-1}\right)\right] \\ &= \frac{1}{p(1-\theta)}\left[(j-1) - \theta\left(\frac{1-\theta^{j-1}}{1-\theta}\right)\right], \end{aligned}$$

and

$$\begin{aligned} v_i &= E[T|X_0 = i] \\ &= \left(\frac{1-\theta^i}{1-\theta^N}\right)\frac{1}{p(1-\theta)}\left[N - \left(\frac{1-\theta^N}{1-\theta}\right)\right] - \frac{1}{p(1-\theta)}\left[i - \left(\frac{1-\theta^i}{1-\theta}\right)\right] \\ &= \frac{1}{p(1-\theta)}\left[N\left(\frac{1-\theta^i}{1-\theta^N}\right) - i\right], \end{aligned}$$

when $\theta \equiv (q/p) \neq 1$.

Finally, we evaluate $W_{ik}$, expressed verbally as the mean number of visits to state $k$ starting from $X_0 = i$ and defined formally in (3.68). Again, we consider the two cases $\theta \equiv (q/p) = 1$ and $\theta \equiv (q/p) \neq 1$.

When $\theta = 1$, then $\rho_j = \theta^j = 1$ and $1 + \cdots + \rho_{i-1} = i$, $\rho_k + \cdots + \rho_{N-1} = N - k$, and (3.69) simplifies to

$$\begin{aligned} W_{ik} &= \begin{cases} \dfrac{i(N-k)}{qN} & \text{for } 0 < i \leq k < N, \\ \dfrac{1}{q}\left[\dfrac{i(N-k)}{N} - (i-k)\right] = \dfrac{k(N-i)}{qN} & \text{for } 0 < k < i < N, \end{cases} \\ &= \frac{i(N-k)}{qN} - \frac{\max\{o, i-k\}}{q}. \end{aligned} \tag{3.55}$$

When $\theta = (q/p) \neq 1$, then $\rho_j = \theta^j$ and

$$\begin{aligned} 1 + \cdots + \rho_{i-1} &= \frac{1-\theta^i}{1-\theta}, \\ \rho_k + \cdots + \rho_{N-1} &= \frac{\theta^k - \theta^N}{1-\theta}, \end{aligned}$$

and

$$q\rho_{k-1} = p\rho_k = p\theta^k.$$

In this case, (3.69) simplifies to

$$W_{ik} = \frac{\left(1-\theta^i\right)\left(\theta^k - \theta^N\right)}{(1-\theta)\left(1-\theta^N\right)}\left(\frac{1}{p\theta^k}\right) \quad \text{for } 0 < i \leq k < N,$$

and

$$W_{ik} = \frac{\left(1-\theta^k\right)\left(\theta^i - \theta^N\right)}{(1-\theta)\left(1-\theta^N\right)}\left(\frac{1}{p\theta^k}\right) \quad \text{for } 0 < k < i < N.$$

We may write the expression for $W_{ik}$ in a single line by introducing the notation $(i-k)^+ = \max\{0, i-k\}$. Then,

$$W_{ik} = \frac{\left(1-\theta^i\right)\left(1-\theta^{N-k}\right)}{p(1-\theta)\left(1-\theta^N\right)} - \frac{1-\theta^{(i-k)^+}}{p(1-\theta)}. \tag{3.56}$$

### 3.6.2 Cash Management

Short-term cash management is the review and control of a corporation's cash balances, short-term loan balances, and short-term marketable security holdings. The objective is to maintain the smallest cash balances that are adequate to meet future disbursements. The corporation cashier tries to eliminate idle cash balances (e.g., by reducing short-term loans or buying treasury bills) but to cover potential cash shortages (by selling treasury bills or increasing short-term loans). The analogous problem for an individual is to maintain an optimal balance between a checking and a savings account.

In the absence of intervention, the corporation's cash level fluctuates randomly as the result of many relatively small transactions. We model this by dividing time into successive, equal length periods, each of short duration, and by assuming that from period to period, the cash level moves up or down one unit, each with a probability of one-half. Let $X_n$ be the cash on hand in period $n$. We are assuming that $\{X_n\}$ is the random walk in which

$$\Pr\{X_{n+1} = k \pm 1 | X_n = k\} = \frac{1}{2}.$$

The cashier's job is to intervene if the cash level ever gets too low or too high. We consider cash management strategies that are specified by two parameters, $s$ and $\mathcal{S}$, where $0 < s < \mathcal{S}$. The policy is as follows: If the cash level ever drops to zero, then sell sufficient treasury bills to replenish the cash level up to $s$. If the cash level ever increases up to $\mathcal{S}$, then invest in treasury bills in order to reduce the cash level to $s$. A typical sequence of cash levels $\{X_n\}$ when $s = 2$ and $\mathcal{S} = 5$ is depicted in Figure 3.5.

We see that the cash level fluctuates in a series of statistically similar cycles, each cycle beginning with $s$ units of cash on hand and ending at the next intervention,

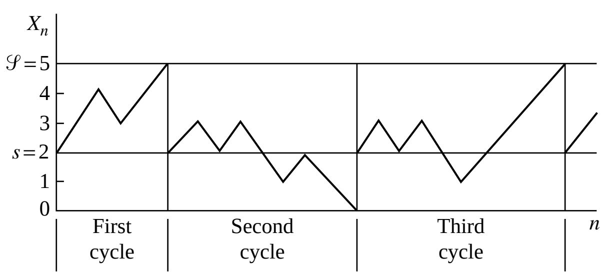



**Figure 3.5** Several typical cycles in a cash inventory model.

whether a replenishment or reduction in cash. We begin our study by evaluating the mean length of a cycle and the mean total unit periods of cash on hand during a cycle. Later, we use these quantities to evaluate the long run performance of the model.

Let $T$ denote the random time at which the cash on hand first reaches the level $\mathcal{S}$ or 0. That is, $T$ is the time of the first transaction. Let $v_s = E[T|X_0 = s]$ be the mean time to the first transaction, or the mean cycle length. From (3.52), we have

$$v_s = s(\mathcal{S} - s). \tag{3.57}$$

Next, fix an arbitrary state $k(0 < k < \mathcal{S})$ and let $W_{sk}$ be the mean number of visits to $k$ up to time $T$ for a process starting at $X_0 = s$. From (3.55), we have

$$W_{sk} = 2\left[\frac{s}{\mathcal{S}}(\mathcal{S} - k) - (s - k)^+\right]. \tag{3.58}$$

Using this we obtain the mean total unit periods of cash on hand up to time $T$ starting from $X_0 = s$ by weighting $W_{sk}$ by $k$ and summing according to

$$\begin{aligned} W_s &= \sum_{k=1}^{\mathcal{S}-1} kW_{sk} \\ &= 2\left\{\frac{s}{\mathcal{S}}\sum_{k=1}^{\mathcal{S}-1} k(\mathcal{S} - k) - \sum_{k=1}^{s-1} k(s-k)\right\} \\ &= 2\left\{\frac{s}{\mathcal{S}}\left[\frac{\mathcal{S}(\mathcal{S}-1)(\mathcal{S}+1)}{6}\right] - \frac{s(s-1)(s+1)}{6}\right\}^* \\ &= \frac{s}{3}\left[\mathcal{S}^2 - s^2\right]. \end{aligned} \tag{3.59}$$

Having obtained these single cycle results, we will use them to evaluate the long run behavior of the model. Note that each cycle starts from the cash level $s$, and thus, the cycles are statistically independent. Let $K$ be the fixed cost of each transaction. Let $T_i$ be the duration of the $i$th cycle and let $R_i$ be the total opportunity cost of holding cash on hand during that time. Over $n$ cycles the average cost per unit time is

$$\text{Average cost} = \frac{nK + R_1 + \cdots + R_n}{T_1 + \cdots + T_n}.$$

Next, divide the numerator and denominator by $n$, let $n \to \infty$, and invoke the law of large numbers to obtain

$$\text{Long run average cost} = \frac{K + E[R_i]}{E[T_i]}.$$

* Use the sum $\sum_{k=1}^{a-1} k(a-k) = \frac{1}{6}a(a+1)(a-1)$.

Let $r$ denote the opportunity cost per unit time of cash on hand. Then, $E[R_i] = rW_s$, while $E[T_i] = v_s$. Since these quantities were determined in (3.57) and (3.59), we have

$$\text{Long run average cost} = \frac{K + (1/3)rs\left(\mathcal{S}^2 - s^2\right)}{s(\mathcal{S} - s)}. \tag{3.60}$$

In order to use calculus to determine the cost-minimizing values for $\mathcal{S}$ and $s$, it simplifies matters if we introduce the new variable $x = s/\mathcal{S}$. Then, (3.60) becomes

$$\text{Long run average cost} = \frac{K + (1/3)r\mathcal{S}^3 x\left(1 - x^2\right)}{\mathcal{S}^2 x(1 - x)},$$

whence

$$\frac{\text{d(average cost)}}{\text{d}x} = 0 = -\frac{K(1-2x)}{\mathcal{S}^2 x^2 (1-x)^2} + \frac{1}{3} r\mathcal{S},$$

$$\frac{\text{d(average cost)}}{\text{d}\mathcal{S}} = 0 = -\frac{2K}{\mathcal{S}^3 x(1-x)} + \frac{r(1+x)}{3},$$

which yield

$$x_{\text{opt}} = \frac{1}{3} \quad \text{and} \quad \mathcal{S}_{\text{opt}} = 3s_{\text{opt}} = 3\sqrt[3]{\frac{3K}{4r}}.$$

Implementing the cash management strategy with the values $s_{\text{opt}}$ and $\mathcal{S}_{\text{opt}}$ results in the optimal balance between transaction costs and the opportunity cost of holding cash on hand.

### *3.6.3 The Success Runs Markov Chain*

Consider the success runs Markov chain on $N + 1$ states whose transition matrix is

$$\mathbf{P} = \begin{array}{r} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & N \end{array} \\ \left\| \begin{array}{cccccc} 1 & 0 & 0 & 0 & \cdots & 0 \\ p_1 & r_1 & q_1 & 0 & \cdots & 0 \\ p_2 & 0 & r_2 & q_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ p_{N-1} & 0 & 0 & 0 & \cdots & q_{N-1} \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{array} \right\| \end{array}.$$

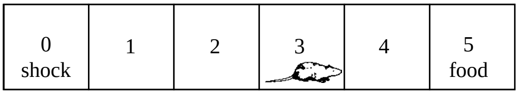
Note that states 0 and $N$ are absorbing; once the process reaches one of these two states it remains there.

Let $T$ be the hitting time to states 0 or $N$,

$$T = \min\{n \geq 0; X_n = 0 \text{ or } X_n = N\}.$$

## Exercises

**3.6.1** A rat is put into the linear maze as shown:




**(a)** Assume that the rat is equally likely to move right or left at each step. What is the probability that the rat finds the food before getting shocked?

**(b)** As a result of learning, at each step the rat moves to the right with probability $p > \frac{1}{2}$ and to the left with probability $q = 1 - p < \frac{1}{2}$. What is the probability that the rat finds the food before getting shocked?

**3.6.2** Customer accounts receivable at Smith Company are classified each month according to

0: Current
1: 30–60 days past due
2: 60–90 days past due
3: Over 90 days past due

Consider a particular customer account and suppose that it evolves month to month as a Markov chain $\{X_n\}$ whose transition probability matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 0.9 & 0.1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.3 & 0 & 0 & 0.7 \\ 0.2 & 0 & 0 & 0.8 \end{array} \right\| \end{array}.$$

Suppose that a certain customer's account is now in state 1: 30–60 days past due. What is the probability that this account will be paid (and thereby enter state 0: Current) before it becomes over 90 days past due? That is, let $T = \min\{n \geq 0; X_n = 0 \text{ or } X_n = 3\}$. Determine $\Pr\{X_T = 0 | X_0 = 1\}$.

**3.6.3** Players A and B each have \$50 at the beginning of a game in which each player bets \$1 at each play, and the game continues until one player is broke. Suppose there is a constant probability $p = 0.492929\ldots$ that Player A wins on any given bet. What is the mean duration of the game?

**3.6.4** Consider the random walk Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

Starting in state 1, determine the mean time until absorption. Do this first using the basic first step approach of equation (3.24), and second using the particular formula for $v_i$ that follows equation (3.54), which applies for a random walk in which $p \neq q$.

## Problems

**3.6.1** The probability of gambler's ruin

$$u_i = \Pr\{X_T = 0 | X_0 = i\} \tag{3.61}$$

satisfies the first step analysis equation

$$u_i = q_i u_{i-1} + r_i u_i + p_i u_{i+1} \quad \text{for } i = 1, \ldots, N-1,$$

and

$$u_0 = 1, \quad u_N = 0.$$

The solution is

$$u_i = \frac{\rho_i + \cdots + \rho_{N-1}}{1 + \rho_1 + \rho_2 + \cdots + \rho_{N-1}}, \quad i = 1, \ldots, N-1, \tag{3.62}$$

where

$$\rho_k = \frac{q_1 q_2 \cdots q_k}{p_1 p_2 \cdots p_k}, \quad k = 1, \ldots, N-1. \tag{3.63}$$

**3.6.2** The mean hitting time

$$v_k = E[T | X_0 = k] \tag{3.64}$$

satisfies the equations

$$v_k = 1 + q_k v_{k-1} + r_k v_k + p_k v_{k+1} \quad \text{and} \quad v_0 = v_N = 0. \tag{3.65}$$

The solution is

$$\begin{aligned} v_k = &\left( \frac{\Phi_1 + \cdots + \Phi_{N-1}}{1 + \rho_1 + \cdots + \rho_{N-1}} \right) (1 + \rho_1 + \cdots + \rho_{k-1}) \\ &- (\Phi_1 + \cdots + \Phi_{k-1}) \quad \text{for } k = 1, \ldots, N-1, \end{aligned} \tag{3.66}$$

where $\rho_i$ is given in (3.63) and

$$\begin{aligned} \Phi_i &= \left( \frac{1}{q_1} + \frac{1}{q_2 \rho_1} + \cdots + \frac{1}{q_i \rho_{i-1}} \right) \rho_i \\ &= \frac{q_2 \cdots q_i}{p_1 \cdots p_i} + \frac{q_3 \cdots q_i}{p_2 \cdots p_i} + \cdots + \frac{q_i}{p_{i-1} p_i} + \frac{1}{p_i} \quad \text{for } i = 1, \ldots, N-1. \end{aligned} \tag{3.67}$$

**3.6.3** Fix a state $k$, where $0 < k < N$, and let $W_{ik}$ be the mean total visits to state $k$ starting from $i$. Formally, the definition is

$$W_{ik} = E\left[\sum_{n=0}^{T-1} \mathbf{1}\{X_n = k\} | X_0 = i\right], \tag{3.68}$$

where

$$\mathbf{1}\{X_n = k\} = \begin{cases} 1 & \text{if } X_n = k, \\ 0 & \text{if } X_n \neq k. \end{cases}$$

Then, $W_{ik}$ satisfies the equations

$$W_{ik} = \delta_{ik} + q_i W_{i-1,k} + r_i W_{ik} + p_i W_{i+1,k} \quad \text{for } i = 1, \ldots, N-1$$

and

$$W_{0k} = W_{Nk} = 0,$$

where

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

The solution is

$$W_{ik} = \begin{cases} \dfrac{(1 + \cdots + \rho_{i-1})(\rho_k + \cdots + \rho_{N-1})}{1 + \cdots + \rho_{N-1}} \left(\dfrac{1}{q_k \rho_{k-1}}\right) & \text{for } i \leq k, \\[2ex] \left[\dfrac{(1 + \cdots + \rho_{i-1})(\rho_k + \cdots + \rho_{N-1})}{1 + \cdots + \rho_{N-1}} \right. \\ \qquad \left. -(\rho_k + \cdots + \rho_{i-1})\right] \left(\dfrac{1}{q_k \rho_{k-1}}\right) & \text{for } i \geq k. \end{cases} \tag{3.69}$$

**3.6.4** The probability of absorption at 0 starting from state $k$

$$u_k = \Pr\{X_T = 0 | X_0 = k\} \tag{3.70}$$

satisfies the equation

$$u_k = p_k + r_k u_k + q_k u_{k+1},$$
$$\text{for } k = 1, \ldots, N-1 \quad \text{and} \quad u_0 = 1, u_N = 0.$$

The solution is

$$u_k = 1 - \left(\frac{q_k}{p_k + q_k}\right) \cdots \left(\frac{q_{N-1}}{p_{N-1} + q_{N-1}}\right) \quad \text{for } k = 1, \ldots, N-1. \tag{3.71}$$

**3.6.5** The mean hitting time

$$v_k = E[T|X_0 = k] \tag{3.72}$$

satisfies the equation

$$v_k = 1 + r_k v_k + q_k v_{k+1} \quad \text{for } k = 1, \ldots, N-1 \quad \text{and} \quad v_0 = v_N = 0.$$

The solution is

$$v_k = \frac{1}{p_k + q_k} + \frac{\pi_{k,k+1}}{p_{k+1} + q_{k+1}} + \cdots + \frac{\pi_{k,N-1}}{p_{N-1} + q_{N-1}}, \tag{3.73}$$

where

$$\pi_{kj} = \left(\frac{q_k}{p_k + q_k}\right)\left(\frac{q_{k+1}}{p_{k+1} + q_{k+1}}\right) \cdots \left(\frac{q_{j-1}}{p_{j-1} + q_{j-1}}\right) \quad \text{for } k < j. \tag{3.74}$$

**3.6.6** Fix a state $j (0 < j < N)$ and let $W_{ij}$ be the mean total visits to state $j$ starting from state $i$ [see equation (3.68)]. Then,

$$W_{iJ} = \begin{cases} \dfrac{1}{p_i + q_i} & \text{for } j = i, \\ \left(\dfrac{q_i}{p_i + q_i}\right) \cdots \left(\dfrac{q_{j-1}}{p_{j-1} + q_{j-1}}\right) \dfrac{1}{p_j + q_j} & \text{for } i < j, \\ 0 & \text{for } i > j. \end{cases} \tag{3.75}$$

**3.6.7** Consider the random walk Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c|cccc|} & 0 & 1 & 2 & 3 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0.3 & 0 & 0.7 & 0 \\ 2 & 0 & 0.1 & 0 & 0.9 \\ 3 & 0 & 0 & 0 & 1 \end{array}.$$

Starting in state 1, determine the mean time until absorption. Do this first using the basic first step approach of equation (3.24) and second using the particular results for a random walk given in equation (3.66).

**3.6.8** Consider the Markov chain $\{X_n\}$ whose transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \left\| \begin{matrix} \alpha & 0 & \beta & 0 \\ \alpha & 0 & 0 & \beta \\ \alpha & \beta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \right\| \end{array},$$

where $\alpha > 0, \beta > 0$, and $\alpha + \beta = 1$. Determine the mean time to reach state 3 starting from state 0. That is, find $E[T|X_0 = 0]$, where $T = \min\{n \geq 0; X_n = 3\}$.

**3.6.9** *Computer Challenge*. You have two urns: A and B, with $a$ balls in A and $b$ balls in B. You pick an urn at random, each urn being equally likely, and move a ball from it to the other urn. You do this repeatedly. The game ends when either of the urns becomes empty. The number of balls in A at the $n$th move is a simple random walk, and the expected duration of the game is $E[T] = ab$ [see equation (3.52)]. Now consider three urns, A, B, and C, with $a, b$, and $c$ balls, respectively. You pick an urn at random, each being equally likely, and move a ball from it to one of the other two urns, each being equally likely. The game ends when one of the three urns becomes empty. What is the mean duration of the game? If you can guess the general form of this mean time by computing it in a variety of particular cases, it is not particularly difficult to verify it by a first step analysis. What about four urns?

## 3.7 Another Look at First Step Analysis*

In this section, we provide an alternative approach to evaluating the functionals treated in Section 3.4. The $n$th power of a transition probability matrix having both transient and absorbing states is directly evaluated. From these $n$th powers, it is possible to extract the mean number of visits to a transient state $j$ prior to absorption, the mean time until absorption, and the probability of absorption in any particular absorbing state $k$. These functionals all depend on the initial state $X_0 = i$, and as a by-product of the derivation, we show that, as functions of this initial state $i$, these functionals satisfy their appropriate first step analysis equations.

Consider a Markov chain whose states are labeled $0, 1, \ldots, N$. States $0, 1, \ldots, r-1$ are transient in that $P_{ij}^{(n)} \to 0$ as $n \to \infty$ for $0 \leq i, j < r$, while states $r, \ldots, N$ are absorbing, or trap, and here $P_{ii} = 1$ for $r \leq i \leq N$. The transition matrix has the form

$$\mathbf{P} = \left\| \begin{matrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{matrix} \right\|, \tag{3.76}$$

* This section contains material at a more difficult level. It is not prerequisite to what follows.

where $\mathbf{0}$ is an $(N-r+1)\times r$ matrix all of whose components are zero, $\mathbf{I}$ is an $(N-r+1)\times(N-r+1)$ identity matrix, and $Q_{ij}=P_{ij}$ for $0\le i,j<r$.

To illustrate the calculations, begin with the four-state transition matrix

$$\mathbf{P}=\begin{array}{c} \\ 0\\1\\2\\3\end{array}\begin{array}{c}\begin{array}{cccc}0&1&2&3\end{array}\\ \left\|\begin{array}{cccc}Q_{00}&Q_{01}&R_{02}&R_{03}\\Q_{10}&Q_{11}&R_{12}&R_{13}\\0&0&1&0\\0&0&0&1\end{array}\right\|\end{array}. \tag{3.77}$$

Straightforward matrix multiplication shows the square of $\mathbf{P}$ to be

$$\mathbf{P}^2=\left\|\begin{array}{cc}\mathbf{Q}^2&\mathbf{R}+\mathbf{QR}\\\mathbf{0}&\mathbf{I}\end{array}\right\|. \tag{3.78}$$

Continuing on to the third power, we have

$$\mathbf{P}^3=\left\|\begin{array}{cc}\mathbf{Q}&\mathbf{R}\\\mathbf{0}&\mathbf{I}\end{array}\right\|\times\left\|\begin{array}{cc}\mathbf{Q}^2&\mathbf{R}+\mathbf{QR}\\\mathbf{0}&\mathbf{I}\end{array}\right\|=\left\|\begin{array}{cc}\mathbf{Q}^3&\mathbf{R}+\mathbf{QR}+\mathbf{Q}^2\mathbf{R}\\\mathbf{0}&\mathbf{I}\end{array}\right\|,$$

and for higher values of $n$,

$$\mathbf{P}^n=\left\|\begin{array}{cc}\mathbf{Q}^n&\left(\mathbf{I}+\mathbf{Q}+\cdots+\mathbf{Q}^{n-1}\right)\mathbf{R}\\\mathbf{0}&\mathbf{I}\end{array}\right\|. \tag{3.79}$$

The consideration of four states was for typographical convenience only. It is straightforward to verify that the $n$th power of $\mathbf{P}$ is given by (3.79) for the general $(N+1)$-state transition matrix of (3.76) in which states $0,1,\ldots,r-1$ are *transient* ($P_{ij}^{(n)}\to 0$ as $n\to\infty$ for $0\le i,j<r$) while states $r,\ldots,N$ are *absorbing* ($P_{ii}=1$ for $r\le i\le N$).

We turn to the interpretation of (3.79). Let $W_{ij}^{(n)}$ be the mean number of visits to state $j$ up to stage $n$ for a Markov chain starting in state $i$. Formally,

$$W_{ij}^{(n)}=E\left[\sum_{l=0}^{n}\mathbf{1}\{X_l=j\}|X_0=i\right], \tag{3.80}$$

where

$$\mathbf{1}\{X_l=j\}=\begin{cases}1 & \text{if } X_l=j,\\ 0 & \text{if } X_l\neq j.\end{cases} \tag{3.81}$$

Now, $E[\mathbf{1}\{X_l = j\}|X_0 = i] = \Pr\{X_l = j|X_0 = i\} = P_{ij}^{(l)}$, and since the expected value of a sum is the sum of the expected values, we obtain from (3.80) that

$$
\begin{aligned}
W_{ij}^{(n)} &= \sum_{i=0}^{n} E[\mathbf{1}\{X_l = j\}|X_0 = i] \\
&= \sum_{l=0}^{n} P_{ij}^{(l)}.
\end{aligned}
\tag{3.82}
$$

Equation (3.82) holds for all states $i, j$, but it has the most meaning when $i$ and $j$ are transient. Because (3.79) asserts that $P_{ij}^{(l)} = Q_{ij}^{(l)}$ when $0 \le i, j < r$, then

$$
W_{ij}^{(n)} = Q_{ij}^{(0)} + Q_{ij}^{(1)} + \cdots + Q_{ij}^{(n)}, \quad 0 \le i, j < r,
$$

where

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

In matrix notation, $\mathbf{Q}^{(\mathbf{0})} = \mathbf{I}$, and because $\mathbf{Q}^{(n)} = \mathbf{Q}^n$, the $n$th power of $\mathbf{Q}$, then

$$
\begin{aligned}
\mathbf{W}^{(n)} &= \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots + \mathbf{Q}^n \\
&= \mathbf{I} + \mathbf{Q}\left(\mathbf{I} + \mathbf{Q} + \cdots + \mathbf{Q}^{n-1}\right) \\
&= \mathbf{I} + \mathbf{Q}\mathbf{W}^{(n-1)}.
\end{aligned}
\tag{3.83}
$$

Upon writing out the matrix equation (3.83) in terms of the matrix entries, we recognize the results of a first step analysis. We have

$$
\begin{aligned}
W_{ij}^{(n)} &= \delta_{ij} + \sum_{k=0}^{r-1} Q_{ik} W_{kj}^{(n-1)} \\
&= \delta_{ij} + \sum_{k=0}^{r-1} P_{ik} W_{kj}^{(n-1)}.
\end{aligned}
$$

In words, the equation asserts that the mean number of visits to state $j$ in the first $n$ stages starting from the initial stage $i$ includes the initial visit if $i = j(\delta_{ij})$ plus the future visits during the $n-1$ remaining stages weighted by the appropriate transition probabilities.

We pass to the limit in (3.83) and obtain for

$$
W_{ij} = \lim_{n \to \infty} W_{ij}^{(n)} = E[\text{Total visits to } j|X_0 = i], \quad 0 \le i, j < r,
$$

the matrix equations

$$\mathbf{W} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots$$

and

$$\mathbf{W} = \mathbf{I} + \mathbf{QW}. \tag{3.84}$$

In terms of its entries, (3.84) is

$$W_{ij} = \delta_{ij} + \sum_{l=0}^{r-1} P_{il} W_{lj} \quad \text{for } i, j = 0, \ldots, r-1. \tag{3.85}$$

Equation (3.85) is the same as equation (3.29), which was derived by a first step analysis.

Rewriting equation (3.84) in the form

$$\mathbf{W} - \mathbf{QW} = (\mathbf{I} - \mathbf{Q})\mathbf{W} = \mathbf{I}, \tag{3.86}$$

we see that $\mathbf{W} = (\mathbf{I} - \mathbf{Q})^{-1}$, the inverse matrix to $\mathbf{I} - \mathbf{Q}$. The matrix $\mathbf{W}$ is often called the *fundamental* matrix associated with $\mathbf{Q}$.

Let $T$ be the time of absorption. Formally, since states $r, r+1, \ldots, N$ are the absorbing ones, the definition is

$$T = \min\{n \geq 0; r \leq X_n \leq N\}.$$

Then, the $(i, j)$th element $W_{ij}$ of the fundamental matrix $\mathbf{W}$ evaluates

$$W_{ij} = E\left[\sum_{n=0}^{T-1} \mathbf{1}\{X = j\} | X_0 = i\right] \quad \text{for } 0 \leq i, j < r. \tag{3.87}$$

Let $v_i = E[T|X_0 = i]$ be the mean time to absorption starting from state $i$. The time to absorption is composed of sojourns in the transient states. Formally,

$$\begin{aligned}\sum_{j=0}^{r-1} \sum_{n=0}^{T-1} \mathbf{1}\{X_n = j\} &= \sum_{n=0}^{T-1} \sum_{j=0}^{r-1} \mathbf{1}\{X_n = j\} \\ &= \sum_{n=0}^{T-1} 1 = T.\end{aligned}$$

It follows from (3.87), then, that

$$\begin{aligned}\sum_{j=0}^{r-1} W_{ij} &= \sum_{j=0}^{r-1} E\left[\sum_{n=0}^{T-1} \mathbf{1}\{X_n = j\} | X_0 = i\right] \\ &= E[T|X_0 = i] = v_i \quad \text{for } 0 \leq i < r.\end{aligned} \tag{3.88}$$

Summing equation (3.85) over transient states $j$ as follows,

$$\sum_{j=0}^{r-1} W_{ij} = \sum_{j=0}^{r-1} \delta_{ij} + \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} P_{ij} W_{kj} \quad \text{for } i = 0, 1, \ldots, r-1,$$

and using the equivalence $v_i = \Sigma_{j=0}^{r-1} W_{ij}$ leads to

$$v_i = 1 + \sum_{k=0}^{r-1} P_{ij} v_k \quad \text{for } i = 0, 1, \ldots, r-1. \tag{3.89}$$

This equation is identical with that derived by first step analysis in (3.28). We turn to the hitting probabilities. Recall that states $k = r, \ldots, N$ are absorbing. Since such a state cannot be left once entered, the probability of absorption in a particular absorbing state $k$ up to time $n$, starting from initial state $i$, is simply

$$\begin{aligned} P_{ik}^{(n)} &= \Pr\{X_n = k | X_0 = i\} \\ &= \Pr\{T \leq n \text{ and } X_T = k | X_0 = i\} \\ &\qquad \text{for } i = 0, \ldots, r-1; k = r, \ldots, N, \end{aligned} \tag{3.90}$$

where $T = \min\{n \geq 0 \colon r \leq X_n \leq N\}$ is the time of absorption. Let

$$\begin{aligned} U_{ik}^{(n)} &= \Pr\{T \leq n \text{ and } X_T = k | X_0 = i\} \\ &\qquad \text{for } 0 \leq i < r \text{ and } r \leq k \leq N. \end{aligned} \tag{3.91}$$

Referring to (3.79) and (3.90), we give the matrix $\mathbf{U}^{(n)}$ by

$$\begin{aligned} \mathbf{U}^{(n)} &= \left(\mathbf{I} + \mathbf{Q} + \cdots + \mathbf{Q}^{n-1}\right)\mathbf{R} \\ &= \mathbf{W}^{(n-1)}\mathbf{R} \qquad \text{[by (3.83)]}. \end{aligned} \tag{3.92}$$

If we pass to the limit in $n$, we obtain the hitting probabilities

$$U_{ik} = \lim_{n\to\infty} U_{ik}^{(n)} = \Pr\{X_T = k | X_0 = i\} \quad \text{for } 0 \leq i < r \text{ and } r \leq k \leq N.$$

Equation (3.92) then leads to an expression of the hitting probability matrix $\mathbf{U}$ in terms of the fundamental matrix $\mathbf{W}$ as simply $\mathbf{U} = \mathbf{WR}$, or

$$U_{ik} = \sum_{j=0}^{r-1} W_{ij} R_{jk} \quad \text{for } 0 \leq i < r \text{ and } r \leq k \leq N. \tag{3.93}$$

Equation (3.93) may be used in conjunction with (3.85) to verify the first step analysis equation for $U_{ik}$. We multiply (3.85) by $R_{jk}$ and sum, obtaining thereby

$$\sum_{j=0}^{r-1} W_{ij}R_{jk} = \sum_{j=0}^{r-1} \delta_{ij}R_{jk} + \sum_{j=0}^{r-1}\sum_{l=0}^{r-1} P_{il}W_{lj}R_{jk},$$

which with (3.93) gives

$$\begin{aligned} U_{ik} &= R_{ik} + \sum_{l=0}^{r-1} P_{il}U_{lk} \\ &= P_{ik} + \sum_{l=0}^{r-1} P_{il}U_{lk} \quad \text{for } 0 \le i < r \text{ and } r \le k \le N. \end{aligned}$$

This equation was derived earlier by first step analysis in (3.26).

## Exercises

**3.7.1** Consider the Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

The transition probability matrix corresponding to the nonabsorbing states is

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left\| \begin{array}{cc} 0.2 & 0.5 \\ 0.2 & 0.6 \end{array} \right\| \end{array}.$$

Calculate the matrix inverse to $\mathbf{I} - \mathbf{Q}$, and from this determine

**(a)** the probability of absorption into state 0 starting from state 1;

**(b)** the mean time spent in each of states 1 and 2 prior to absorption.

**3.7.2** Consider the random walk Markov chain whose transition probability matrix is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

The transition probability matrix corresponding to the nonabsorbing states is

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left\| \begin{array}{cc} 0 & 0.7 \\ 0.3 & 0 \end{array} \right\| \end{array}.$$

Calculate the matrix inverse to $\mathbf{I}-\mathbf{Q}$, and from this determine
**(a)** the probability of absorption into state 0 starting from state 1;
**(b)** the mean time spent in each of states 1 and 2 prior to absorption.

## Problems

**3.7.1** A zero-seeking device operates as follows: If it is in state $m$ at time $n$, then at time $n+1$ its position is uniformly distributed over the states $0, 1, \ldots, m-1$. State 0 is absorbing. Find the inverse of the $\mathbf{I}-\mathbf{Q}$ matrix for the transient states $1, 2, \ldots, m$.

**3.7.2** A zero-seeking device operates as follows: If it is in state $j$ at time $n$, then at time $n+1$ its position is 0 with probability $1/j$, and its position is $k$ (where $k$ is one of the states $1, 2, \ldots, j-1$) with probability $2k/j^2$. State 0 is absorbing. Find the inverse of the $\mathbf{I}-\mathbf{Q}$ matrix.

**3.7.3** Let $X_n$ be an absorbing Markov chain whose transition probability matrix takes the form given in equation (3.76). Let $\mathbf{W}$ be the fundamental matrix, the matrix inverse of $\mathbf{I}-\mathbf{Q}$. Let

$$T = \min\{n \geq 0; r \leq n \leq N\}$$

be the random time of absorption (recall that states $r, r+1, \ldots, N$ are the absorbing states). Establish the joint distribution

$$\Pr\{X_{T-1} = j, X_T = k | X_0 = i\} = W_{ij} P_{jk} \quad \text{for } 0 \leq i, j < r; r \leq k \leq N,$$

whence

$$\Pr\{X_{T-1} = j | X_0 = i\} = \sum_{k=r}^{N} W_{ij} P_{jk} \quad \text{for } 0 \leq i, j < r.$$

**3.7.4** The possible states for a Markov chain are the integers $0, 1, \ldots, N$, and if the chain is in state $j$, at the next step it is equally likely to be in any of the states $0, 1, \ldots, j-1$. Formally,

$$P_{ij} = \begin{cases} 1, & \text{if } i = j = 0, \\ 0 & \text{if } 0 < i \leq j \leq N, \\ 1/i, & \text{if } 0 \leq j < i \leq N. \end{cases}$$

**(a)** Determine the fundamental matrix for the transient states $1, 2, \ldots, N$.
**(b)** Determine the probability distribution for the last positive integer that the chain visits.

**3.7.5** *Computer Challenge*. Consider the partial sums:

$$S_0 = k \quad \text{and} \quad S_m = k + \xi_1 + \cdots + \xi_m, \; k > 0,$$

where $\xi_1, \xi_2, \ldots$ are independent and identically distributed as

$$\Pr\{\xi = 0\} = 1 - \frac{2}{\pi}$$

and

$$\Pr\{\xi = \pm j\} = \frac{2}{\pi(4j^2 - 1)}, j = 1, 2, \ldots.$$

Can you find an explicit formula for the mean time $v_k$ for the partial sums starting from $S_0 = k$ to exit the interval $[0, N] = \{0, 1, \ldots, N\}$? In another context, the answer was found by computing it in a variety of special cases.

**Note:** A simple random walk *on the* integer plane moves according to the rule: If $(X_n, Y_n) = (i, j)$, then the next position is equally likely to be any of the four points $(i+1, j)$, $(i-1, j)$, $(i, j+1)$, or $(i, j-1)$. Let us suppose that the process starts at the point $(X_0, Y_0) = (k, k)$ on the diagonal, and we observe the process only when it visits the diagonal. Formally, we define

$$\tau_1 = \min\{n > 0; X_n = Y_n\},$$

and

$$\tau_m = \min\{n > \tau_{m-1}; X_n = Y_n\}.$$

It is not hard to show that

$$S_0 = k, \; S_m = X_{\tau_m} = Y_{\tau_m}, \; m > 0,$$

is a version of the above partial sum process.

## 3.8 Branching Processes*

Suppose an organism at the end of its lifetime produces a random number $\xi$ of offspring with probability distribution

$$\Pr\{\xi = k\} = p_k \quad \text{for } k = 0, 1, 2, \ldots, \tag{3.94}$$

* Branching processes are Markov chains of a special type. Sections 3.8 and 3.9 are not prerequisites to the later chapters.

where as usual, $p_k \geq 0$ and $\sum_{k=0}^{\infty} p_k = 1$. We assume that all offspring act independently of each other and at the ends of their lifetimes (for simplicity, the lifespans of all organisms are assumed to be the same) individually have progeny in accordance with the probability distribution (3.94), thus propagating their species. The process $\{X_n\}$, where $X_n$ is the population size at the $n$th generation, is a Markov chain of special structure called a *branching process*.

The Markov property may be reasoned simply as follows. In the $n$th generation, the $X_n$ individuals independently give rise to numbers of offspring $\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_{X_n}^{(n)}$, and hence the cumulative number produced for the $(n+1)$st generation is

$$X_{n+1} = \xi_1^{(n)} + \xi_2^{(n)} + \cdots + \xi_{X_n}^{(n)}. \tag{3.95}$$

### 3.8.1 Examples of Branching Processes

There are numerous examples of Markov branching processes that arise naturally in various scientific disciplines. We list some of the more prominent cases.

#### Electron Multipliers

An electron multiplier is a device that amplifies a weak current of electrons. A series of plates are set up in the path of electrons emitted by a source. Each electron, as it strikes the first plate, generates a random number of new electrons, which in turn strike the next plate and produce more electrons, and so forth. Let $X_0$ be the number of electrons initially emitted and $X_1$ be the number of electrons produced on the first plate by the impact due to the $X_0$ initial electrons; in general, let $X_n$ be the number of electrons emitted from the $n$th plate due to electrons emanating from the $(n-1)$st plate. The sequence of random variables $X_0, X_1, X_2, \ldots, X_n, \ldots$ constitutes a branching process.

#### Neutron Chain Reaction

A nucleus is split by a chance collision with a neutron. The resulting fission yields a random number of new neutrons. Each of these secondary neutrons may hit some other nucleus, producing a random number of additional neutrons, and so forth. In this case, the initial number of neutrons is $X_0 = 1$. The first generation of neutrons comprises all those produced from the fission caused by the initial neutron. The size of the first generation is a random variable $X_1$. In general, the population $X_n$ at the $n$th generation is produced by the chance hits of the $X_{n-1}$ individual neutrons of the $(n-1)$st generation.

#### Survival of Family Names

The family name is inherited by sons only. Suppose that each individual has probability $p_k$ of having $k$ male offspring. Then, from one individual there result the $1\text{st}, 2\text{nd}, \ldots, n\text{th}, \ldots$ generations of descendants. We may investigate the distribution of such random variables as the number of descendants in the $n$th generation, or the probability that the family name will eventually become extinct. Such questions will be dealt with beginning in Section 3.8.3.

### *Survival of Mutant Genes*

Each individual gene has a chance to give birth to $k$ offspring, $k = 1, 2, \ldots,$ which are genes of the same kind. Any individual, however, has a chance to transform into a different type of mutant gene. This gene may become the first in a sequence of generations of a particular mutant gene. We may inquire about the chances of survival of the mutant gene within the population of the original genes. In this example, the number of offspring is often assumed to follow a Poisson distribution.

The rationale behind this choice of distribution is as follows. In many populations a large number of zygotes (fertilized eggs) are produced, only a small number of which grow to maturity. The events of fertilization and maturation of different zygotes obey the law of independent binomial trials. The number of trials (i.e., number zygotes) is large. The law of rare events then implies that the number of progeny that mature will approximately follow the Poisson distribution. The Poisson assumption seems quite appropriate in the model of population growth of a rare mutant gene. If the mutant gene carries a biological advantage (or disadvantage), then the probability distribution is taken to be the Poisson distribution with mean $\lambda > 1$ or ($< 1$).

All of the preceding examples possess the following structure. Let $X_0$ denote the size of the initial population. Each individual gives birth to $k$ new individuals with probability $p_k$ *independently of the others*. The totality of all the direct descendants of the initial population constitutes the first generation, whose size we denote by $X_1$. Each individual of the first generation independently bears a progeny set whose size is governed by the probability distribution (3.94). The descendants produced constitute the second generation, of size $X_2$. In general, the $n$th generation is composed of descendants of the $(n-1)$st generation, each of whose members independently produces $k$ progeny with probability $p_k, k = 0, 1, 2, \ldots$. The population size of the $n$th generation is denoted by $X_n$. The $X_n$ forms a sequence of integer-valued random variables that generate a Markov chain in the manner described by (3.95).

## 3.8.2 The Mean and Variance of a Branching Process

Equation (3.95) characterizes the evolution of the branching process as successive random sums of random variables. Random sums were studied in Chapter 2, Section 2.3, and we can use the moment formulas developed there to compute the mean and variance of the population size $X_n$. First some notation. Let $\mu = E[\xi]$ and $\sigma^2 = \text{Var}[\xi]$ be the mean and variance, respectively, of the offspring distribution (3.94). Let $M(n)$ and $V(n)$ be the mean and variance of $X_n$ under the initial condition $X_0 = 1$. Then, direct application of Chapter 2, (2.30) with respect to the random sum (3.95) gives the recursions

$$M(n+1) = \mu M(n) \tag{3.96}$$

and

$$V(n+1) = \sigma^2 M(n) + \mu^2 V(n). \tag{3.97}$$

The initial condition $X_0 = 1$ starts the recursions (3.96) and (3.97) at $M(0) = 1$ and $V(0) = 0$. Then, from (3.96), we obtain $M(1) = \mu 1 = \mu, M(2) = \mu M(1) = \mu^2$, and, in general,

$$M(n) = \mu^n \quad \text{for } n = 0, 1, \ldots. \tag{3.98}$$

Thus, the mean population size increases geometrically when $\mu > 1$, decreases geometrically when $\mu < 1$, and remains constant when $\mu = 1$.

Next, substitution of $M(n) = \mu^n$ into (3.97) gives $V(n+1) = \sigma^2\mu^n + \mu^2 V(n)$, which with $V(0) = 0$ yields

$$\begin{aligned}
V(1) &= \sigma^2, \\
V(2) &= \sigma^2\mu + \mu^2 V(1) = \sigma^2\mu + \sigma^2\mu^2, \\
V(3) &= \sigma^2\mu^2 + \mu^2 V(2) \\
&= \sigma^2\mu^2 + \sigma^2\mu^3 + \sigma^2\mu^4,
\end{aligned}$$

and, in general,

$$\begin{aligned}
V(n) &= \sigma^2\left[\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2}\right] \\
&= \sigma^2\mu^{n-1}\left[1 + \mu + \cdots + \mu^{n-1}\right] \\
&= \sigma^2\mu^{n-1} \times \begin{cases} n & \text{if } \mu = 1, \\ \dfrac{1-\mu^n}{1-\mu} & \text{if } \mu \neq 1. \end{cases}
\end{aligned} \tag{3.99}$$

Thus, the variance of the population size increases geometrically if $\mu > 1$, increases linearly if $\mu = 1$, and decreases geometrically if $\mu < 1$.

### 3.8.3 Extinction Probabilities

Population extinction occurs when and if the population size is reduced to zero. The random time of extinction $N$ is thus the first time $n$ for which $X_n = 0$, and then, obviously, $X_k = 0$ for all $k \geq N$. In Markov chain terminology, 0 is an absorbing state, and we may calculate the probability of extinction by invoking a first step analysis. Let

$$u_n = \Pr\{N \leq n\} = \Pr\{X_n = 0\} \tag{3.100}$$

be the probability of extinction at or prior to the $n$th generation, beginning with a single parent $X_0 = 1$. Suppose that the single parent represented by $X_0 = 1$ gives rise to $\xi_1^{(0)} = k$ offspring. In turn, each of these offspring will generate a population of its own descendants, and if the original population is to die out in $n$ generations, then each of these $k$ lines of descent must die out in $n - 1$ generations. The analysis is depicted in Figure 3.6.

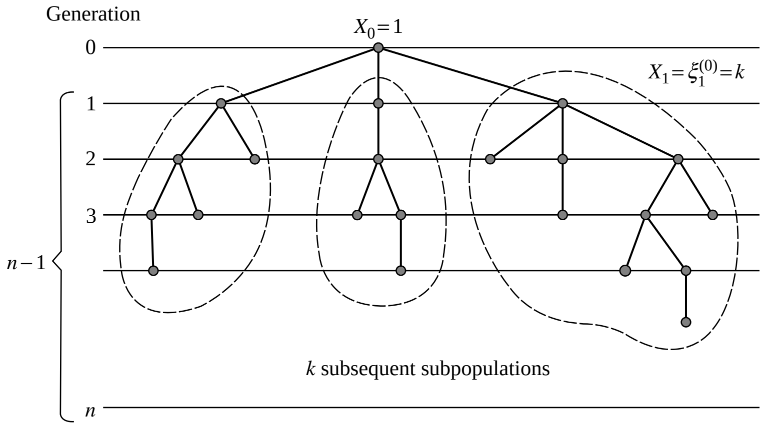



**Figure 3.6** The diagram illustrates that if the original population is to die out by generation $n$, then the subpopulations generated by distinct initial offspring must all die out in $n-1$ generations.

Now, the $k$ subpopulations generated by the distinct offspring of the original parent are independent, and they have the same statistical properties as the original population. Therefore, the probability that any particular one of them dies out in $n-1$ generations is $u_{n-1}$ by definition, and the probability that all $k$ subpopulations die out in $n-1$ generations is the $k$th power $(u_{n-1})^k$ because they are independent. Upon weighting this factor by the probability of $k$ offspring and summing according to the law of total probability, we obtain

$$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k, \quad n = 1, 2, \ldots. \tag{3.101}$$

Of course $u_0 = 0$, and $u_1 = p_0$, the probability that the original parent had no offspring.

**Example** Suppose a parent has no offspring with probability $\frac{1}{4}$ and two offspring with probability $\frac{3}{4}$. Then, the recursion (3.101) specializes to

$$u_n = \frac{1}{4} + \frac{3}{4}(u_{n-1})^2 = \frac{1+3(u_{n-1})^2}{4}.$$

Beginning with $u_0 = 0$, we successively compute

$$\begin{aligned} u_1 &= 0.2500, & u_6 &= 0.3313, \\ u_2 &= 0.2969, & u_7 &= 0.3323, \\ u_3 &= 0.3161, & u_8 &= 0.3328, \\ u_4 &= 0.3249, & u_9 &= 0.3331, \\ u_5 &= 0.3292, & u_{10} &= 0.3332. \end{aligned}$$

We see that the chances are very nearly $\frac{1}{3}$ that such a population will die out by the tenth generation.

## Exercises

**3.8.1** A population begins with a single individual. In each generation, each individual in the population dies with probability $\frac{1}{2}$ or doubles with probability $\frac{1}{2}$. Let $X_n$ denote the number of individuals in the population in the $n$th generation. Find the mean and variance of $X_n$.

**3.8.2** The number of offspring of an individual in a population is $0, 1$, or $2$ with respective probabilities $a > 0$, $b > 0$, and $c > 0$, where $a + b + c = 1$. Express the mean and variance of the offspring distribution in terms of $b$ and $c$.

**3.8.3** Suppose a parent has no offspring with probability $\frac{1}{2}$ and has two offspring with probability $\frac{1}{2}$. If a population of such individuals begins with a single parent and evolves as a branching process, determine $u_n$, the probability that the population is extinct by the $n$th generation, for $n = 1, 2, 3, 4, 5$.

**3.8.4** At each stage of an electron multiplier, each electron, upon striking the plate, generates a Poisson distributed number of electrons for the next stage. Suppose the mean of the Poisson distribution is $\lambda$. Determine the mean and variance for the number of electrons in the $n$th stage.

## Problems

**3.8.1** Each adult individual in a population produces a fixed number $M$ of offspring and then dies. A fixed number $L$ of these remain at the location of the parent. These local offspring will either all grow to adulthood, which occurs with a fixed probability $\beta$, or all will die, which has probability $1 - \beta$. Local mortality is catastrophic in that it affects the entire local population. The remaining $N = M - L$ offspring disperse. Their successful growth to adulthood will occur statistically independently of one another, but at a lower probability $\alpha = p\beta$, where $p$ may be thought of as the probability of successfully surviving the dispersal process. Define the random variable $\xi$ to be the number of offspring of a single parent that survive to reach adulthood in the next generation. According to our assumptions, we may write $\xi$ as

$$\xi = \nu_1 + \nu_2 + \cdots + \nu_N + (M - N)\Theta,$$

where $\Theta, \nu_1, \nu_2, \ldots, \nu_N$ are independent with $\Pr\{\nu_k = 1\} = \alpha, \Pr\{\nu_k = 0\} = 1 - \alpha$, and with $\Pr\{\Theta = 1\} = \beta$ and $\Pr\{\Theta = 0\} = 1 - \beta$. Show that the mean number of offspring reaching adulthood is $E[\xi] = \alpha N + \beta(M - N)$, and since $\alpha < \beta$, the mean number of surviving offspring is maximized by dispersing none ($N = 0$). Show that the probability of having no offspring surviving to adulthood is

$$\Pr\{\xi = 0\} = (1 - \alpha)^N (1 - \beta)$$

and that this probability is made smallest by making $N$ large.

**3.8.2** Let $Z = \sum_{n=0}^{x} X_n$ be the total family size in a branching process whose offspring distribution has a mean $\mu = E[\xi] < 1$. Assuming that $X_0 = 1$, show that $E[Z] = 1/(1-\mu)$.

**3.8.3** Families in a certain society choose the number of children that they will have according to the following rule: If the first child is a girl, they have exactly one more child. If the first child is a boy, they continue to have children until the first girl, and then cease childbearing.

**(a)** For $k = 0, 1, 2, \ldots$, what is the probability that a particular family will have $k$ children in total?

**(b)** For $k = 0, 1, 2, \ldots$, what is the probability that a particular family will have exactly $k$ male children among their offspring?

**3.8.4** Let $\{X_n\}$ be a branching process with mean family size $\mu$. Show that $Z_n = X_n/\mu^n$ is a nonnegative martingale. Interpret the maximal inequality as applied to $\{Z_n\}$.

## 3.9 Branching Processes and Generating Functions*

Consider a nonnegative integer-valued random variable $\xi$ whose probability distribution is given by

$$\Pr\{\xi = k\} = p_k \quad \text{for } k = 0, 1, \ldots. \tag{3.102}$$

The *generating function* $\phi(s)$ associated with the random variable $\xi$ (or equivalently, with the distribution $\{p_k\}$) is defined by

$$\phi(s) = E\left[s^{\xi}\right] = \sum_{k=0}^{\infty} p_k s^k \quad \text{for } 0 \le s \le 1. \tag{3.103}$$

Much of the importance of generating functions derives from the following three results.

First, the relation between probability mass functions (3.102) and generating functions (3.103) is one-to-one. Thus, knowing the generating function is equivalent, in some sense, to knowing the distribution. The relation that expresses the probability mass function $\{p_k\}$ in terms of the generating function $\phi(s)$ is

$$p_k = \frac{1}{k!} \frac{\mathrm{d}^k \phi(s)}{\mathrm{d}s^k}\bigg|_{s=0}. \tag{3.104}$$

For example,

$$\phi(s) = p_0 + p_1 s + p_2 s^2 + \cdots,$$

* This topic is not prerequisite to what follows.

whence

$$p_0 = \phi(0),$$

and

$$\frac{\mathrm{d}\phi(s)}{\mathrm{d}s} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots,$$

whence

$$p_1 = \left.\frac{\mathrm{d}\phi(s)}{\mathrm{d}s}\right|_{s=0}.$$

Second, if $\xi_1, \ldots, \xi_n$ are independent random variables having generating functions $\phi_1(s), \ldots, \phi_n(s)$, respectively, then the generating function of their sum $X = \xi_1 + \cdots + \xi_n$ is simply the product

$$\phi_X(s) = \phi_1(s)\phi_2(s)\cdots\phi_n(s). \tag{3.105}$$

This simple result makes generating functions extremely helpful in dealing with problems involving sums of independent random variables. It is to be expected, then, that generating functions might provide a major tool in the analysis of branching processes.

Third, the moments of a nonnegative integer-valued random variable may be found by differentiating the generating function. For example, the first derivative is

$$\frac{\mathrm{d}\phi(s)}{\mathrm{d}s} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots,$$

whence

$$\left.\frac{\mathrm{d}\phi(s)}{\mathrm{d}s}\right|_{s=1} = p_1 + 2p_2 + 3p_3 + \cdots = E[\xi], \tag{3.106}$$

and the second derivative is

$$\frac{\mathrm{d}^2\phi(s)}{\mathrm{d}s^2} = 2p_2 + 3(2)p_3 s + 4(3)p_4 s^2 + \cdots,$$

whence

$$\begin{aligned}\left.\frac{\mathrm{d}^2\phi(s)}{\mathrm{d}s^2}\right|_{s=1} &= 2p_2 + 3(2)p_3 + 4(3)p_4 + \cdots \\ &= \sum_{k=2}^{\infty} k(k-1)p_k = E[\xi(\xi-1)] \\ &= E\left[\xi^2\right] - E[\xi].\end{aligned}$$

Thus

$$E\left[\xi^2\right] = \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + E[\xi]$$
$$= \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1}$$

and

$$\begin{aligned}\mathrm{Var}[\xi] &= E\left[\xi^2\right] - \{E[\xi]\}^2 \\ &= \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1} - \left\{\left.\frac{d\phi(s)}{ds}\right|_{s=1}\right\}^2.\end{aligned} \tag{3.107}$$

**Example** If $\xi$ has a Poisson distribution with mean $\lambda$ for which

$$p_k = \Pr\{\xi = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, \ldots,$$

then,

$$\begin{aligned}\phi(s) = E\left[s^\xi\right] &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \\ &= e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)} \quad \text{for } |s| < 1.\end{aligned}$$

Then,

$$\frac{d\phi(s)}{ds} = \lambda e^{-\lambda(1-s)}; \qquad \left.\frac{d\phi(s)}{ds}\right|_{s=1} = \lambda;$$
$$\frac{d^2\phi(s)}{ds^2} = \lambda^2 e^{-\lambda(1-s)}; \qquad \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} = \lambda^2.$$

From (3.106) and (3.107), we verify that

$$E[\xi] = \lambda,$$
$$\mathrm{Var}[\xi] = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

### 3.9.1 Generating Functions and Extinction Probabilities

Consider a branching process whose population size at stage $n$ is denoted by $X_n$. Assume that the offspring distribution $p_k = \Pr\{\xi = k\}$ has the generating function

$\phi(s) = E\left[s^{\xi}\right] = \sum_k s^k p_k$. If $u_n = \Pr\{X_n = 0\}$ is the probability of extinction by stage $n$, then the recursion (3.101) in terms of generating functions becomes

$$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k = \phi(u_{n-1}).$$

That is, knowing the generating function $\phi(s)$, we may successively compute the extinction probabilities $u_n$ beginning with $u_0 = 0$ and then $u_1 = \phi(u_0), u_2 = \phi(u_1)$, and so on.

**Example** The extinction probabilities when there are no offspring with probability $p_0 = \frac{1}{4}$ and two offspring with probability $p_2 = \frac{3}{4}$ were computed in the example in Section 3.8.3. We now reexamine this example using the offspring generating function $\phi(s) = \frac{1}{4} + \frac{3}{4}s^2$. This generating function is plotted as Figure 3.7. From the figure, it is clear that the extinction probabilities converge upward to the smallest solution of the equation $u = \phi(u)$. This, in fact, occurs in the most general case. If $u_\infty$ denotes this smallest solution to $u = \phi(u)$, then $u_\infty$ gives the probability that the population eventually becomes extinct at some indefinite, but finite, time. The alternative is that the population grows infinitely large, and this occurs with probability $1 - u_\infty$.

For the example at hand, $\phi(s) = \frac{1}{4} + \frac{3}{4}s^2$, and the equation $u = \phi(u)$ is the simple quadratic $u = \frac{1}{4} + \frac{3}{4}u^2$, which gives

$$u = \frac{4 \pm \sqrt{16 - 12}}{6} = 1, \frac{1}{3}.$$

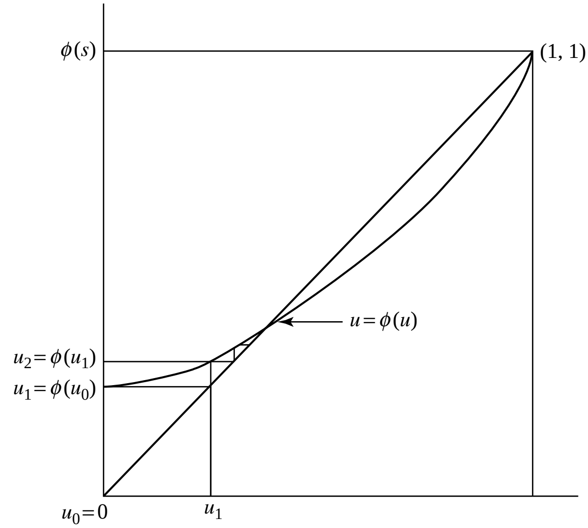



**Figure 3.7** The generating function corresponding to the offspring distribution $p_0 = \frac{1}{4}$ and $p_2 = \frac{3}{4}$. Here $u_k = \Pr\{X_k = 0\}$ is the probability of extinction by generation $k$.

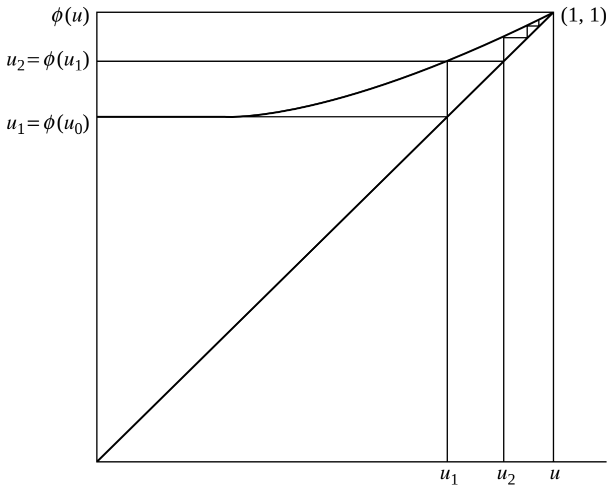



**Figure 3.8** The generating function corresponding to the offspring distribution $p_0 = \frac{3}{4}$ and $p_2 = \frac{1}{4}$.

The smaller solution is $u_\infty = \frac{1}{3}$, which is to be compared with the apparent limit of the sequence $u_n$ computed in the example in Section 3.8.3.

It may happen that $u_\infty = 1$, i.e., the population is sure to die out at some time. An example is depicted in Figure 3.8: The offspring distribution is $p_0 = \frac{3}{4}$ and $p_2 = \frac{1}{4}$. We solve $u = \phi(u) = \frac{3}{4} + \frac{1}{4}u^2$ to obtain

$$u = \frac{4 \pm \sqrt{16 - 12}}{2} = 1, 3.$$

The smaller solution is $u_\infty = 1$, the probability of eventual extinction.

In general, the key is whether or not the generating function $\phi(s)$ crosses the 45° line $\phi(s) = s$, and this, in turn, can be determined from the slope

$$\phi'(1) = \left. \frac{d\phi(s)}{ds} \right|_{s=1}$$

of the generating function at $s = 1$. If this slope is less than or equal to one, then no crossing takes place, and the probability of eventual extinction is $u_\infty = 1$. On the other hand, if the slope $\phi'(1)$ exceeds one, then the equation $u = \phi(u)$ has a smaller solution that is less than one, and extinction is not a certain event.

But the slope $\phi'(1)$ of a generating function at $s = 1$ is the mean $E[\xi]$ of the corresponding distribution. We have thus arrived at the following important conclusion: If the mean offspring size $E[\xi] \le 1$, then $u_\infty = 1$ and extinction is certain. If $E[\xi] > 1$, then $u_\infty < 1$ and the population may grow unboundedly with positive probability.

The borderline case $E[\xi] = 1$ merits some special attention. Here, $E[X_n|X_0 = 1] = 1$ for all $n$, so the mean population size is constant. Yet the population is sure to die out eventually! This is a simple example in which the mean population size alone does not adequately describe the population behavior.

### 3.9.2 Probability Generating Functions and Sums of Independent Random Variables

Let $\xi$ and $\eta$ be independent nonnegative integer-valued random variables having the probability generating functions (p.g.f.s)

$$\phi(s) = E\left[s^{\xi}\right] \quad \text{and} \quad \psi(s) = E\left[s^{\eta}\right] \quad \text{for } |s| < 1.$$

The probability generating function of the sum $\xi + \eta$ is simply the product $\phi(s)\psi(s)$ because

$$\begin{aligned} E\left[s^{\xi+\eta}\right] &= E\left[s^{\xi} s^{\eta}\right] \\ &= E\left[s^{\xi}\right] E\left[s^{\eta}\right] \qquad \text{(because } \xi \text{ and } \eta \text{ are independent)} \\ &= \phi(s)\psi(s). \end{aligned} \tag{3.108}$$

The converse is also true. Specifically, if the product of the p.g.f.s of two independent random variables is a p.g.f. of a third random variable, then the third random variable equals (in distribution) the sum of the other two.

Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed nonnegative integer-valued random variables with p.g.f. $\phi(s) = E\left[s^{\xi}\right]$. Direct induction of (3.108) implies that the sum $\xi_1 + \cdots + \xi_m$ has p.g.f.

$$E\left[s^{\xi_1 + \cdots + \xi_m}\right] = [\phi(s)]^m. \tag{3.109}$$

We extend this result to determine the p.g.f. of a sum of a random number of independent summands. Accordingly, let $N$ be a nonnegative integer-valued random variable, independent of $\xi_1, \xi_2, \ldots,$ with p.g.f. $g_N(s) = E\left[s^N\right]$, and consider the random sum (see Chapter 2, Section 2.3).

$$X = \xi_1 + \cdots + \xi_N.$$

Let $h_X(s) = E\left[s^X\right]$ be the p.g.f. of $X$. We claim that $h_X(s)$ takes the simple form

$$h_X(s) = g_N[\phi(s)]. \tag{3.110}$$

To establish (3.110), consider

$$
\begin{aligned}
h_X(s) &= \sum_{k=0}^{\infty} \Pr\{X = k\} s^k \\
&= \sum_{k=0}^{\infty} \left( \sum_{n=0}^{\infty} \Pr\{X = k | N = n\} \Pr\{N = n\} \right) s^k \\
&= \sum_{k=0}^{\infty} \left( \sum_{n=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k | N = n\} \Pr\{N = n\} \right) s^k \\
&= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k\} \Pr\{N = n\} s^k \\
&\quad \text{[because } N \text{ is independent of } \xi_1, \xi_2, \ldots] \\
&= \sum_{n=0}^{\infty} \left( \sum_{k=0}^{\infty} \Pr\{\xi_1 + \cdots + \xi_n = k\}_{s^k} \right) \Pr\{N = n\} \\
&= \sum_{n=0}^{\infty} \phi(s)^n \Pr\{N = n\} \quad \text{[using (3.109)]} \\
&= g_n[\phi(s)] \quad \text{[by the definition of } g_n(s)].
\end{aligned}
$$

With the aid of (3.110), the basic branching process equation

$$
X_{n+1} = \xi_1^{(n)} + \cdots + \xi_{X_n}^{(n)} \tag{3.111}
$$

can be expressed equivalently and succinctly by means of generating functions. To this end, let $\phi_n(s) = E[s^{X_n}]$ be the p.g.f. of the population size $X_n$ at generation $n$, assuming that $X_0 = 1$. Then easily, $\phi_0(s) = E[s^1] = s$, and $\phi_1(s) = \phi(s) = E[s^{\xi}]$. To obtain the general expression, we apply (3.110) to (3.111) to yield

$$
\phi_{n+1}(s) = \phi_n[\phi(s)]. \tag{3.112}
$$

This expression may be iterated in the manner

$$
\begin{aligned}
\phi_{n+1}(s) &= \phi_{n-1}\{\phi[\phi(s)]\} \\
&= \underbrace{\phi\{\cdots\phi[\phi(s)]\}}_{(n+1)\text{ iterations}} \\
&= \phi[\phi_n(s)].
\end{aligned} \tag{3.113}
$$

That is, we obtain the generating function for the population size $X_n$ at generation $n$, given that $X_0 = 1$, by repeated substitution in the probability generating function of the offspring distribution.

For general initial population sizes $X_0 = k$, the p.g.f. is

$$\sum_{j=0}^{\infty} \Pr\{X_n = j | X_0 = k\} s^j = [\phi_n(s)]^k, \tag{3.114}$$

exactly that of a sum of $k$ independent lines of descents. From this perspective, the branching process evolves as the sum of $k$ independent branching processes, one for each initial parent.

**Example** Let $\phi(s) = q + ps$, where $0 < p < 1$ and $p + q = 1$. The associated branching process is a pure death process. In each period, each individual dies with probability $q$ and survives with probability $p$. The iterates $\phi_n(s)$ in this case are readily determined, e.g., $\phi_2(s) = q + p, (q + ps) = 1 - p^2 + p^2 s$, and generally, $\phi_n(s) = 1 - p^n + p^n s$. If we follow (3.114), the $n$th generation p.g.f. starting from an initial population size of $k$ is $[\phi_n(s)]^k = [1 - p^n + p^n s]^k$.

The probability distribution of the time $T$ to extinction may be determined from the p.g.f. as follows:

$$\begin{aligned}\Pr\{T = n | X(0) = k\} &= \Pr\{X_n = 0 | X_0 = k\} - \Pr\{X_{n-1} = 0 | X_0 = k\} \\ &= [\phi_n(0)]^k - [\phi_{n-1}(0)]^k \\ &= \left(1 - p^n\right)^k - \left(1 - p^{n-1}\right)^k.\end{aligned}$$

### 3.9.3 Multiple Branching Processes

Population growth processes often involve several life history phases (e.g., juvenile, reproductive adult, senescence) with different viability and behavioral patterns. We consider a number of examples of branching processes that take account of this characteristic.

For the first example, suppose that a mature individual produces offspring according to the p.g.f. $\phi(s)$. Consider a population of immature individuals, each of which grows to maturity with probability $p$ and then reproduces independently of the status of the remaining members of the population. With probability $1 - p$, an immature individual will not attain maturity and thus will leave no descendants. With probability $p$, an individual will reach maturity and reproduce a number of offspring determined according to the p.g.f. $\phi(s)$. Therefore, the progeny size distribution (or equivalently the p.g.f.) of a typical immature individual taking account of both contingencies is

$$(1 - p) + p\phi(s). \tag{3.115}$$

If a census is taken of individuals at the adult (mature) stage, the aggregate number of mature individuals contributed by a mature individual will now have p.g.f.

$$\phi(1 - p + ps). \tag{3.116}$$

(The student should verify this finding.)

It is worth emphasis that the p.g.f.s (3.115) and (3.116) have the same mean $p\phi'(1)$ but generally not the same variance, the first being

$$p\left[\phi''(1)+\phi'(1)-(\phi'(1))^2\right]$$

as compared with

$$p^2\phi''(1)+p\phi'(1)-p^2(\phi'(1))^2.$$

**Example** A second example leading to (3.116), as opposed to (3.115), concerns the different forms of mortality that affect a population. We appraise the strength (*stability*) of a population as the probability of indefinite survivorship = 1− probability of eventual extinction.

In the absence of mortality, the offspring number $X$ of a single individual has the p.g.f. $\phi(s)$. Assume, consistent with the postulates of a branching process, that all offspring in the population behave independently governed by the same probability laws. Assume also an adult population of size $X = k$. We consider three types of mortality:

**(a)** *Mortality of Individuals* Let $p$ be the probability of an offspring surviving to reproduce, independently of what happens to others. Thus, the contribution of each litter (family) to the adult population of the next generation has a binomial distribution with parameters $(N,p)$, where $N$ is the progeny size of the parent with p.g.f. $\phi(s)$. The p.g.f. of the adult numbers contributed by a single parent is, therefore, $\phi(q+ps), q = 1-p$, and for the population as a whole is

$$\psi_1(s) = [\phi(q+ps)]^k. \tag{3.117}$$

This type of mortality might reflect predation on adults.

**(b)** *Mortality of Litters* Independently of what happens to other litters, each litter survives with probability $p$ and is wiped out with probability $q = 1-p$. That is, given an actual litter size $\xi$, the effective litter size is $\xi$ with probability $p$, and 0 with probability $q$. The p.g.f. of adults in the following generation is accordingly

$$\psi_2(s) = [q+p\phi(s)]^k. \tag{3.118}$$

This type of mortality might reflect predation on juveniles or on nests and eggs in the case of birds.

**(c)** *Mortality of Generations* An entire generation survives with probability $p$ and is wiped out with probability $q$. This type of mortality might represent environmental catastrophes (e.g., forest fire, flood). The p.g.f. of population size in the next generation in this case is

$$\psi_3(s) = q+p[\phi(s)]^k. \tag{3.119}$$

All the p.g.f.s (3.117) through (3.119) have the same mean but usually different variances.

It is interesting to assess the relative stability of these three models. That is, we need to compare the smallest positive roots of $\psi_i(s) = s, i = 1, 2, 3$, which we will denote by $s_i^*, i = 1, 2, 3$, respectively.

We will show by convexity analysis that

$$\psi_1(s) \le \psi_2(s) \le \psi_3(s).$$

A function $f(x)$ is convex in $x$ if for every $x_1$ and $x_2$ and $0 < \lambda < 1$, then $f[\lambda x_1 + (1-\lambda)x_2] \le \lambda f(x_1) + (1-\lambda) f(x_2)$. In particular, the function $\phi(s) = \sum_{k=0}^{\infty} p_k s^k$ for $0 < s < 1$ is convex in $s$, since for each positive integer $k$, $[(\lambda s_1) + (1-\lambda)s_2]^k \le \lambda s_1^k + (1-\lambda)s_2^k$ for $0 < \lambda, s_1, s_2 < 1$. Now, $\psi_1(s) = [\phi(q+ps)]^k < [q\phi(1) + p\phi(s)]^k = [q + p\phi(s)]^k = \psi_2(s)$, and then $s_1^* < s_2^*$. Thus, the first model is more stable than the second model.

Observe further that due to the convexity of $f(x) = x^k, x > 0$, $\psi_2(s) = [p\phi(s) + q]^k < p[\phi(s)]^k + q \times 1^k = \psi_3(s)$, and thus $s_2^* < s_3^*$, implying that the second model is more stable than the third model. In conjunction we get the ordering $s_1^* < s_2^* < s_3^*$.

## Exercises

**3.9.1** Suppose that the offspring distribution is Poisson with mean $\lambda = 1.1$. Compute the extinction probabilities $u_n = \Pr\{X_n = 0 | X_0 = 1\}$ for $n = 0, 1, \ldots, 5$. What is $u_\infty$, the probability of ultimate extinction?

**3.9.2** Determine the probability generating function for the offspring distribution in which an individual either dies, with probability $p_0$, or is replaced by two progeny, with probability $p_2$, where $p_0 + p_2 = 1$.

**3.9.3** Determine the probability generating function corresponding to the offspring distribution in which each individual produces 0 or $N$ direct descendants, with probabilities $p$ and $q$, respectively.

**3.9.4** Let $\phi(s)$ be the generating function of an offspring random variable $\xi$. Let $Z$ be a random variable whose distribution is that of $\xi$, but conditional on $\xi > 0$. That is,

$$\Pr\{Z = k\} = \Pr\{\xi = k | \xi > 0\} \quad \text{for } k = 1, 2, \ldots.$$

Express the generating function for $Z$ in terms of $\phi$.

## Problems

**3.9.1** One-fourth of the married couples in a far-off society have no children at all. The other three-fourths of couples have exactly three children, with each child equally likely to be a boy or a girl. What is the probability that the male line of descent of a particular husband will eventually die out?

**3.9.2** One-fourth of the married couples in a far-off society have exactly three children. The other three-fourths of couples continue to have children until the first boy and then cease childbearing. Assume that each child is equally likely to be a boy or girl. What is the probability that the male line of descent of a particular husband will eventually die out?

**3.9.3** Consider a large region consisting of many subareas. Each subarea contains a branching process that is characterized by a Poisson distribution with parameter $\lambda$. Assume, furthermore, that the value of $\lambda$ varies with the subarea, and its distribution over the whole region is that of a gamma distribution. Formally, suppose that the offspring distribution is given by

$$\pi(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k = 0, 1, \ldots,$$

where $\lambda$ itself is a random variable having the density function

$$f(\lambda) = \frac{\theta^\alpha \lambda^{\alpha-1} e^{-\theta\lambda}}{\Gamma(\alpha)} \quad \text{for } \lambda > 0,$$

where $\theta$ and $\alpha$ are positive constants. Determine the marginal offspring distribution $p_k = \int \pi(k|\lambda) f(\lambda) d\lambda$.

**Hint:** Refer to the last example of Chapter 2, Section 2.4.

**3.9.4** Let $\phi(s) = 1 - p(1-s)^\beta$, where $p$ and $\beta$ are constants with $0 < p, \beta < 1$. Prove that $\phi(s)$ is a probability generating function and that its iterates are

$$\phi_n(s) = 1 - p^{1+\beta+\cdots+\beta^{n-1}}(1-s)^{\beta^n} \quad \text{for } n = 1, 2, \ldots.$$

**3.9.5** At time 0, a blood culture starts with one red cell. At the end of 1 min, the red cell dies and is replaced by one of the following combinations with the probabilities as indicated:

| | |
|---|---|
| Two red cells | $\frac{1}{4}$ |
| One red, One white | $\frac{2}{3}$ |
| Two white | $\frac{1}{12}$ |

Each red cell lives for 1 min and gives birth to offspring in the same way as the parent cell. Each white cell lives for 1 min and dies without reproducing. Assume that individual cells behave independently.

**(a)** At time $n + \frac{1}{2}$ min after the culture begins, what is the probability that no white cells have yet appeared?

**(b)** What is the probability that the entire culture eventually dies out entirely?

**3.9.6** Let $\phi(s) = as^2 + bs + c$, where $a, b, c$ are positive and $\phi(1) = 1$. Assume that the probability of extinction is $u_\infty$, where $0 < u_\infty < 1$. Prove that $u_\infty = c/a$.

**3.9.7** Families in a certain society choose the number of children that they will have according to the following rule: If the first child is a girl, they have exactly one more child. If the first child is a boy, they continue to have children until the first girl and then cease childbearing. Let $\xi$ be the number of male children in a particular family. What is the generating function of $\xi$? Determine the mean of $\xi$ directly and by differentiating the generating function.

**3.9.8** Consider a branching process whose offspring follow the geometric distribution $p_k = (1-c)c^k$ for $k = 0, 1, \ldots$, where $0 < c < 1$. Determine the probability of eventual extinction.

**3.9.9** One-fourth of the married couples in a distant society have no children at all. The other three-fourths of couples continue to have children until the first girl and then cease childbearing. Assume that each child is equally likely to be a boy or girl.

**(a)** For $k = 0, 1, 2, \ldots$, what is the probability that a particular husband will have $k$ male offspring?

**(b)** What is the probability that the husband's male line of descent will cease to exist by the fifth generation?

**3.9.10** Suppose that in a branching process the number of offspring of an initial particle has a distribution whose generating function is $f(s)$. Each member of the first generation has a number of offspring whose distribution has generating function $g(s)$. The next generation has generating function $f$, the next has $g$, and the distributions continue to alternate in this way from generation to generation.

**(a)** Determine the extinction probability of the process in terms of $f(s)$ and $g(s)$.

**(b)** Determine the mean population size at generation $n$.

**(c)** Would any of these quantities change if the process started with the $g(s)$ process and then continued to alternate?