

This chapter explores the properties of a broadly applicable probability model called a *Markov chain*, named after Russian mathematician A. A. Markov (1856–1922). Markov observed that many real-world phenomena can be modeled as a sequence of “transitions” from one “state” to another, with each transition having some associated uncertainty. For example, a taxi driver might “transition” between several towns (or zones within a large city); each time he drops off a passenger, he can’t be certain where his next fare will want to go. Similarly, a gambler might think of her winnings as transitioning from one “state”—really, a dollar amount—to another; with each round of the game she plays, she cannot be certain whether that dollar amount will go up or down (though, obviously, she hopes it goes up!). The same could be said for modeling the daily closing prices of a stock: each new day, there is uncertainty about whether that stock will “transition” to a higher or lower value, and this uncertainty could be modeled using the tools of probability.

In all of these examples, aside from the probability model for how transitions occur, one extra piece of information is critical: the current “state” (where the taxi driver is, how much money the gambler has). After all, if the gambler is making \$5 wagers, how much money she might have after the next game depends on how much she has now—if she currently holds \$45 in chips, then at the end of the upcoming round she can only have \$40 or \$50 on an even bet. The model structure proposed by Markov applies to situations where *only* knowledge of the current state, and the nature of transitions, is necessary—we don’t care how our gambler arrived at \$45 in chips, only that that’s how much she currently possesses.

Section 6.1 introduces basic notation for Markov chains and provides a rigorous definition of the property alluded to in the previous paragraph. In Sects. 6.2 and 6.3 we explain how the use of matrix notation can facilitate Markov chain computations. Section 6.4 focuses on a special class of Markov chains, so-called *regular chains*, which have a rather exceptional property embodied in the Steady-State Theorem. Section 6.5 considers a different class of Markov chains, those with one or more “inescapable” states, such as a gambler going broke. Finally, Sect. 6.6 discusses the simulation of Markov chains using software.

6.1 Terminology and Basic Properties

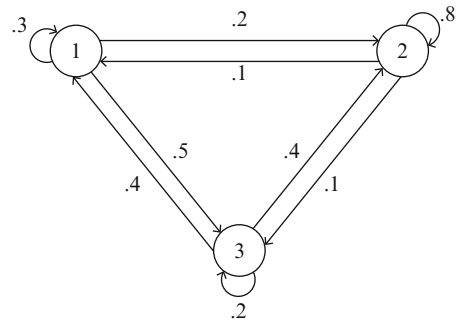
Markov chains provide a model for sequential information that allows future outcomes to depend on previous ones, albeit in a very specific way (the defining *Markov property*). Researchers in numerous

fields employ Markov chains to model the phenomena they study. Recent examples include predicting changes in electricity demand; modeling the motion of sperm whales off the Galapagos Islands; Chinese citizens changing their cell phone service; keeping track of inpatient bed usage at hospitals; monitoring patterns in Web browser histories to deploy better-targeted advertising; the evolution of drought conditions over time; and the dynamics of capital assets.

This first section introduces the basic vocabulary and notation of Markov chains. We begin with the following classic (if slightly artificial) example, which will serve as a thread throughout the chapter.

Example 6.1 A city has three different taxi zones, numbered 1, 2, and 3. A taxi driver operates his cab in all three zones. The probability that his next passenger has a destination in a particular one of these zones depends on where the passenger is picked up. Specifically, whenever the taxi driver is in zone 1, the probability his next passenger is going to zone 1 is .3, to zone 2 is .2, and to zone 3 is .5. Starting in zone 2, the probability his next passenger is going to zone 1 is .1, to zone 2 is .8, and to zone 3 is .1. Finally, whenever he is in zone 3, the probability his next passenger is going to zone 1 is .4, to zone 2 is .4, and to zone 3 is .2. These probabilities are encapsulated in the **state diagram** in Fig. 6.1.

Fig. 6.1 State diagram for Example 6.1



In every such state diagram, the sum of the probabilities on branches exiting any state must equal 1. For example, in Fig. 6.1 the probabilities exiting state 2 (i.e., zone 2) are .1, .8, and .1. We include in this calculation the probability .8 indicated by a “loop” in the state diagram, which simply means that the taxi driver has .8 probability of staying in zone 2 once he has dropped off a fare in zone 2.

Define X_0 to be the zone in which the taxi driver starts and X_n ($n \geq 1$) to be the zone where he drops off his n th fare. Since X_0, X_1, X_2, \dots “occur” in sequence, they are often referred to as a *chain*. More precisely, this particular sequence is a *finite-state, discrete-time, time-homogeneous Markov chain*. Each of these terms will be explained shortly. ■

In Example 6.1, each of the X_n for $n \geq 0$ assumes the value 1, 2, or 3 according to the destination zone. The zones collectively constitute the **states** of our chain, and so the **state space** is {zone 1, zone 2, zone 3}, although we will often drop the state names and just use the integers {1, 2, 3}. States can be identified with physical locations, levels (such as high/medium/low), dollar amounts, or just about anything else. We’ll sometimes refer to the X_n as random variables, even though they are not necessarily numerical (which goes against the definition from Chap. 2). The random variable X_0 is called the **initial state** of the chain. A **discrete-space** chain is one for which the number of possible states is finite or countably infinite. If there are finitely many possible states, we have a **finite-state** chain.

Since time was indexed by the discrete listing $n = 0, 1, 2, \dots$, the sequence of zones the taxi driver visited in Example 6.1 is called a **discrete-time** chain. Section 7.7 gives an overview of **continuous-time** chains, often indexed as $\{X_t; t \in [0, \infty)\}$, which are useful for modeling behavior continuously in time rather than just at discrete time points (e.g., tracking over time the number of people looking at a particular Web site). The taxi driver chain is also **time-homogeneous**, in that the specified probabilities do not change over time. One could imagine a different, more complicated model where the probabilities specified in Example 6.1 apply during morning hours but not in the evening, so that the probability of taking a fare from zone 1 to zone 3 is .5 for $n = 1$ (beginning of the work day) but is .1, say, for $n = 20$ (end of his shift). See Exercises 78 and 79 for examples of nonhomogeneous Markov chains.

Example 6.2 This is a simple version of the famous *Gambler’s Ruin* problem, which we previously considered in Exercise 145 of Chap. 1. Allan and Beth play a succession of independent games for \$1 each. Suppose Allan starts with \$2 and Beth with \$1, and the chance of Allan winning \$1 is p on each game. Ties are not allowed, so the chance of Beth winning \$1 on any particular game is $1 - p$. They compete until one of the two players goes broke (has \$0).

For $n = 0, 1, 2, \dots$, define $X_n =$ the amount of money Allan has after n games. The initial state has been specified as $X_0 = \$2$; Allan’s successive holdings X_0, X_1, X_2, \dots form our chain. The state space for X_n is $\{\$0, \$1, \$2, \$3\}$ or just $\{0, 1, 2, 3\}$, so we again have a finite-state chain. The state space and the specified probabilities are illustrated by the state diagram in Fig. 6.2. Notice we have included two “loops” with probability 1 at \$0 and \$3—these reflect the constraint that the game stops once Allan reaches one of these dollar amounts. That is, once Allan is “at” \$3, he will stay at \$3, and the same goes for \$0. Also, it will always be understood that if no arrow points from state i to state j in such a diagram, then the probability of moving from state i immediately into state j (i.e., in one time step) is zero.

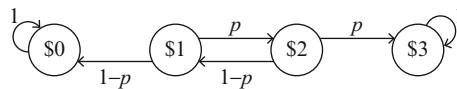


Fig. 6.2 State diagram for Example 6.2

Example 6.3 *A random walk.* Imagine a marker initially placed at 0 on the number line. A fair coin is flipped repeatedly; each head moves the marker one integer to the right, while each tail moves it one integer to the left. Let $X_0 = 0$, the initial state, and $X_n =$ the marker’s position after n coin flips for $n \geq 1$. Each member of the chain can only take on a finite set of values: X_1 is either +1 or -1 , X_2 is one of $-2, 0$, or 2 , and so on. However, the collection of *all* possible states across *all* time indices comprises the entire set of integers: $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$. Thus, this so-called “random walk” is an infinite-state (though still discrete-state) chain; it is partially illustrated in Fig. 6.3.

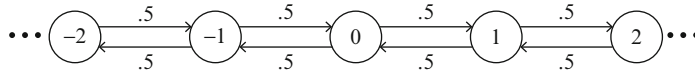


Fig. 6.3 State diagram for Example 6.3

6.1.1 The Markov Property

All of the preceding examples have an important feature known as the *Markov property*. Loosely speaking, it says that in order to know where the chain will go next (say, X_{n+1}), it suffices to know where the chain is now (the value of X_n). In particular, once the current state is specified, the path that brought the chain to that state is irrelevant. Consider, for example, the random walk of Example 6.3: if for any particular n we have $X_n = 4$, then we know $X_{n+1} = 3$ or 5 with probability $.5$ each. It does not matter whether the chain arrived at 4 quickly ($0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$) or by a more circuitous route; the probability distribution of the next state in the chain is the same. This notion is formalized in the following definition.

DEFINITION

Let X_0, X_1, X_2, \dots be a sequence of random variables (a chain) on some discrete state space. The sequence is said to have the **Markov property** if, for any time index n and any set of (not necessarily distinct) states $s_0, s_1, \dots, s_n, s_{n+1}$,

$$P(X_{n+1} = s_{n+1} | X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = P(X_{n+1} = s_{n+1} | X_n = s_n) \quad (6.1)$$

Such a sequence $\{X_n; n = 0, 1, 2, \dots\}$ is called a **Markov chain**.

The conditional probabilities specified in Eq. (6.1) are called the **one-step transition probabilities** of the chain, or sometimes just *transition probabilities*. These are precisely the probabilities specified in Examples 6.1–6.3. It's critical to recognize that these are *conditional* probabilities: they specify the likelihood of the next member of the chain X_{n+1} being in any particular state, *given* the current state of the chain X_n .

Example 6.4 (Example 6.1 continued) The sequence of successive zones visited by our taxi driver is characterized by nine one-step transition probabilities. For example, it is stated that the driver “transitions” from zone 1 to zone 3 with probability $.5$, which means that for any time index n ,

$$P(X_{n+1} = 3 | X_n = 1) = .5$$

This probability does not depend on the value of n , because the chain is time-homogeneous. Instead of writing $P(X_{n+1} = 3 | X_n = 1) = .5$, we will sometimes abbreviate with $P(1 \rightarrow 3) = .5$ to emphasize the idea of transitioning from one state to another. Thus, the complete set of one-step transition probabilities for the taxi driver is

$$\begin{array}{lll} P(1 \rightarrow 1) = .3 & P(1 \rightarrow 2) = .2 & P(1 \rightarrow 3) = .5 \\ P(2 \rightarrow 1) = .1 & P(2 \rightarrow 2) = .8 & P(2 \rightarrow 3) = .1 \\ P(3 \rightarrow 1) = .4 & P(3 \rightarrow 2) = .4 & P(3 \rightarrow 3) = .2 \end{array} \quad \blacksquare$$

Example 6.5 (Example 6.2 continued) The changing fortunes of Allan are governed by six (non-zero) transition probabilities:

$$\begin{aligned}
 P(1 \rightarrow 0) &= 1 - p & P(1 \rightarrow 2) &= p & P(2 \rightarrow 1) &= 1 - p & P(2 \rightarrow 3) &= p \\
 P(0 \rightarrow 0) &= 1 & P(3 \rightarrow 3) &= 1
 \end{aligned}$$

The last two probabilities above correspond to termination of the sequence of games. From a mathematical (if not practical) perspective, they communicate the idea that the chain marches on even when gameplay has ended (e.g., $2 \rightarrow 3 \rightarrow 3 \rightarrow 3 \dots$). That is, the conditional probability $P(3 \rightarrow 3) = P(X_{n+1} = 3 | X_n = 3) = 1$ indicates that if Allan has all \$3 at stake after n games, he will retain his \$3 while some imaginary future gameplay continues (the $(n + 1)$ st game, the $(n + 2)$ nd game, etc.). This convention eliminates the need to “stop” the Markov chain at some particular time point n . We’ll elaborate much more on this in Sect. 6.5.

In addition, there are ten one-step transition probabilities that equal zero; for example, according to the rules of Gambler’s Ruin, $P(1 \rightarrow 3) = 0$, and $P(3 \rightarrow x) = 0$ for $x \in \{0, 1, 2\}$. In general, a finite-state Markov chain with s states is specified by s^2 one-step transition probabilities, although it is quite common for many (if not most) of these to be zero. ■

Example 6.6 Markov chains are often used to model changing weather conditions; research literature in both meteorology and climate science is rife with Markov chain applications. The article “To Ski or Not to Ski: Estimating Transition Matrices to Predict Tomorrow’s Snowfall Using Real Data” (*J. of Statistics Educ.*, vol. 18, no. 3, 2010) provides data for several US cities on the daily transitions between “snow days,” defined by a snow depth of at least 50 mm, and “green days” (snow depth < 50 mm). Let X_n represent the snow status, either S for snow or G for green, on the n th recorded day. For New York City, the following one-step transition probabilities are provided:

$$P(G \rightarrow G) = .964 \quad P(G \rightarrow S) = .036 \quad P(S \rightarrow G) = .224 \quad P(S \rightarrow S) = .776$$

If today is a “green day” in New York, then there is a 96.4% chance that tomorrow’s snow depth will also be below 50 mm, based on the available weather data (which, incidentally, stretches back to the year 1912 for New York). On the other hand, as the author notes, “the presence of a significant snow depth (accumulation) on the current day in Central Park (New York) has an approximately 1 in 5 chance of melting before the next day.” ■

Not all sequences of random variables possess the Markov property. In econometrics (statistical methodology applied to economic scenarios), for example, most models for the closing price X_{n+1} of a stock on the $(n + 1)$ st day of trading incorporate not only the previous day’s closing price X_n but also information from many previous days (the data X_{n-1} , X_{n-2} , and so on). The likelihood that X_{n+1} will be \$5 higher than X_n may depend on the stock’s behavior over all of last week, not just where it closed on day n .

That said, in some instances a model that includes more than a one-time-step dependence can be modified by reconfiguring the state space in such a way that it satisfies the Markov property. This expansion of states is illustrated in the next example.

Example 6.7 The weather model presented in Example 6.6 satisfies the Markov property; in particular, it assumes that one can model tomorrow’s weather based on today’s conditions without incorporating any previous information. A more realistic model might assume that tomorrow’s snow depth depends on today’s *and* yesterday’s weather. Suppose, for example, that tomorrow will be a snow day with probability .8 if both yesterday and today were snow days; with probability .6 if today

was a snow day but yesterday was a green day; with probability .3 if it was green today and snowy yesterday; and with probability .1 if both previous days were green.

Once again let X_n = the “state” of the weather on day n : G for green day, S for snow day. Then the sequence X_0, X_1, X_2, \dots of weather states does *not* satisfy the Markov property, because the conditional distribution of X_{n+1} given all previous weather information depends on both X_n and X_{n-1} (the previous two days’ weather conditions). So, let’s make the following modification: define Y_n to be the ordered pair

$$Y_n = (\text{day } n \text{ weather, day } n + 1 \text{ weather}) = (X_n, X_{n+1})$$

So, for example, if snow depth was ≥ 50 mm on day 4 but < 50 mm on day 5, then $Y_4 = (S, G)$. The weather on day 6 depends on these previous 2 days, but they are now both contained in a single “variable,” Y_4 . In other words, Y_5 can be modeled entirely by knowing Y_4 : Y_5 ’s first entry, X_5 , matches the second entry of Y_4 , and the probability distribution of the second entry of Y_5 (i.e., X_6) is determined by the rules given at the beginning of this example.

With this modification, the sequence Y_0, Y_1, Y_2, \dots forms a Markov chain. The state space of this chain is not $\{S, G\}$, but rather $\{(G, G), (G, S), (S, G), (S, S)\}$. The earlier weather rules can be expressed as one-step transition probabilities for this chain:

$$\begin{aligned} P((S, S) \rightarrow (S, S)) &= .8 & P((S, G) \rightarrow (G, S)) &= .3 \\ P((G, S) \rightarrow (S, S)) &= .6 & P((G, G) \rightarrow (G, S)) &= .1 \end{aligned}$$

Four other transition probabilities can be found by considering the complements of the given transition events. The final eight transition probabilities (with four states, there are $4^2 = 16$ total one-step transition probabilities) are all 0, e.g., $P((S, G) \rightarrow (S, S)) = 0$, because if $Y_n = (S, G)$ then it was “green” on day $n + 1$ ($X_{n+1} = G$), meaning the first entry of Y_{n+1} must also be G . ■

The remainder of this chapter will focus almost exclusively on finite-state, discrete-time, time-homogeneous chains; these are the most commonly encountered models in practice. The case of infinite-state chains, including the random walk of Example 6.3, is considered in several more advanced texts; see, for example, the book *Introduction to Probability Models* by Ross listed in the references.

6.1.2 Exercises: Section 6.1 (1–10)

1. The article “Markov Chain Models of Negotiators’ Communication” (*Encyclopedia of Peace Psychology* 2012: 608-612) describes the following set-up for the back and forth dialogue between two negotiators. If at any stage a negotiator engages in a cooperative strategy, the other negotiator will respond with a cooperative strategy with probability .6. Otherwise, the response is described as a competitive strategy. Similarly, there is probability .7 that a competitive strategy offered at any stage of the negotiations will be met by another competitive strategy. Let X_n = the strategy employed at the n th stage of the negotiation. Identify the state space for the chain, specify its one-step transition probabilities, and draw the corresponding state diagram.
2. Imagine m balls being exchanged between two adjacent chambers (left and right) according to the following rules. At each time step, one of the m balls is randomly selected and moved to the opposite chamber, i.e., if the selected ball is currently in the right chamber, it will be moved to the left one, and vice versa. Let X_n = the number of balls in the left chamber after

the n th exchange. (This is called an *Ehrenfest chain*, a model often used to describe the movement of gas molecules.)

- (a) Identify the state space of this chain.
 - (b) Suppose $m = 3$. Specify the one-step transition probabilities for this chain. [*Hint*: It might be helpful to draw the two chambers and the possible positions of the three balls.]
 - (c) Draw the state diagram corresponding to (b).
 - (d) Generalize the probabilities in (b) to the case of m balls.
3. A certain machine used in a manufacturing process can be in one of three states: fully operational (“full”), partially operational (“part”), or broken. If the machine is fully operational today, there’s a .7 probability it will be fully operational again tomorrow, a .2 chance it will be partially operational tomorrow, and otherwise tomorrow it will be broken. If the machine is partially operational today, there is a .6 probability it will continue to be partially operational tomorrow and otherwise it will be broken (because the machine is never repaired in its partially operational state). Finally, if the machine is broken today, there is a .8 probability it will be repaired to fully operational status tomorrow; otherwise, it remains broken. Let X_n = the state of the machine on day n .
- (a) Identify the state space of this chain.
 - (b) Determine the complete set of one-step transition probabilities, and draw the corresponding state diagram.
4. Michelle will flip a coin until she gets heads four times in a row. Define $X_0 = 0$ and, for $n \geq 1$, X_n = the number of heads in the current streak of heads after the n th flip.
- (a) If the first seven flips result in the sequence *HTHHHTH*, determine the values of X_1, X_2, \dots, X_7 . [*Hint*: Each time Michelle flips tails, the streak is reset to 0.]
 - (b) Is this an example of a Markov chain? Explain why or why not.
 - (c) Identify the state space of the chain. Treat reaching four heads in a row in the same manner that the \$3 state was treated in the Gambler’s Ruin scenario of Example 6.2.
 - (d) Assume $P(H) = p$ for this particular coin. Determine the one-step transition probabilities of this chain, and draw the corresponding state diagram.
5. A single cell has probability p of dividing into two cells and probability $1 - p$ of dying without dividing. Once two new cells have been created, each has the same probability p of splitting in two, independent of the other. In this fashion, cells continue to divide, either indefinitely or until all cells are dead (extinction of the cell line). Let X_n = the number of cells in the n th generation, with $X_0 = 1$ to reflect the initial, single cell.
- (a) What are the possible numerical values of X_1 , and what are their probabilities?
 - (b) What are the possible numerical values of X_2 ?
 - (c) Determine the one-step transition probabilities for this chain. That is, given there are x cells in the n th generation ($X_n = x$), determine the conditional probability distribution of X_{n+1} .
- [*Note*: This is an example of a *branching process*, commonly known as a *Galton-Watson process*. See Exercise 163 at the end of Chap. 4 for information on determining the probability of eventual extinction.]
6. Imagine a set of stacked files, such as papers on your desk. Occasionally, you will need to retrieve one of these files, which you will find by “sequential search”: looking at the first paper in the stack, then the second, and so on until you find the document you require. A sensible *sequential search algorithm* is to place the most recently retrieved file at the top of the stack, the idea being that files accessed more often will “rise to the top” and thus require less searching in the long run. For simplicity’s sake, imagine such a scenario with just three files, labeled A, B, C .

- (a) Let X_n represent the sequence of the entire stack after the n th search. For example, if the files are initially stacked A on top of B on top of C , then $X_0 = ABC$. Determine the state space for this chain.
- (b) If $X_0 = ABC$, list all possible states for X_1 . [*Hint*: One of the three files will be selected and rise to the front of the stack. Is every arrangement listed in (a) possible, starting from ABC ?]
- (c) Suppose that, at any given time, there is probability p_A that file A must be retrieved, p_B that file B must be retrieved, and similarly for p_C ($= 1 - p_A - p_B$). Determine all of the non-zero one-step transition probabilities.
7. Social scientists have used Markov chains to study “social mobility,” the movement of people between social classes, for more than a century. In a typical such model, states are defined as social classes, e.g., lower class, middle class, and upper class. The time index n refers to a familial generation, so if X_n represents a man’s social class, then X_{n-1} is his father’s social class, X_{n-2} his grandfather’s, and so on.
- (a) In this context, what would it mean for X_n to be a Markov chain? In particular, would that imply that a grandfather’s social class has no bearing on his grandson’s? Explain.
- (b) What would it mean for this chain to be time-homogeneous? Does that seem realistic? Explain why or why not.
8. The article “Markov Chain Models for Delinquency: Transition Matrix Estimation and Forecasting” (*Appl. Stochastic Models Bus. Ind.*, 2011: 267-279) classifies loan status into four categories: current (payments are up-to-date), delinquent (payments are behind but still being made), loss (payments have stopped permanently), and paid (the loan has been paid off). Let X_n = the status of a particular loan in its n th month, and assume (as the authors do) that X_n is a Markov chain.
- (a) Suppose that, for one particular loan type, $P(\text{delinquent} \rightarrow \text{current}) = .1$ and $P(\text{current} \rightarrow \text{delinquent}) = .3$. Interpret these probabilities.
- (b) According to the definitions of the “loss” and “paid” states, what are $P(\text{loss} \rightarrow \text{loss})$ and $P(\text{paid} \rightarrow \text{paid})$? [*Hint*: Refer back to Example 6.2.]
- (c) Draw the state diagram for this Markov chain.
- (d) What would it mean for this Markov chain to be time-homogeneous? Does that seem realistic? Explain.
9. The article cited in Exercise 1 also suggests a more complex negotiation model, wherein the strategy employed at the n th stage (cooperative or competitive) is predicted not only by the immediately preceding action but also the one before it. So, negotiator A’s next strategy is determined not only by negotiator B’s most recent move, but also by A’s choice just before that. Again, let X_n = the negotiating strategy used at the n th stage.
- (a) Is X_n a Markov chain? Explain.
- (b) How could you modify this example to create a Markov chain? What additional information would you need to completely specify this chain? [*Hint*: See Example 6.7.]
10. Let X_0, X_1, X_2, \dots be a sequence of *independent* discrete rvs taking values in some common state space.
- (a) Show that X_n satisfies the Markov property. (That is, all sequences of independent rvs on a common state space are trivially discrete-space Markov chains.)
- (b) What additional condition(s), if any, must be satisfied for X_n to be a time-homogeneous Markov chain?

6.2 The Transition Matrix and the Chapman–Kolmogorov Equations

Section 6.1 introduced the notion of a Markov chain and its characteristic one-step transition probabilities. In this section, we will develop a systematic way to determine the probability that a chain moves from one state to another in *two* steps (or three or four ...) by considering all the intermediate paths the chain may have taken. Such calculations are facilitated by aggregating the transition probabilities into a matrix.

6.2.1 The Transition Matrix

The one-step transition probabilities for the taxi driver example were displayed in Example 6.4 as a 3×3 array. It would be more efficient to simply specify the probabilities themselves in that same format, with the understanding that the probability in the i th row and j th column indicates the transition probability $P(i \rightarrow j)$, the chance the taxi driver takes his next fare to zone j given that he picks up the fare in zone i . Such a representation will be critical to understanding how various multistep transition probabilities are calculated.

DEFINITION

Let X_0, X_1, X_2, \dots be a finite-state, time-homogeneous Markov chain, and index the states of the chain by the positive integers $1, 2, \dots, s$. The **(one-step) transition matrix** of the Markov chain is the $s \times s$ matrix \mathbf{P} whose (i, j) th entry is given by

$$p_{ij} = P(i \rightarrow j) = P(X_{n+1} = j | X_n = i)$$

for $i = 1, \dots, s$ and $j = 1, \dots, s$.

Example 6.8 (Example 6.4 continued) The one-step transition matrix for our taxi driver example is

$$\mathbf{P} = \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix}$$

which is identical in format to the display in Example 6.4. The entries are interpreted as the preceding definition suggests, e.g., the upper left entry (first row, first column) of the matrix is

$$p_{11} = P(1 \rightarrow 1) = P(X_{n+1} = 1 | X_n = 1) = .3,$$

i.e., the conditional probability that his next fare is dropped off somewhere in zone 1 given that the taxi is currently in zone 1. ■

Example 6.9 (Example 6.5 continued) For the Gambler's Ruin scenario with a total available fortune of \$3, rather than label the four possible states as 1, 2, 3, 4, it's more natural to use state labels 0, 1, 2, and 3 corresponding to Allan's fortune at any particular time. The transition probabilities specified previously may be written as the following 4×4 matrix:

$$\mathbf{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The labels along the left-hand side of the matrix indicate the ordering of the states for the purpose of creating this matrix; they are not, strictly speaking, a part of \mathbf{P} . For example, $P(X_{n+1} = 1|X_n = 2) = P(\text{Allan loses the next game}) = 1 - p$, while $P(X_{n+1} = 3|X_n = 0) = 0$. ■

Example 6.10 (Example 6.6 continued) The snow depth model has only two states, S (snowy day) and G (“green” day). The one-step transition probabilities given for New York City can be summarized by the following 2×2 matrix:

$$\mathbf{P} = \begin{array}{c} G \\ S \end{array} \begin{bmatrix} .964 & .036 \\ .224 & .776 \end{bmatrix} \quad \blacksquare$$

Notice that the entries of every row in all of the preceding transition matrices sums to 1. This will always be the case: given that the chain is currently in some state i , it has to go somewhere in its next step (even if that entails remaining in state i). That is, for any state i and any time index n , we must have

$$\sum_{j=1}^s p_{ij} = \sum_{j=1}^s P(i \rightarrow j) = \sum_{j=1}^s P(X_{n+1} = j|X_n = i) = 1$$

6.2.2 Computation of Multistep Transition Probabilities

We now turn to the determination of multistep transition probabilities. Given that a Markov chain is currently in state i , what is the probability it will be in state j two steps later (i.e., after two transitions)? Three steps later? We begin with the following definition.

DEFINITION

Let X_0, X_1, X_2, \dots be a time-homogeneous Markov chain. For any positive integer k , the **k -step transition probabilities** are defined by

$$P^{(k)}(i \rightarrow j) = P(X_{n+k} = j|X_n = i) \quad (6.2)$$

where i and j range across the states of the chain (typically $1, \dots, s$). For $k = 1$, we will typically revert to the previous notation: $P^{(1)}(i \rightarrow j) = P(i \rightarrow j)$.

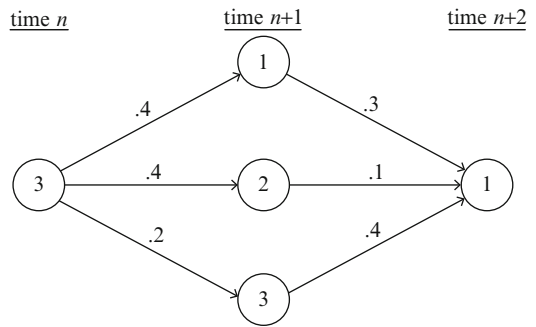
The superscript (k) in Expression (6.2) does not indicate taking the k th power; it is simply notation representing “in k steps.” The next example illustrates how these k -step transition probabilities can be calculated, and how they can be represented compactly in terms of powers of the one-step transition matrix.

Example 6.11 (Example 6.8 continued) Suppose our taxi driver just dropped off a fare in zone 3, so that that is his current state. What is the probability that his *second* fare, counting from now, takes him

to zone 1? That is, we wish to determine $P(X_{n+2} = 1|X_n = 3) = P^{(2)}(3 \rightarrow 1)$. The calculation method is suggested by Fig. 6.4. Consider all the possible destinations of the $(n + 1)$ st fare, i.e., all the intermediate steps the taxi driver could take from zone 3 to zone 1, and then employ the Law of Total Probability (applied here to conditional probabilities).

The partitioning events in the Law of Total Probability are the possible states at time $n + 1$:

Fig. 6.4 Transitioning from state 3 to state 1 in two time steps



$$\begin{aligned}
 P^{(2)}(3 \rightarrow 1) &= P(X_{n+2} = 1|X_n = 3) \\
 &= P(X_{n+1} = 1|X_n = 3)P(X_{n+2} = 1|X_n = 3, X_{n+1} = 1) \\
 &\quad + P(X_{n+1} = 2|X_n = 3)P(X_{n+2} = 1|X_n = 3, X_{n+1} = 2) \\
 &\quad + P(X_{n+1} = 3|X_n = 3)P(X_{n+2} = 1|X_n = 3, X_{n+1} = 3)
 \end{aligned}$$

By the Markov property, $P(X_{n+2} = 1|X_n = 3, X_{n+1} = 1) = P(X_{n+2} = 1|X_{n+1} = 1)$, and the other two probabilities involving conditioning on X_n and X_{n+1} simplify analogously. Thus,

$$\begin{aligned}
 P^{(2)}(3 \rightarrow 1) &= P(X_{n+1} = 1|X_n = 3)P(X_{n+2} = 1|X_{n+1} = 1) \\
 &\quad + P(X_{n+1} = 2|X_n = 3)P(X_{n+2} = 1|X_{n+1} = 2) \\
 &\quad + P(X_{n+1} = 3|X_n = 3)P(X_{n+2} = 1|X_{n+1} = 3) \\
 &= P(3 \rightarrow 1)P(1 \rightarrow 1) + P(3 \rightarrow 2)P(2 \rightarrow 1) + P(3 \rightarrow 3)P(3 \rightarrow 1) \\
 &= (.4)(.3) + (.4)(.1) + (.2)(.4) = .24
 \end{aligned}$$

For later reference, the last expression could be written in terms of the elements of the transition matrix \mathbf{P} ; specifically, it's $p_{31}p_{11} + p_{32}p_{21} + p_{33}p_{31}$.

Similarly, the conditional probability that his second fare wants to be dropped off in zone 2 is computed by

$$\begin{aligned}
 P^{(2)}(3 \rightarrow 2) &= P(3 \rightarrow 1)P(1 \rightarrow 2) + P(3 \rightarrow 2)P(2 \rightarrow 2) + P(3 \rightarrow 3)P(3 \rightarrow 2) \\
 &= \sum_{j=1}^3 P(3 \rightarrow j)P(j \rightarrow 2) = \sum_{j=1}^3 p_{3j}p_{j2} \\
 &= (.4)(.2) + (.4)(.8) + (.2)(.4) = .48
 \end{aligned}$$

Finally, the probability the taxi driver finds himself back in zone 3 after two fares is

$$P^{(2)}(3 \rightarrow 3) = \sum_{j=1}^3 p_{3j}p_{j3} = (.4)(.5) + (.4)(.1) + (.2)(.2) = .28$$

This makes sense, since the taxi driver must arrive in one of the three zones at time $n + 2$, and $1 - (.24 + .48) = .28$. ■

The sums of products of matrix entries that appear repeatedly in the preceding example should look familiar: they are the same manner of computation that arises when one matrix is multiplied by another (or, here, a matrix is multiplied by itself). Indeed, consider what happens if we multiply the one-step transition matrix \mathbf{P} from Example 6.8 by itself:

$$\mathbf{P}^2 = \mathbf{P}\mathbf{P} = \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} = \begin{bmatrix} .31 & .42 & .27 \\ .15 & .70 & .15 \\ .24 & .48 & .28 \end{bmatrix}$$

The entries in the bottom row—.24, .48, .28—are precisely the two-step transition probabilities computed in Example 6.11. Specifically, the (3, 1) entry of \mathbf{P}^2 is $P^{(2)}(3 \rightarrow 1) = .24$, the (3, 2) entry of \mathbf{P}^2 is $P^{(2)}(3 \rightarrow 2) = .48$, and the (3, 3) entry of \mathbf{P}^2 is $P^{(2)}(3 \rightarrow 3) = .28$. It should come as no surprise that the other six entries of \mathbf{P}^2 follow the same pattern: the (i, j) entry of \mathbf{P}^2 is equal to $P^{(2)}(i \rightarrow j)$. Hence, we can obtain all nine two-step transition probabilities with a single matrix computation (which, of course, can be facilitated by Matlab or other matrix-capable software).

The foregoing result can be generalized to an arbitrary fixed number of steps: to find the *three*-step transition probabilities, for example, one only needs to compute the matrix \mathbf{P}^3 . It is not necessary to consider explicitly the many different paths by which the Markov chain could transition from state i to state j in three steps and add up all the corresponding probabilities (this is, secretly, what the threefold matrix multiplication does). The most general result is often referred to as the set of *Chapman–Kolmogorov Equations*.

CHAPMAN–KOLMOGOROV EQUATIONS

If a Markov chain has one-step transition matrix \mathbf{P} , then the k -step transition probabilities are the entries of the matrix \mathbf{P}^k . Specifically,

$$P^{(k)}(i \rightarrow j) = \text{the } (i, j) \text{ entry of } \mathbf{P}^k$$

Example 6.12 (Example 6.11 continued) Back to our intrepid taxi driver: if he just dropped off a fare in zone 2, what is the probability that he will be in zone 1 two fares later? That is, we wish to determine the two-step transition probability $P(X_{n+2} = 1 | X_n = 2) = P^{(2)}(2 \rightarrow 1)$. According to the Chapman–Kolmogorov Equations, this is simply the (2, 1) entry of the foregoing matrix \mathbf{P}^2 :

$$P^{(2)}(2 \rightarrow 1) = .15$$

Now consider a longer-term question: If the taxi driver starts the day in zone 3 and transports ten fares before lunch, what is the probability he ends up “back home” (i.e., in zone 3) for lunch? The goal is to find $P(X_{10} = 3 | X_0 = 3) = P^{(10)}(3 \rightarrow 3)$, which could involve summing up a terrifying number of intermediate travel options (19,683 of them, to be precise!). But the Chapman–Kolmogorov Equations, coupled with computer software, makes light work of the problem. With the aid of Matlab, the tenth power of \mathbf{P} is found to be

$$\mathbf{P}^{10} = \begin{bmatrix} .2004 & .5993 & .2004 \\ .1998 & .6004 & .1998 \\ .2002 & .5996 & .2002 \end{bmatrix}$$

The desired probability is just the (3, 3) entry of this 10-step transition matrix: $P^{(10)}(3 \rightarrow 3) = .2002$. ■

Example 6.13 The report “Research and Application by Markov Chain Operators in the Mobile Phone Market” (Second International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011) details an analysis of customer loyalty and movement between China’s three major cell phone service providers: (1) China Telecom, (2) China Unicom, and (3) China Mobile. A “transition” in this setting refers to an opportunity for a customer to renew his or her contract with a current provider or else switch to one of the other two companies. The report includes the following one-step transition matrix, with the companies numbered as above:

$$\mathbf{P} = \begin{bmatrix} .84 & .06 & .10 \\ .08 & .82 & .10 \\ .10 & .04 & .86 \end{bmatrix}$$

The entries along the main diagonal indicate customer loyalty, e.g., 84% of China Telecom customers stick with that company when their contract expires.

Suppose a customer is currently with China Unicom. What is the probability she will be with the same service provider three contracts from now? In other words, what is $P^{(3)}(2 \rightarrow 2)$? According to the Chapman–Kolmogorov Equations, we need the (2, 2) entry of \mathbf{P}^3 . That matrix is computed to be

$$\mathbf{P}^3 = \begin{bmatrix} .6310 & .1352 & .2338 \\ .1920 & .5742 & .2338 \\ .2267 & .1006 & .6727 \end{bmatrix}$$

from which we may extract $P^{(3)}(2 \rightarrow 2) = .5742$.

It’s important to distinguish this probability from the answer to a more restrictive question: what’s the chance she stays with China Unicom for *all* of her next three cell phone contracts? This probability can be represented as $P(X_{n+1} = 2 \cap X_{n+2} = 2 \cap X_{n+3} = 2 | X_n = 2)$ or, less formally, as $P(2 \rightarrow 2 \rightarrow 2 \rightarrow 2)$. Applying the Markov property gives $[P(2 \rightarrow 2)]^3 = p_{22}^3 = (.82)^3 = .5514$. This probability is slightly lower than $P^{(3)}(2 \rightarrow 2) = .5742$, since the latter accounts for the possibility that the customer switches companies at some intermediate stage(s) but ends up back with China Unicom three contracts later. ■

Example 6.14 (Example 6.9 continued) Suppose in our earlier Gambler’s Ruin example that $p = .55$; that is, Allan has a 55% chance of winning any particular \$1 game. The one- and two-step transition matrices are as follows:

$$\mathbf{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ .45 & 0 & .55 & 0 \\ 0 & .45 & 0 & .55 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{P}^2 = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ .45 & .2475 & 0 & .3025 \\ .2025 & 0 & .2475 & .55 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

As before, Allan starts with \$2. Looking at the \$2 row (i.e., the third row) of \mathbf{P}^2 , there is a .2025 probability he has gone broke after two games. This is easy to compute by hand: since he could only lose \$2 in two games by losing twice, the chance is $(.45)^2 = .2025$. The chance that he is back to where he started after two games (i.e., $X_2 = \$2$) is the (\$2, \$2) entry of \mathbf{P}^2 : .2475. This also could have occurred in just one way: $\$2 \rightarrow \$1 \rightarrow \$2$, for which the two-step transition probability is $(.45)(.55) = .2475$. Notice that the (\$2, \$1) entry of \mathbf{P}^2 is 0, i.e., $P^{(2)}(\$2 \rightarrow \$1) = 0$. Since exactly \$1 exchanges hands at the end of each game, it's impossible for Allan to transition from \$2 to \$1 in exactly two steps. Finally, observe that the (\$2, \$3) entry of both matrices is .55, so $P(\$2 \rightarrow \$3) = P^{(2)}(\$2 \rightarrow \$3) = .55$. That's because the game ends when Allan has all \$3 at stake, which he could achieve in one step with probability $p = .55$. Having done so, he will "stay at \$3" in the imaginary second game/step, i.e., from a mathematical perspective, the observed sequence of the Markov chain steps X_0, X_1 , and X_2 is $\$2 \rightarrow \$3 \rightarrow \$3$, with the second transition occurring with probability 1.

A natural concern from Allan's perspective is the likelihood that he will *eventually* win. One way to estimate that probability is to look at the chance Allan has arrived at the \$3 "state" after some large number of steps. (This works because once he has \$3, he will always remain at \$3.) Matlab can easily calculate high powers of small matrices; we requested \mathbf{P}^{75} :

$$\mathbf{P}^{75} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5980 & 0 & .0000 & .4020 \\ .2691 & .0000 & 0 & .7309 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The two entries that read .0000 indicate that the probability is not strictly 0, but rather is 0 to four decimal places. From this matrix, we have that

$$P(\text{Allan eventually has } \$3 \mid X_0 = \$2) \approx P(\text{Allan has } \$3 \text{ after 75 steps} \mid X_0 = \$2) \\ = P^{(75)}(\$2 \rightarrow \$3) = .7309$$

Had Allan started with just \$1, he would have a roughly .4020 chance of eventually winning all the money.

In Sect. 6.5, we will present an exact method for determining the probability that Allan eventually wins (or loses) his competition with Beth. ■

6.2.3 Exercises: Section 6.2 (11–22)

- The authors of the article "The Fate of Priority Areas for Conservation in Protected Areas: A Fine-Scale Markov Chain Approach" (*Envir. Mgmt.*, 2011: 263–278) postulated the following model for landscape changes in the forest regions of Italy. Each "pixel" on a map is classified as forested (F) or non-forested (NF). For any specific pixel, X_n represents its status n years after 2000 (so X_1 corresponds to 2001, X_2 to 2002, and so on). Their analysis showed that a pixel has a

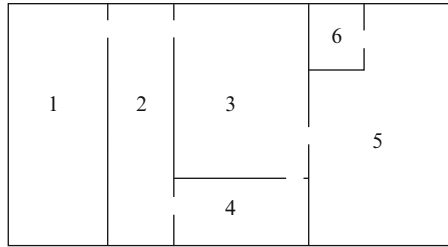
90% chance of being forested next year if it is forested this year and an 11% chance of being forested next year if it non-forested this year; moreover, data in the twenty-first century are consistent with the assumptions of a Markov chain.

- (a) Construct the one-step transition matrix for this chain, with states $1 = F$ and $2 = NF$.
 - (b) If a map pixel was forested in the year 2000, what is the probability it was still forested in 2002? 2013?
 - (c) If a map pixel was non-forested in the year 2000, what is the probability it was still non-forested in 2002? 2013?
 - (d) The article’s authors use this model to project forested status for several Italian regions in the years 2050 and 2100. Comment on the assumptions required for these projections to be valid.
12. A large automobile insurance company classifies its customers into four risk categories (1 being the lowest risk, aka best/safest drivers, 4 being the worst/highest risk; premiums are assigned accordingly). Each year, upon renewal of a customer’s insurance policy, the risk category may change depending on the number of accidents in the previous year. Actuarial data suggest the following: category 1 customers stay in category 1 with probability .9 and move to categories 2, 3, 4 with probabilities .07, .02, and .01, respectively. Category 2 customers shift to category 1 (based on having no accidents last year) with probability .8 and rise to risk categories 3 and 4 with probabilities .15 and .05, respectively. Similarly, category 3 customers transition to 2 and 4 with probabilities .7 and .3, while category 4 customers stay in that risk category with probability .4 and move to category 3 otherwise.
- (a) Let X_n denote a customer’s risk category for his/her n th year with the insurance company. Construct the one-step transition matrix for this Markov chain.
 - (b) If a customer starts in category 1, what is the probability she falls into risk category 2 five years later?
 - (c) If a customer is currently in risk category 4, determine the probability he will be a category 1 driver in k years, for $k = 1, 2, 3, 4, 5, 6$.
 - (d) What is the probability that a driver currently in category 1 remains in that category for each of the next 5 years?
13. The article cited in Example 6.6 also gives the following one-step transition matrix, with the same definitions of states, for Willow City, ND:

$$\mathbf{P} = \begin{matrix} G \\ S \end{matrix} \begin{bmatrix} .933 & .067 \\ .012 & .988 \end{bmatrix}$$

- (a) Contrast Willow City with New York City: where is snow more likely to stay on the ground for an extended time period? Explain.
 - (b) If today is a snowy day in Willow City, what is the probability it will also be a snowy day there 2 days from now? three days from now?
 - (c) If today is a snowy day in Willow City, what is the probability it will continue to be snowy for the next 4 days in a row?
14. I (author Carlton) have a six-room house whose configuration is depicted in the accompanying diagram. When my sister and her family visit, I often play hide-and-seek with my young nephew, Lucas. Consider the following situation: Lucas counts to ten in Room 1, while I run and hide in Room 6. Lucas’ “strategy,” as much as he has one, is such that standing in any room of the house, he is equally likely to next visit any of the adjacent rooms, regardless of where he

has searched previously. (The exception, of course, is if he enters Room 6, in which case he discovers me and the round of hide-and-seek is over.)



- (a) Let X_n = the n th room Lucas visits (with $X_0 = 1$, his starting point). Construct the one-step transition matrix for the corresponding Markov chain.
 - (b) What is the probability that his third room-to-room transition will take him into Room 2?
 - (c) What is the fewest number of time steps (i.e., room transitions) required for Lucas to find me?
 - (d) What is the probability that, after 12 time steps, he still hasn't found me?
15. Refer back to Exercise 1 in the previous section. Consider two negotiators, A and B, who employ strategies according to the Markov chain model described.
- (a) Construct the one-step transition matrix for the Markov chain X_n = strategy employed at the n th stage of a negotiation, assuming the states are (1) cooperative and (2) competitive.
 - (b) If negotiator A employs a cooperative strategy at some stage, what is the probability she uses a competitive strategy the next time? [Don't forget that A's turns are two time steps apart, since B counter-negotiates in between.]
 - (c) Now introduce a third state, (3) end of the negotiation. Assume that a Markov chain model with the following one-step transition matrix applies:

$$\mathbf{P} = \begin{bmatrix} .6 & .2 & .2 \\ .3 & .4 & .3 \\ 0 & 0 & 1 \end{bmatrix}$$

Given that the initial strategy presented is cooperative, what is the probability the negotiations end within three time steps?

- (d) Refer back to (c). Given that the initial strategy presented is competitive, what is the probability the negotiations end within three time steps?
16. Sarah, a statistician at a large Midwestern polling agency, owns four umbrellas. Initially, two of them are at her home and two are at her office. Each morning, she takes an umbrella with her to work (assuming she has any at home) if and only if it's currently raining, which happens on 20% of mornings. Each evening, she takes an umbrella from work to home (again, assuming any are available) if and only if it's raining when she leaves work, which happens on 30% of all evenings. Assume weather conditions, including morning and evening on the same day, are independent (in the Midwest, that's not unrealistic). Let X_n = the number of umbrellas Sarah has at home at the end of her n th work day (i.e., once she's back at home).
- (a) Identify the state space for this chain.
 - (b) Assume Sarah has two umbrellas at home tonight. By considering all possible weather conditions tomorrow morning and tomorrow evening, determine the one-step transition probabilities for the number of umbrellas she'll have at home tomorrow night.

- (c) Repeat the logic of (b) to determine the complete one-step transition matrix for this chain. Be careful when considering the two extreme cases!
17. Refer back to the previous exercise.
- (a) Given that Sarah has two umbrellas at home (and two at work) as of Sunday night, what is the probability she'll have exactly two umbrellas at home the following Friday night? What is the probability she'll have *at least* two umbrellas at home the following Friday night?
- (b) Given that Sarah has two umbrellas at home Sunday night, what are the chances she won't have an umbrella to take with her to work the following Thursday morning when a surprise thunderstorm moves through the area?
- (c) Assume again that Sarah has two umbrellas at home at the start of the week. Determine the *expected* number of umbrellas she has at home at the end of Monday and at the end of Tuesday. [*Hint*: X_n is a discrete rv; if $X_0 = 2$, then the probability distribution of X_n appears in the corresponding row of \mathbf{P}^n .]
18. A box always contains three marbles, each of which is green or yellow. At regular intervals, one marble is selected at random from the box and removed, while another is put in its place according to the following rules: a green marble is replaced by a yellow marble with probability .3 (and otherwise by another green marble), while a yellow marble is equally likely to be replaced by either color. Let X_n = the number of green marbles in the box after the n th swap.
- (a) What are the possible values of X_n ?
- (b) Construct the one-step transition matrix for this Markov chain.
- (c) If all three marbles currently in the box are green, what is the probability the same will be true three swaps from now?
- (d) If all three marbles currently in the box are green, what is the probability that the fourth marble selected from the box will be green? [*Hint*: Use part (c). Be careful not to confuse the color of the marble *selected* on the fourth swap with the color of the one that replaces it!]
19. A Markov chain model for customer visits to an auto repair shop is described in the article “Customer Lifetime Value Prediction by a Markov Chain Based Data Mining Model: Application to an Auto Repair and Maintenance Company in Taiwan” (*Scientia Iranica*, 2012: 849-855). Customers make between 0 and 4 visits to the repair shop each year; for any customer that made exactly i visits last year, the number of visits s/he will make next year follows a Poisson distribution with parameter μ_i . (The event “4 visits” is really “4 or more visits,” so the probability of 4 visits next year is calculated as $1 - \sum_{x=0}^3 p(x; \mu_i)$ from the appropriate Poisson pmf.) Parameter values cited in the article, which were estimated from real data, appear in the accompanying table.

i	0	1	2	3	4
μ_i	1.938696	1.513721	1.909567	2.437809	3.445738

- (a) Construct the one-step transition matrix for the chain X_n = number of repair shop visits by a randomly selected customer in the n th observed year.
- (b) If a customer made two visits last year, what is the probability that s/he makes two visits next year and two visits the year after that?
- (c) If a customer made no visits last year, what is the probability s/he makes a total of exactly two visits in the next 2 years?
20. The four vans in a university's vanpool are maintained at night by a single mechanic, who can service one van per night (assuming any of them need repairs). Suppose that there is a 10% chance that a van working today will need service by tonight, independent of the status of the

other vans. We wish to model X_n = the number of vans *unavailable* for service at the beginning of the n th day.

- (a) Suppose all four vans were operational as of this morning. Find the probability that exactly j of them will be unusable tomorrow morning for $j = 0, 1, 2, 3$. [*Hint*: The number of unusable vans for tomorrow will be 1 less than the number that break down today, unless that's 0, because the mechanic can fix only one van per night. What is the probability distribution of Y = the number of vans that break down today, assuming all 4 worked this morning?]
 - (b) Suppose three vans were operational as of this morning, and one was broken. Find the probabilities $P(1 \rightarrow j)$ for this chain. [*Hint*: Assume the broken van will be fixed tonight. Then the number of unavailable vans tomorrow morning is the number that break down today, out of the three currently functioning.]
 - (c) Use reasoning similar to that of (a) and (b) to determine the complete one-step transition matrix for this Markov chain.
21. Refer back to the previous exercise.
- (a) If all four vans were operational as of Monday morning, what is the probability exactly three vans will be usable Wednesday morning? Thursday morning? Friday morning?
 - (b) A *backlog* occurs whenever $X_n \geq 1$, indicating that some vans will be temporarily out of commission because the mechanic could not get to them the previous night. Assuming there was no backlog as of Monday morning, what is the probability a backlog exists Tuesday morning? Answer the same question for Wednesday, Thursday, and Friday mornings.
 - (c) How do the probabilities in (b) change if there was a backlog of 1 van as of Monday morning?
22. Consider a Markov chain with state space $\{1, 2, \dots, s\}$. Show that, for any positive integers m and n and any states i and j ,

$$P_{ij}^{(m+n)} = \sum_{k=1}^s P_{ik}^{(m)} P_{kj}^{(n)}$$

This is an alternative version of the Chapman–Kolmogorov Equations. [*Hint*: Write the left-hand side as $P(X_{m+n} = j | X_0 = i)$, and consider all the possible states after m transitions.]

6.3 Specifying an Initial Distribution

Thus far, every probability we have considered in this chapter (i.e., all the one-, two-, and higher-step transition probabilities) has been conditional. For example the entries of any one-step transition matrix indicate $P(X_{n+1} = j | X_n = i)$. In this section, we briefly explore unconditional probabilities, which result from specifying a distribution for the rv X_0 , the initial state of the chain. We will consider two cases: modeling the initial state X_0 as a random variable, and treating X_0 as having a fixed/known value.

Example 6.15 (Example 6.11 continued) The never-ending saga of the taxi driver continues! Imagine this poor fellow sleeps in his taxi, so from his perspective each new day starts in a “random” zone. Specifically, suppose for now that he has a 20% chance of waking up in zone 1, a 50% chance of

waking up in zone 2, and a 30% chance of waking up in zone 3. That is, we have assigned the following **initial distribution** to the Markov chain:

$$\begin{array}{c|ccc}
 i & 1 & 2 & 3 \\
 \hline
 P(X_0 = i) & .2 & .5 & .3
 \end{array} \tag{6.3}$$

Notice that, unlike the conditional probabilities that comprise the transition matrix of the Markov chain, this initial distribution (6.3) specifies the *unconditional* (aka marginal) distribution for the rv X_0 . In what follows, we will sometimes refer to the bottom row of (6.3) as the “initial probability vector” of X_0 .

Now consider the rv X_1 , the destination of the taxi driver’s first fare. The probability his first fare wants to go somewhere in zone 3 can be determined via the Law of Total Probability:

$$\begin{aligned}
 P(X_1 = 3) &= P(X_0 = 1)P(X_1 = 3|X_0 = 1) + P(X_0 = 2)P(X_1 = 3|X_0 = 2) \\
 &\quad + P(X_0 = 3)P(X_1 = 3|X_0 = 3) \\
 &= \sum_{i=1}^3 [P(X_0 = i)P(i \rightarrow 3)] = \sum_{i=1}^3 [P(X_0 = i)p_{i3}] \\
 &= (.2)(.5) + (.5)(.1) + (.3)(.2) = .21
 \end{aligned}$$

As indicated in the intermediate step, this unconditional probability can be computed by taking the product of the initial probability vector $[.2 \ .5 \ .3]$, regarded as a 1×3 matrix, with the third column of the one-step transition matrix \mathbf{P} . Similarly, the (unconditional) probability that his first fare wants to be dropped off in zone 2 is

$$\begin{aligned}
 P(X_1 = 2) &= \sum_{i=1}^3 [P(X_0 = i)P(i \rightarrow 2)] = \sum_{i=1}^3 [P(X_0 = i)p_{i2}] \\
 &= (.2)(.2) + (.5)(.8) + (.3)(.4) = .56
 \end{aligned}$$

The foregoing computation is the product of the initial probability vector with the second column of \mathbf{P} . Finally, the probability that the first fare is taken to zone 1 equals .23, which can be computed either as a similar product or by observing that $1 - (.21 + .56) = .23$. All together, the unconditional pmf of the rv X_1 is

$$\begin{array}{c|ccc}
 i & 1 & 2 & 3 \\
 \hline
 P(X_1 = i) & .23 & .56 & .21
 \end{array}$$

Clearly, the most efficient way to determine the distribution of X_1 is to compute all three products simultaneously through matrix multiplication. If we multiply the transition matrix \mathbf{P} on the left by a 1×3 row vector containing the initial probabilities for X_0 , we obtain

$$[.2 \ .5 \ .3]\mathbf{P} = [.2 \ .5 \ .3] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} = [.23 \ .56 \ .21]$$

■

The method illustrated in the preceding example can be generalized to find the unconditional distribution of the state X_n in the chain after any number of transitions n , starting with a specified initial distribution for X_0 .

THEOREM

Let $X_0, X_1, \dots, X_n, \dots$ be a Markov chain with state space $\{1, \dots, s\}$ and one-step transition matrix \mathbf{P} . Let $\mathbf{v}_0 = [v_{01} \dots v_{0s}]$ be a $1 \times s$ vector specifying the initial distribution of the chain, i.e., $v_{0k} = P(X_0 = k)$ for $k = 1, \dots, s$. If \mathbf{v}_1 denotes the vector of marginal (i.e., unconditional) probabilities associated with X_1 , then

$$\mathbf{v}_1 = \mathbf{v}_0 \mathbf{P}$$

More generally, if \mathbf{v}_n denotes the $1 \times s$ vector of marginal probabilities for X_n , then

$$\mathbf{v}_n = \mathbf{v}_0 \mathbf{P}^n$$

Proof The formula $\mathbf{v}_1 = \mathbf{v}_0 \mathbf{P}$ can be established using the same computational approach displayed in Example 6.15. Now consider \mathbf{v}_2 , the vector of unconditional probabilities for X_2 . By the same reasoning as in Example 6.15, we have

$$\mathbf{v}_2 = \mathbf{v}_1 \mathbf{P}$$

The substitution $\mathbf{v}_1 = \mathbf{v}_0 \mathbf{P}$ then yields $\mathbf{v}_2 = (\mathbf{v}_0 \mathbf{P}) \mathbf{P} = \mathbf{v}_0 \mathbf{P}^2$. Continuing by induction, we have for general n that $\mathbf{v}_n = \mathbf{v}_{n-1} \mathbf{P} = (\mathbf{v}_0 \mathbf{P}^{n-1}) \mathbf{P} = \mathbf{v}_0 \mathbf{P}^n$, as claimed. ■

With the aid of software such as Matlab, the unconditional distributions of future states of the Markov chain can be computed very quickly once the initial distribution is specified. For example, as a continuation of Example 6.15, the probability vector for X_2 , the destination of the driver's second fare, is given by

$$\mathbf{v}_2 = \mathbf{v}_1 \mathbf{P} = [.23 \quad .56 \quad .21] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} = [.209 \quad .578 \quad .213]$$

or, equivalently,

$$\mathbf{v}_2 = \mathbf{v}_0 \mathbf{P}^2 = [.2 \quad .5 \quad .3] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix}^2 = [.209 \quad .578 \quad .213]$$

That is, assuming that the initial distribution specified in Example 6.15 is correct, the taxi driver has a 20.9% chance of taking his second fare to zone 1, a 57.8% chance of taking him/her to zone 2, and a 21.3% chance of being in zone 3 after two fares.

Example 6.16 As you probably learned in high school biology, Austro-Hungarian scientist Gregor Mendel studied the inheritance of characteristics within plant species, particularly peas. Suppose one particular pea plant can either be green or yellow, which is determined by a single gene with green (G) dominant over yellow (g). That is, the genetic material determining a plant's color (its "genotype") can be one of three pairings— GG , Gg , or gg —depending on which types were passed on by

the parent plants. To say that green is “dominant” over yellow means that the plant’s visible color—its “phenotype”—will be green unless that gene is completely absent from the plant (so plants with GG or Gg genotype appear green, while only gg plants are yellow).

Consider cross-breeding with a yellow plant, whose genotype is therefore known to be gg . Mendel’s laws of genetic recombination can be expressed by the following transition matrix, where X_n is the genotype of an n th-generation plant resulting from cross-breeding with a gg plant:

$$\mathbf{P} = \begin{matrix} GG \\ Gg \\ gg \end{matrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & .5 & .5 \\ 0 & 0 & 1 \end{bmatrix}$$

For example, crossing $GG \times gg$ yields Gg with probability 1, while $Gg \times gg$ results in Gg or gg with probability .5 each.

Suppose our initial population of plants (to be cross-bred with the pure yellow specimens) has the following genotype distribution: 70% GG , 20% Gg and 10% gg . The initial probability vector associated with this “0th generation” is $\mathbf{v}_0 = [.7 \ .2 \ .1]$. The probabilities associated with the first generation of cross-bred plants is $\mathbf{v}_1 = \mathbf{v}_0\mathbf{P} = [0 \ .8 \ .2]$, meaning that 80% of first-generation plants are expected to be Gg and the remaining 20% gg . Notice that GG plants cannot exist past the first generation, since cross-breeding with gg plants makes such a recombination impossible.

Similarly, the second-generation probabilities are given by $\mathbf{v}_2 = \mathbf{v}_1\mathbf{P} = \mathbf{v}_0\mathbf{P}^2 = [0 \ .4 \ .6]$, so that within two generations gg plants should be the majority (60% gg compared to 40% Gg). As cross-breeding with pure gg plants continues, that genotype will increase in relative proportion (80% in generation 3, 90% in generation 4), until eventually the dominant G allele dies out. ■

6.3.1 A Fixed Initial State

The case in which the initial state X_0 is fixed or known rather than random can be handled by forming a “degenerate” initial probability distribution.

Example 6.17 (Example 6.15 continued) Suppose that our taxi driver lives in zone 3 and always goes home at night, which means that he starts each new day in zone 3. Starting with certainty in zone 3 means that $P(X_0 = 3) = 1$, while $P(X_0 = 1) = P(X_0 = 2) = 0$. Written as a pmf, the distribution of X_0 is

i	1	2	3
$P(X_0 = i)$	0	0	1

Equivalently, the probability vector for X_0 is $\mathbf{v}_0 = [0 \ 0 \ 1]$. From the original description of the Markov chain (Example 6.1), the initial state being zone 3 implies that $X_1 = 1$ with probability .4, $X_1 = 2$ with probability .4, and $X_1 = 3$ with probability .2. This same result can be obtained by applying the theorem of this section:

$$\mathbf{v}_1 = \mathbf{v}_0\mathbf{P} = [0 \ 0 \ 1] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} = [.4 \ .4 \ .2]$$

Notice that left-multiplying \mathbf{P} by the vector $[0 \ 0 \ 1]$ simply extracts the third row of \mathbf{P} . Similarly, the pmf of X_5 , the destination of the fifth passenger, is given by

$$\begin{aligned} \mathbf{v}_5 &= \mathbf{v}_0 \mathbf{P}^5 = [0 \quad 0 \quad 1] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix}^5 = [0 \quad 0 \quad 1] \begin{bmatrix} .2115 & .5767 & .2118 \\ .1938 & .6125 & .1938 \\ .2073 & .5858 & .2070 \end{bmatrix} \\ &= [.2073 \quad .5858 \quad .2070] \end{aligned}$$

The matrix \mathbf{P}^5 was computed by Matlab. The row vector \mathbf{v}_5 is simply the third row of \mathbf{P}^5 , because the chain begins in zone 3 with probability 1. So, starting the day at home in zone 3, the taxi driver finds himself in zone 1, 2, or 3 after five fares with probabilities .2073, .5858, and .2070, respectively. ■

Example 6.18 (Example 6.14 continued) As before, we can use a high power of the one-step transition matrix, say \mathbf{P}^{75} , to approximate the long-term behavior of our Gambler's Ruin Markov chain. Suppose as before that $p = .55$ and Allan's initial stake is \$2. We can express the latter as $\mathbf{v}_0 = [0 \ 0 \ 1 \ 0]$; recall that the states, in order, are \$0, \$1, \$2, \$3. Then the probability distribution of X_{75} is

$$\mathbf{v}_{75} = \mathbf{v}_0 \mathbf{P}^{75} = [0 \ 0 \ 1 \ 0] \mathbf{P}^{75} = \text{the third (i.e., \$2) row of } \mathbf{P}^{75} = [.2691 \ .0000 \ 0 \ .7309]$$

If Allan begins the competition with \$2 (and Beth with \$1), there is a 73.09% chance he will end up with all the money within 75 games, and a 26.91% chance he will end up broke after 75 games. As discussed previously, the competition will almost certainly end long before a 75th game, but for purposes of forecasting long-run behavior we imagine that when either player goes broke, game-play continues but no further money is exchanged.

Suppose instead that Allan's initial stake is just \$1, while Beth starts with \$2. Then Allan's initial "distribution" is specified by $\mathbf{v}_0 = [0 \ 1 \ 0 \ 0]$, meaning $P(X_0 = \$1) = 1$ while $P(X_0 = \$0, \$2, \$3) = 0$. After 75 plays, we now have

$$\mathbf{v}_{75} = \mathbf{v}_0 \mathbf{P}^{75} = [0 \ 1 \ 0 \ 0] \mathbf{P}^{75} = \text{the second (i.e., \$1) row of } \mathbf{P}^{75} = [.5980 \ 0 \ .0000 \ .4020]$$

Starting with \$1, Allan has a 40.2% chance of winning the competition (i.e., ending up with \$3) and a 59.8% chance of being "ruined." ■

6.3.2 Exercises: Section 6.3 (23–30)

23. Refer back to Exercise 1 of this chapter. Suppose that Negotiator A goes first and that 75% of the time she begins negotiations with a cooperative strategy. (Consider this to be time index 0.)
 - (a) Determine the (unconditional) probability that Negotiator B's first strategy will also be cooperative.
 - (b) Determine the (unconditional) probability that Negotiator B's second strategy will be cooperative. [*Hint*: Which time index corresponds to his second move?]
24. Refer back to the Ehrenfest chain model described in Exercise 2 with $m = 3$ balls. The possible states of the chain $X_n =$ number of balls in the left chamber after the n th exchange are $\{0, 1, 2, 3\}$.
 - (a) Suppose that all four possible initial states are equally likely. Determine the probability distributions of X_1 and X_2 .
 - (b) Suppose instead that each of the three balls is initially equally likely to be placed in the left or right chamber. In this situation, what is the initial distribution of the chain?

- (c) Using the initial distribution specified in (b), determine the unconditional distributions of X_1 and X_2 . What do you notice?
25. Information bits (0s and 1s) in a binary communication system travel through a long series of relays. At each relay, a “bit-switching” error might occur. Suppose that at each relay, there is a 4% chance of a 0 bit being switched to a 1 bit and a 5% chance of a 1 becoming a 0. Let $X_0 =$ a bit’s initial parity (0 or 1), and let $X_n =$ the bit’s parity after traversing the n th relay.
- (a) Construct the one-step transition matrix for this chain. [*Hint*: There are only two states, 0 and 1.]
- (b) Suppose the input stream to this relay system consists of 80% 0s and 20% 1s. Determine the proportions of 0s and 1s exiting the first relay.
- (c) Under the same conditions as (b), determine the proportions of 0s and 1s exiting the fifth relay.
26. Refer to the genetic recombination scenario of Example 6.16. Suppose that plants will now be cross-bred with known hybrids (i.e., those with genotype Gg). Mendel’s laws imply the following transition matrix for such breeding:

$$\mathbf{P} = \begin{matrix} & GG & Gg & gg \\ GG & \begin{bmatrix} .5 & .5 & 0 \end{bmatrix} \\ Gg & \begin{bmatrix} .25 & .5 & .25 \end{bmatrix} \\ gg & \begin{bmatrix} 0 & .5 & .5 \end{bmatrix} \end{matrix}$$

Again assume the initial population genotype distribution of plants to be cross-bred with these hybrids is 70% GG , 20% Gg , and 10% gg .

- (a) Determine the genotype distribution of the first generation of plants resulting from this cross-breeding experiment.
- (b) Determine the genotype distributions of the second, third, and fourth generations.
27. Refer to the weather scenario described in Example 6.6 and Example 6.10. Suppose today’s weather forecast for New York City gives a 20% chance of experiencing a snowy day.
- (a) Let X_0 denote today’s weather condition. Express the information provided as an initial probability vector for X_0 .
- (b) Determine the (unconditional) likelihoods of a snowy day and a green day tomorrow, using the one-step transition probabilities specified in Example 6.6.
- (c) Based on today’s forecast and the transition probabilities, what is the chance New York City will experience a “green day” 1 week (7 days) from now?
28. The article “Option Valuation Under a Multivariate Markov Chain Model” (Third International Joint Conference on Computational Science and Optimization, 2010) includes information on the dynamic movement of certain assets between three states: (1) up, (2) middle, and (3) down. For a particular class of assets, the following one-step transition probabilities were estimated from available data:

$$\mathbf{P} = \begin{bmatrix} .4069 & .3536 & .2395 \\ .3995 & .5588 & .0417 \\ .5642 & .0470 & .3888 \end{bmatrix}$$

Suppose that the initial valuation of this asset class found that 31.4% of such assets were in the “up” dynamic state, 40.5% were “middle,” and the remainder were “down.”

- (a) What is the initial probability vector for this chain?
- (b) Determine the unconditional probability distribution of X_1 , the asset dynamic state one time step after the initial valuation.
- (c) Determine the unconditional probability distribution of X_2 , the asset dynamic state two time steps after the initial valuation.

29. Refer back to Exercise 23, and now suppose that Negotiator A always opens talks with a competitive strategy.
- What is the probability vector for X_0 , Negotiator A's initial strategy?
 - Without performing any matrix computation, determine the distribution of X_1 , Negotiator B's first strategy choice.
 - What is the probability Negotiator A's second strategy is cooperative? competitive?
30. Transitions between sleep stages are described in the article "Multinomial Logistic Estimation of Markov-Chain Models for Modeling Sleep Architecture in Primary Insomnia Patients" (*J. Pharmacokinet. Pharmacodyn.*, 2010:137–155). The following one-step transition probabilities for the five stages awake (AW), stage 1 sleep (ST1), stage 2 sleep (ST2), slow-wave sleep (SWS), and rapid-eye movement sleep (REM) were obtained from a graph in the article:

$$\mathbf{P} = \begin{array}{l} \text{AW} \\ \text{ST1} \\ \text{ST2} \\ \text{SWS} \\ \text{REM} \end{array} \begin{bmatrix} .90 & .09 & .01 & .00 & .00 \\ .21 & .40 & .34 & .02 & .03 \\ .02 & .02 & .84 & .09 & .03 \\ .02 & .02 & .22 & .72 & .02 \\ .04 & .04 & .05 & .00 & .87 \end{bmatrix}$$

The time index of the Markov chain corresponds to half-hour intervals (i.e., $n = 1$ is 30 min after the beginning of the study, $n = 2$ is 60 min in, etc.). Initially, all patients in the study were awake.

- Let \mathbf{v}_0 denote the probability vector for X_0 , the initial state of a patient in the sleep study. Determine \mathbf{v}_0 .
- Without performing any matrix computations, determine the distribution of patients' sleep states 30 min (one time interval) into the study.
- Determine the distribution of patients' sleep states 4 h into the study. [*Hint*: What time index corresponds to the 4-h mark?]

6.4 Regular Markov Chains and the Steady-State Theorem

In previous sections, we have alluded to the long-term behavior of certain Markov chains. In some cases, such as Gambler's Ruin, we anticipate that the chain will eventually reach, and remain in, one of several "absorbing" states (we'll discuss these in Sect. 6.5). Our taxi driver, in contrast, should continually move around, but perhaps something can be said about how much time he will spend in each of the three zones over the course of many, many fares. It turns out that the taxi driver example belongs to a special class of Markov chains, called *regular chains*, for which the long-run behavior "stabilizes" in some sense and can be determined analytically.

6.4.1 Regular Chains

DEFINITION

A finite-state Markov chain with one-step transition matrix \mathbf{P} is said to be a **regular chain** if there exists a positive integer n such that all of the entries of the matrix \mathbf{P}^n are positive.

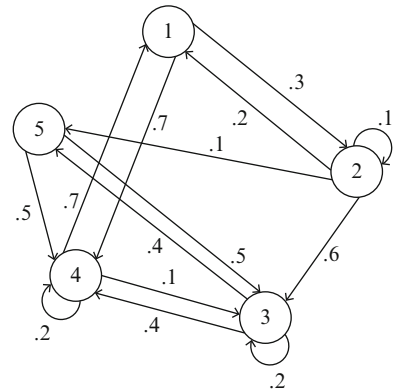
In other words, for a regular Markov chain there is some positive integer n such that every state can be reached from every state (including itself) in exactly n steps.

It's straightforward to show that if all the entries of \mathbf{P}^n are positive, then so are all of the entries of \mathbf{P}^{n+1} , \mathbf{P}^{n+2} , and so on (Exercise 37). Our taxi driver example is a regular chain, since all nine entries of \mathbf{P} itself are positive. The next example shows that a regular chain may have some one-step transition probabilities equal to zero.

Example 6.19 Internet users' browser histories can be modeled as Markov chains, where the "states" are different Web sites (or classes of Web sites) and transitions occur when users move from one Web site to another. The article "Evaluating Variable-Length Markov Chain Models for Analysis of User Web Navigation Sessions" (*IEEE Trans. Knowl. Data Engr.* 2007: 441-452) discusses increasingly complex models of this type. Suppose for simplicity that Web sites are grouped into five categories: (1) social media, (2) e-mail, (3) news and sports, (4) online retailers, and (5) other (use your imagination). Consider a Markov chain model for users' transitions between these five categories whose state diagram is depicted in Fig. 6.5.

Notice that, according to this model, not every state can access all five states in one step, because many one-step transition probabilities are zero. The one-step transition matrix \mathbf{P} of this Markov chain is as follows:

Fig. 6.5 State diagram for Example 6.19



$$\mathbf{P} = \begin{bmatrix} 0 & .3 & 0 & .7 & 0 \\ .2 & .1 & .6 & 0 & .1 \\ 0 & 0 & .2 & .4 & .4 \\ .7 & 0 & .1 & .2 & 0 \\ 0 & 0 & .5 & .5 & 0 \end{bmatrix}$$

Eleven of the twenty-five entries in \mathbf{P} are zero. However, consider several higher powers of this matrix:

$$\mathbf{P}^2 = \begin{bmatrix} .55 & .03 & .25 & .14 & .03 \\ .02 & .07 & .23 & .43 & .25 \\ .28 & 0 & .28 & .36 & .08 \\ .14 & .21 & .04 & .57 & .04 \\ .35 & 0 & .15 & .30 & .20 \end{bmatrix}, \quad \mathbf{P}^3 = \begin{bmatrix} .104 & .168 & .097 & .528 & .103 \\ .315 & .013 & .256 & .317 & .099 \\ .252 & .084 & .132 & .420 & .112 \\ .441 & .063 & .211 & .248 & .037 \\ .210 & .105 & .160 & .465 & .060 \end{bmatrix}$$

Since every entry of \mathbf{P}^3 is positive, by definition we have a regular Markov chain. Every state can reach every state (including itself) in exactly three moves. ■

In contrast, Gambler's Ruin is *not* a regular Markov chain. It is not possible for Allan to go from \$2 to \$1 in an even number of moves, so the (\$2, \$1) entry of \mathbf{P}^n is zero whenever n is even. Similarly, Allan cannot go from \$2 back to \$2 in an odd number of steps, so the (\$2, \$2) entry of \mathbf{P}^n equals zero for every odd exponent n . Thus, there exists no positive integer n for which all sixteen entries of \mathbf{P}^n are positive. (In fact, six other entries of \mathbf{P}^n must always be 0: $P^{(n)}(0 \rightarrow j) = 0$ for states $j \neq 0$ and $P^{(n)}(3 \rightarrow j) = 0$ for $j \neq 3$, since both \$0 and \$3 are “absorbing” states.) Another non-regular Markov chain, one that does not have any absorbing states, is given in the following example.

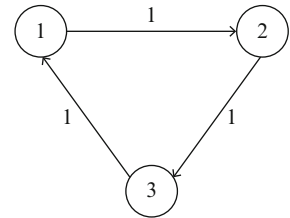
Example 6.20 Unlike our taxi driver, bus drivers follow a well-defined route. Consider a bus route from campus (state 1), to the nearby student housing complex (state 2), to downtown (state 3), and then back to campus. The associated Markov chain cycles endlessly: $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \dots$ Figure 6.6 shows the corresponding state diagram.

The one-step transition matrix for this chain is

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Direct computation shows that

Fig. 6.6 State diagram for Example 6.20



$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{P}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I},$$

where \mathbf{I} denotes the 3×3 identity matrix. Hence, $\mathbf{P}^4 = \mathbf{P}^3\mathbf{P} = \mathbf{IP} = \mathbf{P}$; $\mathbf{P}^5 = \mathbf{P}^3\mathbf{P}^2 = \mathbf{IP}^2 = \mathbf{P}^2$; $\mathbf{P}^6 = \mathbf{P}^3\mathbf{P}^3 = \mathbf{II} = \mathbf{I}$; and so on. That is, the n -step transition matrix \mathbf{P}^n equals one of \mathbf{P} , \mathbf{P}^2 , or \mathbf{I} for every positive integer n , and all three of these contain some zero entries. Therefore, this is *not* a regular Markov chain. ■

6.4.2 The Steady-State Theorem

What's so special about regular chains? The transition matrices of regular Markov chains exhibit a rather interesting property. Consider a very high power of the transition matrix for our taxi driver, computed with the aid of Matlab:

$$\mathbf{P} = \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} \Rightarrow \mathbf{P}^{100} = \begin{bmatrix} .2000 & .6000 & .2000 \\ .2000 & .6000 & .2000 \\ .2000 & .6000 & .2000 \end{bmatrix}$$

Notice that every row of \mathbf{P}^{100} is identical: roughly, each one is [.2 .6 .2]. What's more, raising \mathbf{P} to even higher powers yields the same matrix to several decimal places. That is, \mathbf{P}^{101} , \mathbf{P}^{102} , and so on are all roughly equal to \mathbf{P}^{100} . Something similar occurs for the regular Markov chain of Example 6.19:

$$\mathbf{P} = \begin{bmatrix} 0 & .3 & 0 & .7 & 0 \\ .2 & .1 & .6 & 0 & .1 \\ 0 & 0 & .2 & .4 & .4 \\ .7 & 0 & .1 & .2 & 0 \\ 0 & 0 & .5 & .5 & 0 \end{bmatrix} \Rightarrow \mathbf{P}^{100} = \begin{bmatrix} .2844 & .0948 & .1659 & .3791 & .0758 \\ .2844 & .0948 & .1659 & .3791 & .0758 \\ .2844 & .0948 & .1659 & .3791 & .0758 \\ .2844 & .0948 & .1659 & .3791 & .0758 \\ .2844 & .0948 & .1659 & .3791 & .0758 \end{bmatrix}$$

Again, every row of \mathbf{P}^{100} is the same, and replacing 100 by an even higher power gives the same result (i.e., to several decimal places $\mathbf{P}^{100} = \mathbf{P}^{101} = \mathbf{P}^{102} = \dots$). These are two examples of the central result in the theory of Markov chains, the so-called *Steady-State Theorem*.

STEADY-STATE THEOREM

Let \mathbf{P} be the one-step transition matrix of a finite-state, *regular* Markov chain. Then the matrix limit

$$\mathbf{\Pi} = \lim_{n \rightarrow \infty} \mathbf{P}^n \quad (6.4)$$

exists. Moreover, the rows of the limiting matrix $\mathbf{\Pi}$ are identical, with all positive entries.

The proof of the Steady-State Theorem is beyond the scope of this book; interested readers may consult the text by Karlin and Taylor listed in the references.

If we let $\boldsymbol{\pi} = [\pi_1 \cdots \pi_s]$ denote each of the identical rows of the limiting matrix $\mathbf{\Pi}$ in Eq. (6.4), $\boldsymbol{\pi}$ is called the **steady-state distribution** of the Markov chain. Thus, for the taxi driver example, the steady-state distribution is $\boldsymbol{\pi} = [.2 \ .6 \ .2]$, while the steady-state distribution for the Web browsing Markov chain in Example 6.19 is $\boldsymbol{\pi} = [.2844 \ .0948 \ .1659 \ .3791 \ .0758]$.

A Markov chain does not have to be regular for the limit of \mathbf{P}^n to exist as $n \rightarrow \infty$. For example, computing progressively larger powers of the one-step transition matrix for the Gambler's Ruin scenario of Example 6.14 shows that, for large n ,

$$\mathbf{P}^n \approx \mathbf{P}^{75} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5980 & 0 & .0000 & .4020 \\ .2691 & .0000 & 0 & .7309 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

That is, the limit of \mathbf{P}^n exists and is, at least to four decimal places, equal to the matrix displayed above. However, unlike in the case of a regular Markov chain, the rows of this limiting matrix are not identical and the matrix includes several zeros. We will consider in more detail Markov chains of this type in the next section.

The transition matrix of a "periodic" Markov chain, such as the one in Example 6.20, does not have a limit. This is not surprising, since periodic functions in general do not have long-run limits but rather cycle through their possible values.

6.4.3 Interpreting the Steady-State Distribution

The steady-state distribution $\boldsymbol{\pi}$ of a regular Markov chain can be interpreted in several ways. We present four different interpretations here; verifications of the second and fourth statements can be found in the Karlin and Taylor text.

1. If the “current” state of the Markov chain is observed after a large number of transitions, there is an approximate probability π_j of the chain being in state j . That is, for large n , $P(X_n = j) \approx \pi_j$. Moreover, this holds regardless of the initial distribution of the chain (i.e., the unconditional distribution of the initial state X_0).

The first sentence is essentially the definition of $\boldsymbol{\pi}$ stemming from the Steady-State Theorem.

2. The long-run proportion of time the Markov chain visits the j th state is π_j .

To be more precise, for any state j let $N_j(n)$ denote the number of times the chain visits state j in its first n transitions; that is,

$$N_j(n) = \#\{1 \leq k \leq n : X_k = j\}$$

Then it can be shown that $N_j(n)/n$, the *proportion* of time the Markov chain spends in state j among the first n transitions, converges in probability to π_j .

3. If we assign $\boldsymbol{\pi}$ to be the initial distribution of X_0 , then the distribution of X_n is also $\boldsymbol{\pi}$ for any subsequent number of transitions n . For this reason, $\boldsymbol{\pi}$ is customarily referred to as the **stationary distribution** of the Markov chain.

To prove Statement 3, first let $\boldsymbol{\Pi}$ denote the matrix in Eq. (6.4), each of whose rows is $\boldsymbol{\pi}$. Now write $\mathbf{P}^{n+1} = \mathbf{P}^n \mathbf{P}$ and take the limit of both sides as $n \rightarrow \infty$:

$$\mathbf{P}^{n+1} = \mathbf{P}^n \mathbf{P} \Rightarrow \lim_{n \rightarrow \infty} \mathbf{P}^{n+1} = \lim_{n \rightarrow \infty} [\mathbf{P}^n \mathbf{P}] = \left[\lim_{n \rightarrow \infty} \mathbf{P}^n \right] \mathbf{P} \Rightarrow \boldsymbol{\Pi} = \boldsymbol{\Pi} \mathbf{P}$$

Each side of the last equation is an $s \times s$ matrix; equating the top rows of these two matrices, we have $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}$. (You could just as well equate any other row, since all the rows of $\boldsymbol{\Pi}$ are the same.)

Now, assign the steady-state distribution to X_0 : $\mathbf{v}_0 = \boldsymbol{\pi}$. Then the (unconditional) distribution of X_1 , using the results of Sect. 6.3, is $\mathbf{v}_1 = \boldsymbol{\pi} \mathbf{P}$, which we have established equals $\boldsymbol{\pi}$. Continuing by induction, we have for any n that the unconditional distribution of X_n is $\mathbf{v}_n = \mathbf{v}_{n-1} \mathbf{P} = \boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$, completing the proof.

4. The expected number of transitions required to return to the j th state, beginning in the j th state, is equal to $1/\pi_j$. This is called the **mean recurrence time** for state j .

Compare this result to the mean of a geometric rv from Chap. 2: the expected number of trials (replications) required to first observe an event whose probability is p equals $1/p$. The difference is that the geometric model assumes the trials are independent, while a Markov chain model assumes that successive states of the chain are dependent (as specified by the Markov property). But if we think of “return to the j th state” as our event of interest, then Statement 1 implies that (at least for large n) the probability of this event is roughly π_j , and so it seems reasonable that the average number of tries/steps it will take to achieve this event will be $1/\pi_j$.

Example 6.21 The steady-state distribution for the taxi driver example is given by the 1×3 vector $\boldsymbol{\pi} = [.2 \ .6 \ .2]$. For now, this relies on the computation of \mathbf{P}^{100} above; shortly, we will present a derivation of this vector that does not require raising \mathbf{P} to a high power. From the preceding descriptions, we conclude all of the following:

1. Regardless of where the taxi driver starts his day, for large n there is about a 20% chance his n th fare will be dropped off in zone 1, a 60% chance that that fare will go to zone 2, and a 20% chance for zone 3.
2. In the long run, the taxi driver drops off about 20% of his fares in zone 1, about 60% in zone 2, and about 20% in zone 3.
3. Suppose the taxi driver sleeps in his cab, thus waking up each day in a “random” zone, and we assign to X_0 (his point of origin tomorrow, say) the initial distribution $\mathbf{v}_0 = \boldsymbol{\pi} = [.2 \ .6 \ .2]$. The unconditional distribution of X_1 , the destination of tomorrow’s first fare, is

$$\mathbf{v}_1 = \mathbf{v}_0 \mathbf{P} = [.2 \ .6 \ .2] \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix}$$

By direct computation, the first entry of \mathbf{v}_1 is $(.2)(.3) + (.6)(.1) + (.2)(.4) = .2$; the second entry is $(.2)(.2) + (.6)(.8) + (.2)(.4) = .6$; and the last is $.2$. That is, $\mathbf{v}_1 = [.2 \ .6 \ .2] = \boldsymbol{\pi}$, and so X_1 has the same distribution as X_0 . The same will hold for X_2, X_3 , and so on.

4. If the driver starts from his home in zone 3, then on the average the number of fares he handles until he is brought back to zone 3 is given by $1/\pi_3 = 1/(.2) = 5$. That is, the mean recurrence time for state 3 (zone 3) is five transitions. ■

6.4.4 Efficient Computation of Steady-State Probabilities

The preceding examples of regular Markov chains and the resulting steady-state distributions may suggest that one determines $\boldsymbol{\pi}$ by computing a high power of the transition matrix \mathbf{P} , preferably with software, and then extracting any row of the resulting matrix (all of which will be the same, according to the Steady-State Theorem). Fortunately there is a more direct technique for determining $\boldsymbol{\pi}$. The method was hinted at in the proof of Statement 3 above: the steady-state distribution $\boldsymbol{\pi}$ satisfies the matrix equation $\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$. In fact, something stronger is true.

THEOREM

Let \mathbf{P} be the one-step transition matrix of a regular Markov chain on the state space $\{1, \dots, s\}$. The steady-state distribution of the Markov chain is the *unique* solution $\boldsymbol{\pi} = [\pi_1 \cdots \pi_s]$ to the system of equations formed by

$$\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi} \quad \text{and} \quad \pi_1 + \cdots + \pi_s = 1 \tag{6.5}$$

Proof Statement 3 above and the fact that $\boldsymbol{\pi}$ is a probability vector (because it’s the limit of probability vectors) ensures that $\boldsymbol{\pi}$ itself satisfies Eq. (6.5). We must show that any *other* vector satisfying both equations in Eq. (6.5) is, in fact, $\boldsymbol{\pi}$. To that end, let \mathbf{w} be any $1 \times s$ vector satisfying the two conditions $\mathbf{w} \mathbf{P} = \mathbf{w}$ and $\sum w_i = 1$. Similar to earlier derivations, we have $\mathbf{w} \mathbf{P}^2 = (\mathbf{w} \mathbf{P}) \mathbf{P} = \mathbf{w} \mathbf{P} = \mathbf{w}$ and, by induction, $\mathbf{w} \mathbf{P}^n = \mathbf{w}$ for any positive integer n . Taking the limit of both sides as $n \rightarrow \infty$, the Steady-State Theorem implies that $\mathbf{w} \boldsymbol{\Pi} = \mathbf{w}$.

Now expand $\mathbf{w}\mathbf{\Pi}$:

$$\begin{aligned}\mathbf{w}\mathbf{\Pi} &= [w_1 \cdots w_s] \begin{bmatrix} \pi_1 & \cdots & \pi_s \\ \vdots & \vdots & \vdots \\ \pi_1 & \cdots & \pi_s \end{bmatrix} = [(w_1\pi_1 + \cdots + w_s\pi_1) \quad \cdots \quad (w_1\pi_s + \cdots + w_s\pi_s)] \\ &= [(\sum w_i)\pi_1 \quad \cdots \quad (\sum w_i)\pi_s] = (\sum w_i)[\pi_1 \cdots \pi_s] = (\sum w_i)\boldsymbol{\pi}\end{aligned}$$

Since $\sum w_i = 1$ by assumption, we have $\mathbf{w}\mathbf{\Pi} = \boldsymbol{\pi}$. It was established above that $\mathbf{w}\mathbf{\Pi} = \mathbf{w}$, and so we conclude that $\mathbf{w} = \boldsymbol{\pi}$, as originally claimed. ■

Example 6.22 Consider again the Markov chain model for snowy days (S) and non-snowy or “green” days (G) in New York City, begun in Example 6.6. The one-step transition matrix was given by

$$\mathbf{P} = \begin{matrix} G \\ S \end{matrix} \begin{bmatrix} .964 & .036 \\ .224 & .776 \end{bmatrix}$$

Since all the entries of \mathbf{P} are positive, this is a regular Markov chain. The preceding theorem can be used to determine the steady-state probabilities $\boldsymbol{\pi} = [\pi_1 \ \pi_2]$. The equations in Eq. (6.5), written out long-hand, are

$$\begin{aligned}.964\pi_1 + .224\pi_2 &= \pi_1 \\ .036\pi_1 + .776\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1\end{aligned}$$

Substituting $\pi_2 = 1 - \pi_1$ into the first equation gives $.964\pi_1 + .224(1 - \pi_1) = \pi_1$; solving for π_1 produces $\pi_1 = .224/.260 = .8615$ and then $\pi_2 = 1 - .8615 = .1385$. For the season to which this model applies, in the long run New York City has at least 50 mm of snow on 86.15% of days and less than 50 mm on the other 13.85% of days.

It’s important to note that the top two equations alone, i.e., those provided by the relationship $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, do not uniquely determine the value of the vector $\boldsymbol{\pi}$. The first equation is equivalent to $.224\pi_2 = .036\pi_1$ (subtract $.964\pi_1$ from both sides), but so is the second equation (subtract $.776\pi_2$ from both sides). The final equation, requiring the entries of $\boldsymbol{\pi}$ to sum to 1, is necessary to obtain a unique solution. ■

Expression (6.5) may be reexpressed as a single matrix equation. Taking a transpose,

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi} \Rightarrow \mathbf{P}^T\boldsymbol{\pi}^T = \boldsymbol{\pi}^T = \mathbf{I}\boldsymbol{\pi}^T \Rightarrow (\mathbf{P}^T - \mathbf{I})\boldsymbol{\pi}^T = \mathbf{0},$$

where $\mathbf{0}$ is an $s \times 1$ vector of zeros. The requirement $\pi_1 + \cdots + \pi_s = 1$ can be rendered in matrix form as $[1 \cdots 1]\boldsymbol{\pi}^T = [1]$, and so the system of Eq. (6.5) can be expressed with the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & \cdots & 1 & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{array} \right] \quad (6.6)$$

Example 6.23 (Example 6.21 continued) To analytically determine the steady-state distribution of our taxi driver example, first construct the matrix $\mathbf{P}^T - \mathbf{I}$:

It should be clear that every regular chain is irreducible (do you see why?). However, the reverse is not true: an irreducible Markov chain need not be a regular chain. Consider the cyclic chain of Example 6.20: the bus can access any of the three locations it visits (campus, housing, downtown) from any other location, so the chain is irreducible. However, as discussed earlier in this section, the chain is definitely not regular. The Ehrenfest chain model developed in Exercise 2 is another example of an irreducible but not regular chain; see Exercise 43 at the end of this section.

It can be shown that *any finite-state, irreducible Markov chain has a stationary distribution*. That is, if \mathbf{P} is the transition matrix of an irreducible chain, there exists a row vector $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$; moreover, there is a unique such vector satisfying the additional constraint $\sum \pi_i = 1$. For example, the cyclic bus route chain of Example 6.20 has stationary distribution $\boldsymbol{\pi} = [1/3 \ 1/3 \ 1/3]$, as seen by the computation

$$\boldsymbol{\pi}\mathbf{P} = [1/3 \ 1/3 \ 1/3] \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = [1/3 \ 1/3 \ 1/3] = \boldsymbol{\pi}$$

So, if the bus is equally likely to be at any of its three locations right now, it is also equally likely to be at any of those three places after the next transition (the “stationary” interpretation of $\boldsymbol{\pi}$). This is true even though the chain is not regular, so the Steady-State Theorem does not apply.

If an s -state Markov chain is irreducible but not regular, then every state can access every other state but there exists no integer n for which all s^2 probabilities $P^{(n)}(i \rightarrow j)$ are positive. The only way this can occur is if the chain exhibits some sort of “periodic” behavior, e.g., when one group of states can access some states only in an even number of steps and others only in an odd number of steps. Formally, the **period** of a state i is defined as the greatest common divisor of all positive integers n such that $P^{(n)}(i \rightarrow i) > 0$; if that gcd equals 1, then state i is called *aperiodic*. All three states in the cyclic chain above have period 3, because for every state the period is $\gcd(3, 6, 9, \dots) = 3$. It can be shown that every state in an irreducible chain has the same period; the chain is called aperiodic if that common period is 1 and is called periodic otherwise.

As noted previously, for any regular Markov chain there exists an integer n such that all the entries of \mathbf{P}^n , \mathbf{P}^{n+1} , \mathbf{P}^{n+2} , and so on are positive. Since the gcd of the set $\{n, n+1, n+2, \dots\}$ is 1, it immediately follows that every regular Markov chain is aperiodic. The following theorem characterizes regularity for finite-state chains.

THEOREM

A finite-state Markov chain is regular if, and only if, it is both irreducible and aperiodic.

The “only if” direction of the theorem is established in the earlier paragraphs of this sub-section. The converse statement, that all irreducible and aperiodic finite-state chains are regular, can be proved using a result called the *Frobenius coin-exchange theorem* (we will not present the proof here).

6.4.6 Exercises: Section 6.4 (31–43)

31. Refer back to Mendel’s plant breeding experiments in Example 6.16 and Exercise 26.

- (a) Do the genotypes formed by successive cross-breeding with pure recessive plants gg , as in Example 6.16, form a regular Markov chain?
- (b) Do the genotypes formed by successive cross-breeding with hybrid plants Gg , as in Exercise 26, form a regular Markov chain?
32. Refer back to Exercise 2. Assume $m = 3$ balls are being exchanged between the two chambers. Is the Markov chain $X_n =$ number of balls in the left chamber a regular chain?
33. Refer back to Example 6.13 regarding cell phone contracts in China.
- (a) Determine the steady-state probabilities of this chain.
- (b) In the long run, what proportion of Chinese cell phone users will have contracts with China Mobile?
- (c) A certain cell phone customer currently has a contract with China Telecom. On the average, how many contract changes will s/he make before signing with China Telecom again?
34. The article “Markov Chain Model for Performance Analysis of Transmitter Power Control in Wireless MAC Protocol” (Twenty-first International Conference on Advanced Networking and Applications, 2007) describes a Markov chain model for the state of a communication channel using a particular “slotted non-persistent” (SNP) protocol. The channel’s possible states are (1) idle, (2) successful transmission, and (3) collision. For particular values of the authors’ proposed four-parameter model, the following transition matrix results:

$$\mathbf{P} = \begin{bmatrix} .50 & .40 & .10 \\ .02 & .98 & 0 \\ .12 & 0 & .88 \end{bmatrix}$$

- (a) Verify that \mathbf{P} is the transition matrix of a regular Markov chain.
- (b) Determine the steady-state probabilities for this channel.
- (c) What proportion of time is this channel idle, in the long run?
- (d) What is the average number of time steps between successive collisions?
35. Refer back to Exercise 3.
- (a) Construct the one-step transition matrix \mathbf{P} of this chain.
- (b) Show that $X_n =$ the machine’s state (full, part, broken) on the n th day is a regular Markov chain.
- (c) Determine the steady-state probabilities for this chain.
- (d) On what proportion of days is the machine fully operational?
- (e) What is the average number of days between breakdowns?
36. Refer back to Exercise 6, and assume three files A, B, C are to be repeatedly requested. Suppose that 60% of requests are for file A , 10% for file B , and 30% for C . Let $X_n =$ the stacked order of the files (e.g., ABC) after the n th request.
- (a) Construct the transition matrix \mathbf{P} for this chain. (The one-step transition probabilities were established in Exercise 6(c).)
- (b) Determine the steady-state probability for the stack ABC .
- (c) Show that, in general, the steady-state probability for ABC is given by

$$\pi_{ABC} = \frac{p_A \cdot p_B}{p_B + p_C}$$

where $p_A = P(\text{file } A \text{ is requested})$ and p_B and p_C are defined similarly. (The other five steady-state probabilities can be deduced by changing the subscripts appropriately.)

37. Let \mathbf{P} be the one-step transition matrix of a Markov chain. Show that if all the entries of \mathbf{P}^n are positive for some positive integer n , then so are all the entries of \mathbf{P}^{n+1} , \mathbf{P}^{n+2} , and so on. [Hint: Write $\mathbf{P}^{n+1} = \mathbf{P} \cdot \mathbf{P}^n$ and consider how the (i, j) th entry of \mathbf{P}^{n+1} is obtained.]
38. Refer back to Exercise 19.
- Consider a new customer. By definition, s/he made no visits to the repair shop last year. What is his/her expected number of visits this year?
 - Now suppose a car owner has been a customer of this repair shop for many years. What is the expected number of shop visits s/he will make next year?
39. Consider a Markov chain with just two states, 0 and 1, with one-step transition probabilities $\alpha = P(0 \rightarrow 1)$ and $\beta = P(1 \rightarrow 0)$.
- Assuming $0 < \alpha < 1$ and $0 < \beta < 1$. Determine the steady-state probabilities of states 0 and 1 in terms of α and β .
 - What happens if α and/or β equals 0 or 1?
40. *Occupational prestige* describes how particular jobs are regarded by society and is often used by sociologists to study class. The article “Social Mobility in the United States as a Markov Process” (*J. for Economic Educators*, v. 8 no. 1 (2008): 15-37) investigates the occupational prestige of fathers and sons. Data provided in the article can be used to derive the following transition matrix for occupational prestige classified as low (L), medium (M), or high (H):

$$\mathbf{P} = \begin{matrix} & \begin{matrix} \text{L} & \text{M} & \text{H} \end{matrix} \\ \begin{matrix} \text{L} \\ \text{M} \\ \text{H} \end{matrix} & \begin{bmatrix} .5288 & .2096 & .2616 \\ .3688 & .2530 & .3782 \\ .2312 & .1738 & .5950 \end{bmatrix} \end{matrix}$$

- Which occupational prestige “state” is the most likely to self-replicate (i.e., father and son are in the same category)? Which is the least likely?
 - Determine the steady-state distribution of this Markov chain.
 - Interpret the distribution in (b), assuming the model specified by the matrix is valid across many generations.
- [Note: The authors actually used 11 categories of occupational prestige; we have collapsed these into three categories for simplicity.]
41. The two ends of a wireless communication system can each be inactive (0) or active (1). Suppose the two nodes act independently, each as a Markov chain with the transition probabilities specified in Exercise 39. Let X_n = the “combined” state of the two relays at the n th time step. The state space for this chain is $\{00, 01, 10, 11\}$, e.g., state 01 corresponds to an inactive transmitter with an active receiver. (Performance analysis of such systems is described in “Energy-Efficient Markov Chain-Based Duty Cycling Schemes for Greener Wireless Sensor Networks,” *ACM J. on Emerging Tech. in Computing Systems* (2012):1-32.)
- Determine the transition matrix for this chain. [Hint: Use independence to uncouple the two states, e.g., $P(00 \rightarrow 10) = P(0 \rightarrow 1) \cdot P(0 \rightarrow 0)$.]
 - Determine the steady state distribution of this chain.
 - As the authors note, “a connection is feasible only when both wireless nodes are active.” What proportion of time is a connection feasible under this model?
42. A particular gene has three expressions: AA , Aa , and aa . When two individuals mate, one half of each parent’s gene is contributed to the offspring (and each half is equally likely to be donated). For example, an AA mother can only donate A while an Aa father is equally likely to donate A or a , resulting in a child that is either AA or Aa . Suppose that the population proportions of AA , Aa , and aa individuals are p , q , and r , respectively (so $p + q + r = 1$). Consider the offspring of a

randomly selected individual; specifically, let X_n = the gene expression of the oldest child in his or her n th generation of descendants (whom we assume will have at least one offspring).

- (a) Assume the n th-generation individual's mate is selected at random from the genetic population described above. Show the following: $P(X_{n+1} = AA|X_n = AA) = p + q/2$, $P(X_{n+1} = Aa|X_n = AA) = q/2 + r$, and $P(X_{n+1} = aa|X_n = AA) = 0$. [Hint: Apply the Law of Total Probability.]
- (b) Using the same method as in (a), determine the other one-step transition probabilities and construct the transition matrix \mathbf{P} of this chain.
- (c) Verify that X_n is a regular Markov chain.
- (d) Suppose there exists some $\alpha \in [0, 1]$ such that $p = \alpha^2$, $q = 2\alpha(1 - \alpha)$, and $r = (1 - \alpha)^2$. (In this context, $\alpha = P(A \text{ allele})$.) Show that $\boldsymbol{\pi} = [p \ q \ r]$ is the stationary distribution of this chain. (This fact is called the *Hardy-Weinberg law*; it establishes that the rules of genetic recombination result in a long-run stable distribution of genotypes.)
43. Refer back to the Ehrenfest chain model of Exercises 2 and 24. Once again assume that $m = 3$ balls are being exchanged between the two chambers.
- (a) Explain why this is an irreducible chain, but not a regular chain.
- (b) Explain why each state has period equal to 2.
- (c) Show that the vector $[1/8 \ 3/8 \ 3/8 \ 1/8]$ is a stationary distribution for this chain. (Thus, even though the chain is not regular and the transition matrix \mathbf{P} does not have a limit, there still exists a stationary distribution due to irreducibility.)

6.5 Markov Chains with Absorbing States

The Gambler's Ruin scenario, begun in Example 6.2, has the feature that the chain "terminates" when it reaches either of two states (\$0 or \$3 in our version of the competition). As we've noted previously, it's mathematically advantageous to imagine that the Markov chain actually continues in these cases, just never leaving the state 0 or 3; one such sample path is

$$2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \rightarrow 0 \rightarrow 0 \rightarrow 0 \rightarrow \dots$$

In this section, we first investigate states from which a Markov chain can never exit and the time it takes to arrive in one of those states.

DEFINITION

A state j of a Markov chain is called an **absorbing state** if

$$P(j \rightarrow j) = 1.$$

Equivalently, j is an absorbing state if the (j, j) th entry of the one-step transition matrix of the chain is 1.

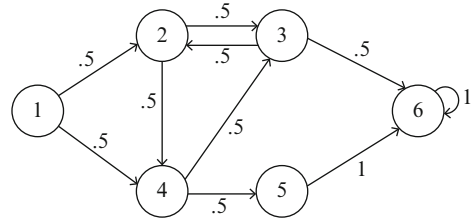
The states 0 and 3 are both absorbing states in our Gambler's Ruin example. In contrast, the taxi driver example has no absorbing states. The next example shows that some care must be taken in identifying absorbing states.

Example 6.25 Anyone who has applied for a bank loan knows that the process of eventual approval (or rejection) involves many steps and, occasionally, a lot of complex negotiation. Figure 6.7 illustrates the possible route of a set of loan documents from (1) document initiation to (6) final approval or rejection. The intermediate steps (2)–(5) represent various exchanges between underwriters, loan officers, and the like. In this particular chain, two such individuals (at states 3 and 5) have the authority to make a final decision, though the agent at state 3 may elect to return the documents for further discussion.

The one-step transition matrix of this chain is

$$\mathbf{P} = \begin{bmatrix} 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & .5 & 0 & 0 & 0 & .5 \\ 0 & 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Fig. 6.7 State diagram for Example 6.25



Although the number 1 appears twice in \mathbf{P} , only state 6 is an absorbing state of this chain. Indeed, $p_{66} = P(6 \rightarrow 6) = 1$; however, state 5 is *not* an absorbing state because $p_{55} = P(5 \rightarrow 5) = 0$. Rather, the fifth row of \mathbf{P} indicates that if the chain ever enters state 5, it will necessarily pass in the next transition into state 6 (where, as it happens, it will be “absorbed”). ■

To be clear, a Markov chain may have no absorbing states (the taxi driver), a single absorbing state (Example 6.25), or multiple absorbing states (Gambler’s Ruin).

6.5.1 Time to Absorption

When a Markov chain has one or more absorbing states, it is natural to ask how long it will take to reach an absorbing state. Of course, the answer depends on where (i.e., in which state) the Markov chain begins. For any *non-absorbing* state i , define a random variable T_i by

T_i = number of transitions until the Markov chain reaches an absorbing state, starting in state i

This rv T_i is called the **time to absorption from state i** ; the possible values of T_i are 1, 2, 3, 4, As we shall now illustrate, the distribution of T_i can be approximated from the k -step transition matrices \mathbf{P}^k for $k = 1, 2, 3, \dots$. For simplicity’s sake, consider first a Markov chain with a single absorbing state, which we will call a . Then the (i, a) th entry of \mathbf{P} is the probability of transitioning directly from state i into the absorbing state a , which is therefore also the probability that T_i equals 1:

$$P(i \rightarrow a) = P(T_i = 1)$$

Since T_i is always a positive integer, this also equals $P(T_i \leq 1)$, a fact which will prove important shortly. Now consider the (i, a) th entry of \mathbf{P}^2 , which represents $P^{(2)}(i \rightarrow a)$. There are two ways the Markov chain could transition from i to a in two steps:

$$i \rightarrow \text{any non-absorbing state} \rightarrow a \quad (T_i = 2), \quad \text{or}$$

$$i \rightarrow a \rightarrow a \quad (T_i = 1)$$

Therefore, the two-step probability $P^{(2)}(i \rightarrow a)$ does not represent the chance T_i equals 2, but rather the chance that T_i is *at most* 2. That is,

$$P^{(2)}(i \rightarrow a) = P(T_i \leq 2).$$

Following the same pattern, the k -step transition probability $P^{(k)}(i \rightarrow a)$ is equal to $P(T_i \leq k)$ for any positive integer k .

If the Markov chain has two absorbing states a_1 and a_2 , say, then the chance of being absorbed from state i in one step is simply the sum $P(i \rightarrow a_1) + P(i \rightarrow a_2)$, since those two events are mutually exclusive (you can only arrive in one state). Similarly, the probability $P(T_i \leq 2)$ is determined by adding $P^{(2)}(i \rightarrow a_1) + P^{(2)}(i \rightarrow a_2)$, and so on. The general result is stated in the following theorem.

THEOREM

Consider a finite-state Markov chain, and let A denote the (non-empty) set of absorbing states. For any state $i \notin A$, define $T_i =$ the number of transitions, starting in state i , until the chain arrives in some absorbing state. Then the cdf of T_i is given by

$$F_{T_i}(k) = P(T_i \leq k) = \sum_{a \in A} P^{(k)}(i \rightarrow a) \quad k = 1, 2, 3, \dots$$

In the special case of a single absorbing state, a , this simplifies to

$$F_{T_i}(k) = P(T_i \leq k) = P^{(k)}(i \rightarrow a)$$

The probability distribution of T_i (i.e., the pmf of the rv T_i) can then be determined from the cdf.

Example 6.26 (Example 6.25 continued) Let's consider the rv T_1 , the absorption time from state 1 (i.e., the number of steps from loan document initiation to the bank's final decision). From the one-step transition matrix \mathbf{P} , we know that

$$F_{T_1}(1) = P(T_1 \leq 1) = P(T_1 = 1) = P(1 \rightarrow 6) = p_{16} = 0.$$

The $(1,6)$ entry of \mathbf{P}^2 is also zero, so $F_{T_1}(2) = P(T_1 \leq 2) = P^{(2)}(1 \rightarrow 6) = 0$. Software was used to obtain the matrices $\mathbf{P}^3, \dots, \mathbf{P}^{12}$, resulting in the following values for the $(1,6)$ entry.

k	1	2	3	4	5	6	7	8	9	10	11	12
$F_{T_1}(k)$	0	0	.5	.6875	.75	.8594	.8984	.9336	.9570	.9707	.9810	.9873

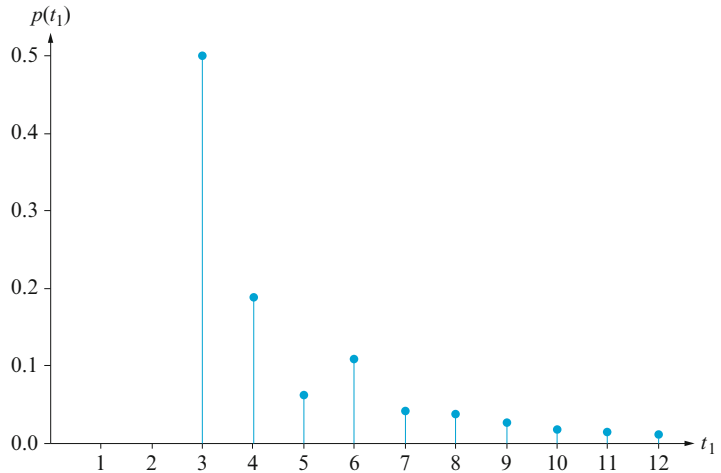
The accompanying table is, of course, an incomplete description of the cdf of T_1 , since this process could theoretically be continued indefinitely. Next, because the rv T_1 is integer-valued, its pmf is easily determined from the cdf:

$$P(T_1 = 2) = P(T_1 \leq 2) - P(T_1 \leq 1) = F_{T_1}(2) - F_{T_1}(1) = 0 - 0 = 0$$

$$P(T_1 = 3) = P(T_1 \leq 3) - P(T_1 \leq 2) = F_{T_1}(3) - F_{T_1}(2) = .5 - 0 = .5$$

$$P(T_1 = 4) = P(T_1 \leq 4) - P(T_1 \leq 3) = F_{T_1}(4) - F_{T_1}(3) = .6875 - .5 = .1875$$

Fig. 6.8 The (incomplete) pmf of T_1 from Example 6.26



The first 12 probabilities in the pmf of T_1 are as follows (their sum is .9873):

k	1	2	3	4	5	6	7	8	9	10	11	12
$p_{T_1}(k)$	0	0	.5	.1875	.0625	.1094	.0390	.0352	.0234	.0137	.0103	.0063

This incomplete pmf is graphed in Fig. 6.8. Notice that T_1 must be at least 3, which is consistent with the state diagram in Fig. 6.7: it takes at least three steps to get from state 1 to state 6 (one of $1 \rightarrow 2 \rightarrow 3 \rightarrow 6$, $1 \rightarrow 4 \rightarrow 3 \rightarrow 6$, or $1 \rightarrow 4 \rightarrow 5 \rightarrow 6$).

A call to the bank determines that the documents are in the hands of the underwriter indicated by state 4. So, let’s now consider the rv $T_4 =$ time to absorption (completion of the process) starting from state 4. Based on the state diagram, it seems reasonable to anticipate that it will typically take less time to reach state 6 starting from state 4 than it did when the chain began in state 1. Reading off the (4, 6) entries of $\mathbf{P}, \mathbf{P}^2, \dots, \mathbf{P}^{12}$ yields the cdf values in the accompanying table; subtraction as before then gives the corresponding pmf values.

k	1	2	3	4	5	6	7	8	9	10	11	12
$F_{T_4}(k)$	0	.75	.75	.8125	.9063	.9219	.9531	.9688	.9785	.9863	.9907	.9939
$p_{T_4}(k)$	0	.75	0	.0625	.0938	.0156	.0312	.0157	.0097	.0078	.0044	.0032

Notice that, starting in state 4, the chain is quite likely to be absorbed into state 6 in exactly two steps (either $4 \rightarrow 5 \rightarrow 6$ or $4 \rightarrow 3 \rightarrow 6$, with probabilities .5 and .25, respectively), and that it is impossible to move from 4 to 6 in exactly three steps. ■

Example 6.27 In the Gambler’s Ruin scenario with $p = .55$, how many games will Allan and Beth play against each other before one player goes broke? Recall that the transition matrix \mathbf{P} is set from Allan’s perspective, and that he begins with \$2. Thus, the rv of interest is T_2 , the number of transitions (aka games), starting from Allan having \$2, until the competition ends because Allan either has \$0 or \$3. The one- and two-step transition matrices of this chain appear in Example 6.14. Hence

$$P(T_2 \leq 1) = P(2 \rightarrow 0) + P(2 \rightarrow 3) = 0 + .55 = .55$$

$$P(T_2 \leq 2) = P^{(2)}(2 \rightarrow 0) + P^{(2)}(2 \rightarrow 3) = .2025 + .55 = .7525$$

In general, the cumulative probability $P(T_2 \leq k)$ can be determined by adding the (2,0) and (2,3) entries of the k -step transition matrix \mathbf{P}^k . These values were determined with the aid of software for $k = 1$ through 10 and are summarized in the accompanying table.

k	1	2	3	4	5	6	7	8	9	10
$F_{T_2}(k)$.55	.7525	.8886	.9387	.9724	.9848	.9932	.9962	.9983	.9991
$p_{T_2}(k)$.55	.2025	.1361	.0501	.0337	.0124	.0084	.0030	.0021	.0008

It’s important to notice that T_2 indicates the number of steps required to enter *some* absorbing state (here, either \$0 or \$3), not the number of steps to enter a particular such state. ■

6.5.2 Mean Time to Absorption

With $T_i =$ time to absorption starting from state i , the expected value of T_i is called the **mean time to absorption (mtta) from state i** :

$\mu_i = E(T_i) =$ expected number of transitions until the Markov chain reaches an absorbing state, starting in state i

For each of the preceding examples, the incomplete pmf can be used to approximate the mttA from state i . Consider the Markov chain in Example 6.26:

$$\begin{aligned} \mu_1 &= E(T_1) = \sum_{k=1}^{\infty} k \cdot p_{T_1}(k) \approx \sum_{k=1}^{12} k \cdot p_{T_1}(k) \\ &= 1(0) + 2(0) + 3(.5) + 4(.1875) + \dots + 11(.0103) + 12(.0063) \\ &= 4.31 \end{aligned}$$

To a hopefully reasonable approximation, on average the chain requires 4.31 transitions, starting in state 1, to be absorbed into state 6. Similarly, the mean time to absorption from state 4 is approximately

$$\begin{aligned} \mu_4 &= \sum_{k=1}^{\infty} k \cdot p_{T_4}(k) \approx \sum_{k=1}^{12} k \cdot p_{T_4}(k) \\ &= 1(0) + 2(.75) + \dots + 11(.0044) + 12(.0032) = 2.91 \end{aligned}$$

For the Gambler’s Ruin competition with $p = .55$ and Allan’s initial stake at \$2, the pmf displayed in Example 6.27 gives

$$\mu_2 \approx 1(.55) + 2(.2025) + \cdots + 10(.0008) = 1.92$$

That is, if Allan starts with \$2 and $p = .55$, the expected length of the Gambler's Ruin competition is approximated to be 1.92 games.

In all such approximations, two things should be clear. First, the estimated means are *smaller than* the correct values, since the sums used are truncated versions of the correct summations and every term is nonnegative. So, in the Gambler's Ruin scenario, $\mu_2 > 1.92$. Second, the more terms we include in the truncated sum, the closer the approximation will be to the correct mean time to absorption from that state. Of course, additional terms require overcoming the practical hurdle of computing successively higher powers of the matrix \mathbf{P} . With software, one could in practice use this method to get a very good approximation to the mta.

Exercise 56 presents a different approximation method that always yields a better approximation to the mean time to absorption; moreover, it relies directly on the cdf values and thus does not require computing differences to form the pmf. But this is still an approximation; what we would really like is an explicit method for determining the exact mean time to absorption from various states in the chain. The following theorem provides such a result.

MTTA THEOREM

Suppose a finite-state Markov chain with one-step transition matrix \mathbf{P} has r non-absorbing states (and at least one absorbing state). Suppose further that there exists a path from every non-absorbing state into some absorbing state.

Let \mathbf{Q} be the $r \times r$ sub-matrix of \mathbf{P} corresponding to the non-absorbing states of the chain. Then the mean times to absorption from these states are given by the matrix formula

$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{1},$$

where $\mu_i =$ mta from the i th state in the \mathbf{Q} sub-matrix, $\boldsymbol{\mu} = [\mu_1 \cdots \mu_r]^T$, \mathbf{I} is the $r \times r$ identity matrix, and $\mathbf{1} = [1 \cdots 1]^T$.

This theorem not only provides the exact mean times to absorption (as opposed to the earlier approximations) but also computes *all* of them simultaneously. A proof of the MTTA Theorem will be presented shortly, but first we illustrate its use with our two ongoing examples.

Example 6.28 For the bank loan Markov chain in Example 6.25, state 6 is the only absorbing state, so there are $r = 5$ non-absorbing states. The sub-matrix corresponding to these non-absorbing states is

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .5 & .5 & 0 \\ 0 & .5 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

This can be obtained by “crossing out” the row and column of \mathbf{P} corresponding to absorbing state 6. Let $\mu_i = E(T_i)$ be the mean time to absorption from state i for $i = 1, 2, 3, 4, 5$. Then, according to the MTTA Theorem,

$$\begin{aligned}
 \boldsymbol{\mu} &= (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{1} \\
 &= \begin{bmatrix} 1 & -.5 & 0 & -.5 & 0 \\ 0 & 1 & -.5 & -.5 & 0 \\ 0 & -.5 & 1 & 0 & 0 \\ 0 & 0 & -.5 & 1 & -.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 1 & 1 & 1 & .5 \\ 0 & 1.6 & 1.2 & .8 & .4 \\ 0 & .8 & 1.6 & .4 & .2 \\ 0 & .4 & .8 & 1.2 & .6 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 4.0 \\ 3.0 \\ 3.0 \\ 1.0 \end{bmatrix}
 \end{aligned}$$

The inverse of $\mathbf{I} - \mathbf{Q}$ was determined using software.

So, for example, the mean time to absorption from state 1 is $\mu_1 = 4.5$ transitions, slightly larger than our earlier approximation of 4.31. On the average, it takes 4.5 steps to arrive at a loan decision starting from the time the loan documents are initiated. Similarly, the earlier estimate $\mu_4 \approx 2.91$ was a little off from the correct answer of $\mu_4 = 3$. The last entry of the vector $\boldsymbol{\mu}$ is obvious from the design of the chain: since state 5 transitions immediately into state 6 with certainty, T_5 is identically equal to 1, and so its mean is 1. ■

Example 6.29 Consider once again our Gambler's Ruin scenario, this time with an arbitrary probability p that Allan triumphs over Beth in any one game. The only two non-absorbing states are \$1 and \$2, so the required sub-matrix \mathbf{Q} consists of the "center four" entries of the original 4×4 transition matrix:

$$\mathbf{P} = \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{Q} = \frac{1}{2} \begin{bmatrix} 0 & p \\ 1-p & 0 \end{bmatrix}$$

There is a simple inverse formula for a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad (6.7)$$

Applying Eq. (6.7) and the MTTA Theorem,

$$\begin{aligned}
 \boldsymbol{\mu} &= (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{1} \\
 &= \begin{bmatrix} 1 & -p \\ -1+p & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{(1)(1) - (-p)(-1+p)} \begin{bmatrix} 1 & p \\ 1-p & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
 &= \frac{1}{1-p+p^2} \begin{bmatrix} 1+p \\ 2-p \end{bmatrix}
 \end{aligned}$$

Hence

$$\mu_1 = \frac{1+p}{1-p+p^2} \quad \mu_2 = \frac{2-p}{1-p+p^2}$$

Since we have always started Allan with \$2, let's explore μ_2 further. If $p = 1$, so Allan cannot lose, then $\mu_2 = (2 - 1)/(1 - 1 + 1^2) = 1$. This is logical, since Allan would automatically transition from \$2 to \$3 in 1 step/game and the competition would be over. Similarly, substituting $p = 0$ into this expression gives $\mu_2 = 2$, reflecting the fact that if Allan cannot win games then the chain must necessarily proceed along the path $2 \rightarrow 1 \rightarrow 0$, a total of two transitions. For $p = .55$, the numerical case illustrated earlier, we have

$$\mu_2 = \frac{2 - .55}{1 - .55 + .55^2} = \frac{1.45}{.7525} = \frac{580}{301} = 1.92691$$

which is quite close to our previous approximation of 1.92.

For what value of p is the competition expected to take the longest? Using calculus, one can find the maximum of μ_2 with respect to p , which turns out to occur at $p = 2 - \sqrt{3} \approx .268$. If Allan begins with \$2 and has a .268 probability of winning each game, the expected duration of the competition is maximized, specifically with $\mu_2 = 1 + 2/\sqrt{3} \approx 2.155$ games. ■

Proof of the MTTA Theorem For notational ease, let $1, 2, \dots, r$ be the non-absorbing states of the chain. Also, let A denote the set of absorbing states (which, if the Markov chain has s total states, could be enumerated as $r + 1, \dots, s$). Starting in any non-absorbing state i , consider the first transition of the chain. If the chain transitions into any member of A , then it has been “absorbed” in one step and so $T_i = 1$. On the other hand, if the chain transitions into any *non*-absorbing state j (including back into i itself), then the expected number of steps to absorption is $1 + E(T_j)$, where the 1 accounts for the step just taken and T_j represents the time to absorption starting from the *new* state j . Apply the Law of Total Expectation:

$$\begin{aligned} E(T_i) &= 1 \cdot P(i \rightarrow A) + \sum_{j=1}^r (1 + E(T_j)) \cdot P(i \rightarrow j) \\ &= P(i \rightarrow A) + \sum_{j=1}^r P(i \rightarrow j) + \sum_{j=1}^r E(T_j) \cdot P(i \rightarrow j) \end{aligned}$$

Since the state space of the Markov chain is $A \cup \{1, 2, \dots, r\}$, the first two terms in the expression above must sum to 1. Thus, we have $\mu_i = 1 + \sum_{j=1}^r \mu_j P(i \rightarrow j)$ for $i = 1, 2, \dots, r$. Stacking these equations and rewriting slightly, we have

$$\begin{aligned} \mu_1 &= P(1 \rightarrow 1)\mu_1 + \dots + P(1 \rightarrow r)\mu_r + 1 \\ \mu_2 &= P(2 \rightarrow 1)\mu_1 + \dots + P(2 \rightarrow r)\mu_r + 1 \\ &\vdots \\ \mu_r &= P(r \rightarrow 1)\mu_1 + \dots + P(r \rightarrow r)\mu_r + 1 \end{aligned}$$

This stack can be written more compactly as $\boldsymbol{\mu} = \mathbf{Q}\boldsymbol{\mu} + \mathbf{1}$. Solving for $\boldsymbol{\mu}$ yields the desired result. ■

The MTTA Theorem requires that every non-absorbing state can reach (at least) one absorbing state. That is, the set of absorbing states must be *accessible* from every non-absorbing state. What would happen if this were not the case?

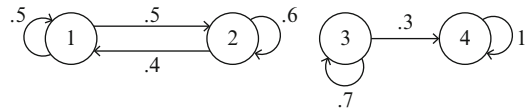
Example 6.30 In the Markov chain depicted in Fig. 6.9, 4 is an absorbing state, but it is only accessible from state 3. It is clear that the chain will eventually be absorbed into state 4 if $X_0 = 3$ and will never be absorbed into state 4 if $X_0 = 1$ or 2. So, where does the MTTA Theorem break down?

The one-step transition matrix \mathbf{P} for this chain, the resulting sub-matrix \mathbf{Q} for the non-absorbing states, and the matrix $\mathbf{I} - \mathbf{Q}$ required for calculating mean times to absorption are

$$\mathbf{P} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} \begin{bmatrix} .5 & .5 & 0 & 0 \\ .4 & .6 & 0 & 0 \\ 0 & 0 & .7 & .3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{Q} = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \begin{bmatrix} .5 & .5 & 0 \\ .4 & .6 & 0 \\ 0 & 0 & .7 \end{bmatrix} \quad \mathbf{I} - \mathbf{Q} = \begin{bmatrix} .5 & -.5 & 0 \\ -.4 & .4 & 0 \\ 0 & 0 & .3 \end{bmatrix}$$

The matrix $\mathbf{I} - \mathbf{Q}$ is not invertible; this can be seen by noting that the first and second rows are multiples of each other, or by computing the determinant and discovering that $\det(\mathbf{I} - \mathbf{Q}) = 0$. Because $(\mathbf{I} - \mathbf{Q})^{-1}$ does not exist, the formula from the MTTA Theorem cannot be applied.

Fig. 6.9 State diagram for Example 6.30



Recall that the cdf of T_1 can be determined from the appropriate entries of the k -step transition matrices; specifically, since the only absorbing state of this chain is state 4,

$$F_{T_1}(k) = P(T_1 \leq k) = P^{(k)}(1 \rightarrow 4) = \text{the } (1, 4) \text{ entry of } \mathbf{P}^k$$

The $(1, 4)$ entry of the matrix \mathbf{P} above is 0, so $P(T_1 \leq 1) = 0$. But since state 4 is not accessible from state 1, the $(1, 4)$ entry of every transition matrix \mathbf{P}^k is 0. Thus, $P(T_1 \leq k) = 0$ for all positive integers k and $p_{T_1}(k) = 0 - 0 = 0$ for all k . Since the probabilities associated with T_1 sum to zero and not 1, T_1 is not actually a valid rv (and so, in particular, has no mean). ■

In general, when the set of absorbing states is not accessible from every non-absorbing state, the matrix $\mathbf{I} - \mathbf{Q}$ will be singular (i.e., not invertible). If a subset of the non-absorbing states can access the absorbing states (that’s true for state 3 in Example 6.30), we can apply the MTTA Theorem if we define \mathbf{Q} to be the sub-matrix of \mathbf{P} corresponding to those states that can access the absorbing states.

6.5.3 Mean First Passage Times

We now briefly turn our attention back to regular Markov chains. In Sect. 6.4, we saw that one interpretation of the probability π_i from the Steady-State Theorem is that $1/\pi_i$ represents the expected number of transitions necessary for the chain to return to state i given that it starts there—the mean recurrence time for state i . With a clever use of the MTTA Theorem, we can also determine the expected number of transitions required for the chain to transition from a state i to a *different* state j —the **mean first passage time** from i to j .

Example 6.31 (Example 6.23 continued) For the ubiquitous taxi driver example, it was found that the steady-state probability for zone 3 is $\pi_3 = .2$ and, thus, the expected number of fares until the driver returns to zone 3 is $1/\pi_3 = 1/.2 = 5$.

But suppose the taxi driver just dropped off a fare in zone 1 (or zone 2). He wonders how long it will take him to get back home to zone 3 for lunch. More precisely, he wishes to know the expected number of fares required to reach zone 3, *starting from some other state* (i.e., different than zone 3).

The trick to answering the taxi driver's question—i.e., to determine the mean first passage time for zone 3 when beginning in zone 1 or zone 2—is to pretend that zone 3 is an absorbing state, and then invoke the MTTA Theorem. Modify the original one-step transition matrix \mathbf{P} of the Markov chain so that zone 3 is absorbing state, and label the new matrix $\tilde{\mathbf{P}}$:

$$\mathbf{P} = \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ .4 & .4 & .2 \end{bmatrix} \Rightarrow \tilde{\mathbf{P}} = \begin{bmatrix} .3 & .2 & .5 \\ .1 & .8 & .1 \\ 0 & 0 & 1 \end{bmatrix}$$

Now proceed as before: the sub-matrix for the non-absorbing states, which in $\tilde{\mathbf{P}}$ are zone 1 and zone 2, is

$$\mathbf{Q} = \begin{bmatrix} .3 & .2 \\ .1 & .8 \end{bmatrix},$$

from which

$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{1} = \begin{bmatrix} .7 & -.2 \\ -.1 & .2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \cdots = \begin{bmatrix} 3.33 \\ 6.67 \end{bmatrix}$$

Thus, starting in zone 1, the average number of trips required for the taxi driver to get home to zone 3 is 3.33, while it takes twice that long on the average if he's starting from zone 2. ■

6.5.4 Probabilities of Eventual Absorption

As discussed in Example 6.29 in the context of Gambler's Ruin, when a Markov chain has multiple absorbing states one can only speak of the mean time to absorption into the *set* of absorbing states, not any particular absorbing state (e.g., not time to \$0 separate from time to \$3). We can, however, ask about the *probability* of eventual absorption into state \$0, as opposed to eventual absorption into state \$3.

DEFINITION

Let a be an absorbing state of a Markov chain and let i be a non-absorbing state. The **probability of eventual absorption into a from state i** , denoted $\pi(i \rightarrow a)$, is defined by

$$\pi(i \rightarrow a) = \lim_{n \rightarrow \infty} P^{(n)}(i \rightarrow a)$$

That is, $\pi(i \rightarrow a)$ is defined to be the limit of the (i, a) entry of \mathbf{P}^n as $n \rightarrow \infty$. This is consistent with our previous efforts to determine the probability of eventual absorption by examining \mathbf{P}^{75} or \mathbf{P}^{100} . But rather than approximate these probabilities by taking a high power of \mathbf{P} , we now present an explicit method for determining them.

Before illustrating the method for determining $\pi(i \rightarrow a)$, a few observations are in order. First, if state a is not accessible from state i , then $P^{(n)}(i \rightarrow a) = 0$ for all n and the limit is also zero, i.e., $\pi(i \rightarrow a) = 0$ when i cannot access a . This occurred in Example 6.30, with state 4 not being accessible from states 1 or 2.

Second, if the Markov chain has a single absorbing state a , then $\pi(i \rightarrow a) = 1$ for every state i that can access a . That is, a chain with an accessible absorbing state will always eventually be absorbed. This would be the case, for instance, in Example 6.25: it is a sure thing that the chain will eventually arrive at (and stay in) state 6, irrespective of where the chain begins. So, the interesting cases of determining $\pi(i \rightarrow a)$ are for Markov chains with multiple absorbing states, such as Gambler's Ruin.

Third, suppose we extended the preceding definition to non-absorbing states. That is, what can be said about

$$\lim_{n \rightarrow \infty} P^{(n)}(i \rightarrow j)$$

when j is *not* an absorbing state? If the Markov chain has any absorbing states (and assuming at least one of these is accessible from i), then the chain will eventually get absorbed and so $P^{(n)}(i \rightarrow j) \rightarrow 0$. If we have a regular Markov chain—which, in particular, means there are no absorbing states—then the Steady State Theorem tells us $P^{(n)}(i \rightarrow j) \rightarrow \pi_j$, a steady-state probability that is independent of i . For other cases, such as the cyclic chain of Example 6.20, the limit of $P^{(n)}(i \rightarrow j)$ may not exist at all.

On to the calculation: as in the proof of the MTTA Theorem, rearrange the states so that the non-absorbing states of the Markov chain are $1, 2, \dots, r$ and the absorbing states are $r + 1, \dots, s$. Then the one-step transition matrix \mathbf{P} can be partitioned as follows:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 \\ \vdots \\ r \\ r+1 \\ \vdots \\ s \end{matrix} & \left[\begin{array}{c|c} & \\ \hline \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{O} & \mathbf{I} \end{array} \right] \end{matrix} \quad (6.8)$$

Expression (6.8) is sometimes called the **canonical form** of a Markov chain. In Eq. (6.8), \mathbf{Q} is the $r \times r$ sub-matrix for the non-absorbing states, as before. The matrix \mathbf{O} in the lower left of Eq. (6.8) consists entirely of zeros, since that quadrant of \mathbf{P} indicates the probabilities of transitioning from an absorbing state ($r + 1, \dots, s$) to a *non*-absorbing state ($1, \dots, r$). Similarly, \mathbf{I} is the $(s - r) \times (s - r)$ identity matrix, since its diagonal entries correspond to $P(a \rightarrow a)$ for the absorbing states and its off-diagonal entries to impossible events (transitions from one absorbing state to another). The “remainder” matrix \mathbf{R} indicates the transition probabilities from the non-absorbing states into the absorbing states and can have (fairly) arbitrary entries.

The probabilities of eventual absorption into every absorbing state from every non-absorbing state are provided by the following theorem.

THEOREM

Consider a Markov chain with non-absorbing states $1, \dots, r$ and absorbing states $r + 1, \dots, s$. Define sub-matrices \mathbf{Q} and \mathbf{R} of the one-step transition matrix \mathbf{P} as in Eq. (6.8). Suppose further than every absorbing state is accessible from every non-absorbing state. Then the probabilities of eventual absorption are given by

$$\mathbf{\Pi} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R},$$

where \mathbf{I} is the $r \times r$ identity matrix and $\mathbf{\Pi}$ is an $r \times (s - r)$ matrix whose entries are the probabilities $\pi(i \rightarrow a)$ for $i = 1, \dots, r$ and $a = r + 1, \dots, s$.

Some guidance for the proof of this theorem can be found in Exercise 57.

Example 6.32 (Example 6.29 continued) To apply the previous theorem to our Gambler's Ruin example, we need to reorder the states, so that non-absorbing states \$1 and \$2 come first while absorbing states \$0 and \$3 come last. The canonical form of \mathbf{P} , along with the relevant sub-matrices \mathbf{Q} and \mathbf{R} , are

$$\mathbf{P} = \begin{array}{c} 1 \\ 2 \\ 0 \\ 3 \end{array} \left[\begin{array}{cc|cc} 0 & p & 1-p & 0 \\ 1-p & 0 & 0 & p \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right], \mathbf{Q} = \begin{bmatrix} 0 & p \\ 1-p & 0 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1-p & 0 \\ 0 & p \end{bmatrix}$$

Applying the previous theorem, along with the inverse formula (6.7) for a 2×2 matrix,

$$\begin{aligned} \mathbf{\Pi} &= (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R} = \begin{bmatrix} 1 & -p \\ -1+p & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1-p & 0 \\ 0 & p \end{bmatrix} \\ &= \frac{1}{1-p+p^2} \begin{bmatrix} 1 & p \\ 1-p & 1 \end{bmatrix} \begin{bmatrix} 1-p & 0 \\ 0 & p \end{bmatrix} = \begin{bmatrix} \frac{1-p}{1-p+p^2} & \frac{p^2}{1-p+p^2} \\ \frac{1-2p+p^2}{1-p+p^2} & \frac{p}{1-p+p^2} \end{bmatrix} \end{aligned}$$

Reading off the entries of the matrix $\mathbf{\Pi}$, we have

$$\begin{aligned} \pi(\$1 \rightarrow \$0) &= \frac{1-p}{1-p+p^2} & \pi(\$1 \rightarrow \$3) &= \frac{p^2}{1-p+p^2} \\ \pi(\$2 \rightarrow \$0) &= \frac{1-2p+p^2}{1-p+p^2} & \pi(\$2 \rightarrow \$3) &= \frac{p}{1-p+p^2} \end{aligned}$$

In particular, if Allan starts with \$2, the probability he will eventually win the competition is $\pi(\$2 \rightarrow \$3) = p/(1 - p + p^2)$. As a check, this probability equals zero when $p = 0$ (Allan never wins games) and equals one when $p = 1$ (Allan always wins games). If $p = .55$, as in several of the previous examples in this chapter,

$$\pi(\$2 \rightarrow \$3) = \frac{.55}{1 - .55 + .55^2} = \frac{.55}{.7525} = \frac{220}{301} \approx .7309$$

Notice that this is, to four decimal places, the probability we approximated by computing \mathbf{P}^{75} with software and thereby obtaining $P^{(75)}(\$2 \rightarrow \$3)$. ■

The matrices \mathbf{R} and $\mathbf{\Pi}$ in Example 6.32 are square, but this is not necessarily the case in other scenarios. In general, \mathbf{Q} is an $r \times r$ matrix (hence, square), but the dimensions of both \mathbf{R} and $\mathbf{\Pi}$ are $r \times (s - r)$.

6.5.5 Exercises: Section 6.5 (44–58)

44. Explain why a Markov chain with one or more absorbing states cannot be a regular chain.
45. A local community college offers a three-semester athletics training (AT) program. Suppose that at the end of each semester, 75% of students successfully move on to the next semester (or to graduation from the third semester) and 25% are required to repeat the most recent semester.
 - (a) Construct a transition matrix to represent this scenario. The four states are (1) first semester, (2) second semester, (3) third semester, (4) graduate.
 - (b) What is the probability a student graduates the program within three semesters? Four semesters? Five semesters?
 - (c) What is the average number of semesters required to graduate from this AT program?
 - (d) According to this model, what is the probability of eventual graduation? Does that seem realistic?
46. Refer back to the previous exercise. Now suppose that at the end of each semester, 75% of students successfully move on to the next semester (or to graduation from the third semester), 15% flunk out of the program, and 10% repeat the most recent semester.
 - (a) Construct a transition matrix to represent this updated situation by adding a fifth state, (5) flunk out. [*Hint*: Two of the five states are absorbing.]
 - (b) What is the probability a student exits the program, either by graduating or flunking out, within three semesters? Four semesters? Five semesters?
 - (c) What is the average number of semesters students spend in this program before exiting (again, either by graduating or flunking out)?
 - (d) What proportion of students that enter the program eventually graduate? What proportion eventually flunk out?
 - (e) Given that a student has passed the first two semesters (and, so, is currently in her third-semester courses), what is the probability she will eventually graduate?
47. The article “Utilization of Two Web-Based Continuing Education Courses Evaluated by Markov Chain Model” (*J. Am. Med. Inform. Assoc.* 2012: 489-494) compared students’ flow between pages of an online course for two different Web layouts in two different health professions classes. In the first week of the classes, students could visit (1) the homepage, (2) the syllabus, (3) the introduction, and (4) chapter 1 of the course content. Each student was tracked until s/he either reached chapter 1 or exited without reading chapter 1 (call the latter state 5). For one version of the Web content in one class, the following transition matrix was estimated:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ .21 & 0 & .33 & .05 & .41 \\ .09 & .15 & 0 & .67 & .09 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

When students log into the course, they are always forced to begin on the homepage.

- (a) Identify the absorbing state(s) of this chain.
 - (b) Let T_1 = the number of transitions students take, starting from the homepage, until the either arrive at chapter 1 or exit early. Determine $P(T_1 \leq k)$ for $k = 1, 2, \dots, 10$.
 - (c) Use (b) to approximate the pmf of T_1 , and then approximate the mean time to absorption starting from the class homepage.
 - (d) Determine the (true) mean time to absorption starting from the homepage.
 - (e) What proportion of students eventually got to chapter 1 in the first week? What proportion exited the course without visiting chapter 1?
48. Refer back to the previous exercise. After some content redesign, the same Web-based health professions course was run a second time. The first-week transition probabilities for the revised course were as follows:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ .15 & 0 & .43 & .06 & .36 \\ .09 & .16 & 0 & .66 & .09 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- (a) How did the redesign affect the average amount of time students spent in the course (at least as measured by the number of Web page visits within a session)?
 - (b) Did the redesign improve the chances that students would get to the chapter 1 content before exiting the system?
49. In Exercise 4, we introduced a game in which Michelle will flip a coin until she gets heads four times in a row. Define $X_0 = 0$ and, for $n \geq 1$, X_n = the number of heads in the current streak of heads after the n th flip.
- (a) Construct the one-step transition matrix \mathbf{P} for this chain, on the state space $\{0, 1, 2, 3, 4\}$. What is special about state 4?
 - (b) Let T_0 denote the total number of coin flips required by Michelle to achieve four heads in a row. Construct the cdf of T_0 , $P(T_0 \leq k)$, for $k = 1, 2, \dots, 15$. [Hint: The cdf values for $k = 1, 2, 3$ should be obvious.]
 - (c) Michelle will win a prize if she can get four heads in a row within 10 coin flips. What is the probability she wins the prize?
 - (d) Use (b) to construct an incomplete pmf of T_0 . Then use this incomplete pmf to approximate both the mean and standard deviation of T_0 .
 - (e) What is the (exact) expected number of coin flips required for Michelle to get four heads in a row?
50. Refer back to Exercise 8. The article referenced in that exercise provides the following transition matrix for the states (1) current, (2) delinquent, (3) loss, and (4) paid, for a certain class of loans:

$$\mathbf{P} = \begin{bmatrix} .95 & .04 & 0 & .01 \\ .15 & .75 & .07 & .03 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- (a) Identify the absorbing state(s).
 - (b) Determine the mean time to absorption for a loan customer who is current on payments, and for a customer who is delinquent.
 - (c) If a loan customer is current on payments, what is the probability s/he will eventually pay off the loan? What is the probability the loan company will suffer a loss on this account?
 - (d) Answer (c) for customers who are delinquent on their loans.
51. Refer back to Exercise 15(c). Calculate and interpret the mean times to absorption for this chain. For which opening strategy, cooperative or competitive, is the negotiation process longer on the average?
 52. Refer back to Exercise 14. Assuming Lucas begins searching for his uncle in room 1 and his uncle is hiding in room 6, what is the expected number of rooms Lucas will visit in order to “win” this round of hide-and-seek?
 53. Modify the Gambler’s Ruin example of this section to a \$4 total stake. That is, Allan may start with $x_0 = \$1, \$2, \text{ or } \$3$, and Beth has $\$(4 - x_0)$ initially. As usual, let p denote the probability Allan wins any single game.
 - (a) Construct the one-step transition matrix.
 - (b) Determine the mean times to absorption for each of Allan’s possible starting values, as functions of p .
 - (c) Determine the probability Allan eventually wins, starting with \$1 or \$2 or \$3, as functions of p .
 54. Refer back to the Ehrenfest chain model introduced in Exercise 2. Suppose there are $m = 3$ balls being exchanged between the two chambers. If the left chamber is currently empty, what is the expected number of exchanges until it is full (i.e., all 3 balls are on the left side)?
 55. Refer back to Exercise 40. If a man has a low-prestige occupation, what is the expected number of generations required for him to have an offspring with a high-prestige occupation?
 56. Exercise 48 of Chap. 2 established the following formula for the mean of a rv X whose possible values are positive integers:

$$E(X) = 1 + \sum_{x=1}^{\infty} [1 - F(x)],$$

where $F(x)$ is the cdf of X . Hence, if the values $F(1), F(2), \dots, F(x^*)$ are known for some integer x^* , the mean of X can be approximated by $1 + \sum_{x=1}^{x^*} [1 - F(x)]$.

- (a) Refer back to Example 6.26. Use the given cdf values and the above expression with $x^* = 12$ to approximate $E(T_1)$, the mean time to absorption starting in state 1. How does this compare to the pmf-based approximation in the example? How does it compare to the exact answer, 4.5?
- (b) Repeat part (a), starting in state 4 of the bank loan Markov chain.
- (c) Will this method always under-approximate the true mean of the rv, or can you tell? Explain.

[Note: It can be shown that this “cdf method” of approximating the mean will always produce a higher value than the truncated sum of $x \cdot p(x)$.]

57. This exercise outlines a proof of the formula $\boldsymbol{\Pi} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R}$ for the probabilities of eventual absorption. You will want to refer back to Eq. (6.8), as well as the proof of the MTTA Theorem.
- (a) Starting in a non-absorbing state i , the chain will eventually be absorbed into absorbing state a if either (1) the chain transitions immediately into a , or (2) the chain transitions into any non-absorbing state and then eventually is absorbed into state a . Use this to explain why

$$\pi(i \rightarrow a) = P(i \rightarrow a) + \sum_{j \in A'} P(i \rightarrow j)\pi(j \rightarrow a),$$

where A' denotes the set of non-absorbing states of the chain.

- (b) The equation in (a) holds for all $i \in A'$ and all $a \in A$. Show that this collection of equations can be rewritten in matrix form as $\boldsymbol{\Pi} = \mathbf{R} + \mathbf{Q}\boldsymbol{\Pi}$, and then solve for $\boldsymbol{\Pi}$. (You may assume the matrix $\mathbf{I} - \mathbf{Q}$ is invertible.)
58. The matrix $(\mathbf{I} - \mathbf{Q})^{-1}$ arises in several contexts in this section. This exercise provides an interpretation of its entries. Consider a Markov chain with at least one absorbing state, and assume that every non-absorbing state can access at least one absorbing state. As before, A and A' will denote the sets of absorbing and non-absorbing states, respectively.
- (a) Consider any two non-absorbing states i and j . Let μ_{ij} denote the expected number of visits to state j , starting in state i , before the chain is absorbed. (When $j = i$, the initial state is counted as one visit.) Mimicking the proof of the MTTA Theorem, show that

$$\begin{aligned} \mu_{ii} &= \sum_{a \in A} 1 \cdot P(i \rightarrow a) + \sum_{k \in A'} (1 + \mu_{ki}) \cdot P(i \rightarrow k) \\ &= 1 + \sum_{k \in A'} \mu_{ki} \cdot P(i \rightarrow k) \end{aligned}$$

- (b) Using similar reasoning, show that for $i \neq j$,

$$\mu_{ij} = \sum_{k \in A'} \mu_{kj} P(i \rightarrow k)$$

- (c) Let \mathbf{M} be the square matrix whose (i, j) th entry is μ_{ij} . Combine (a) and (b) to establish the equation $\mathbf{M} = \mathbf{I} + \mathbf{Q}\mathbf{M}$, and solve for \mathbf{M} .

6.6 Simulation of Markov chains

A typical Markov chain simulation requires two elements: the one-step transition matrix, \mathbf{P} , and an indication of the initial state X_0 (either as a fixed state value or as a rv with a probability distribution). The actual simulation of a single realization of the chain X_0, X_1, X_2, \dots then amounts to repeated selections from the transitional probability distributions specified by elements of \mathbf{P} . Simulation of Markov chains allow us to confirm theoretical results and, more importantly, determine properties of Markov chains that are not covered by the theorems of this chapter or other theoretical results.

The main step in any Markov chain simulation is to simulate a value for the next step, X_{n+1} , based on the transition probabilities coming out of the current step X_n . Let's start with the initial state X_0 . Suppose for one particular run of the simulation, X_0 has been assigned the state i , either because that's the fixed initial state or because a single draw from some initial distribution \mathbf{v}_0 yielded i . Conditional

on $X_0 = i$, the distribution of X_1 is determined by the transition probabilities $P(i \rightarrow j)$ for $j = 1, 2, 3, \dots$, which appear in the i th row of \mathbf{P} . Thus, one needs to extract the i th row of \mathbf{P} and use it as the basis for a single discrete simulation. If the result of this simulation is $X_1 = j$, then the j th row of \mathbf{P} can be accessed to simulate X_2 , and so on.

Example 6.33 Let's simulate a typical day in the life of our taxi driver. Although a real taxi driver does not have the same number of fares each day, for purposes of this first simulation we'll assume that he takes exactly 25 fares in 1 day.

Suppose first that the driver begins each day in a random zone X_0 , as in Example 6.15, specifically with the initial distribution $p(1) = .2$, $p(2) = .5$, $p(3) = .3$. We begin by simulating a single value from this initial distribution. Once that is determined, our program should then simulate a single value of X_1 using the row of \mathbf{P} corresponding to the value of X_0 , then do the same for X_2 based on the

a	b
----------	----------

```

P=[.3 .2 .5; .1 .8 .1; .4 .4 .2];
states=[1 2 3];
v0=[.2 .5 .3];
X=randsample(states,1,true,v0);
current=X;
for i=1:25
    nextstate=
    randsample(states,1,true,P(current,:));
    X=[X nextstate];
    current=nextstate;
end

```

```

P <- matrix(c(.3,.2,.5,.1,.8,.1,
.4,.4,.2),nrow=3,ncol=3,byrow=TRUE)
states <- c(1,2,3)
v0 <- c(.2,.5,.3)
X <- sample(states,1,TRUE,v0)
current <- X
for (i in 1:25){
    nextstate <-
    sample(states,1,TRUE,P[current,])
    X <- c(X,nextstate)
    current <- nextstate
}

```

Fig. 6.10 Code for Example 6.33: (a) Matlab; (b) R

simulated value of X_1 , and so on. Figure 6.10 shows Matlab and R code for such a simulation.

In Matlab, `P(current, :)` calls for the row of \mathbf{P} specified by the numerical index `current`; the code `P[current,]` performs the same task in R. The output of both of these programs is a vector, `X`, containing the sequence of states for the Markov chain (beginning with X_0). For example, one run of the above program in R yielded the following output:

```

> X
[1] 2 1 3 1 2 2 2 2 1 1 1 1 3 1 3 3 1 3 2 2 2 2 3 3 1 3

```

The randomly selected initial state was $X_0 = 2$, followed by $X_1 = 1$, $X_2 = 3$, \dots , and finally $X_{25} = 3$. (The symbol `[1]` at the left is not the initial state; this is just R's way of denoting the beginning of `X`.) If we weren't interested in the initial state of the chain, the code could easily be modified not to store X_0 , in which case the indices of the output vector would match the time indices of the Markov chain (i.e., the subscripts on X_1, X_2, \dots, X_{25}).

To make X_0 a fixed initial state instead of a true random variable, one need simply replace the two lines of code specifying the initial probability vector and the first random selection. In the Matlab code, the third and fourth lines could be replaced by the statement `X=3`; to fix the taxi driver's initial state as zone 3. A similar comment applies to the R code. And, again, one could choose whether or not to store the initial state as part of the output vector. ■

It is important not to confuse the *number of transitions of the chain* with the *number of runs of the simulation*. In Example 6.33, both programs simulate the chain through 25 transitions, but only a

single run. If it's our desire to keep track of the chain's behavior across many different runs, analogous to the simulations described at the ends of Chaps. 1–4, then we must add an additional layer of code, typically in the form of a surrounding “for” or “while” loop.

Example 6.34 As an illustration of the Steady-State Theorem, consider the model for Web users' browser histories discussed in Example 6.19 (refer back to that example to see the one-step transition matrix). Let's simulate the distribution of X_{75} , the Web site category of a user's 75th visited page. For variety's sake, suppose users are equally likely to start surfing the Web in any one of the five Web site categories; recall that the initial distribution of a regular chain will not affect its long-run behavior. The programs displayed in Fig. 6.11 perform 10,000 runs of simulating this Markov chain up through X_{75} . Purely to save space, the code to create \mathbf{P} has been suppressed in Fig. 6.11, but it is very similar to what appears in Fig. 6.10.

In the fourth line of code, we have employed a shortcut version of the `randsample` and `sample` functions in Matlab and R, respectively, to randomly and uniformly select a single random integer from 1 to 5 (this is the initial state). Both programs store the state of the Markov chain after 75 transitions in the vector named `X75` for each of 10,000 runs. (Notice that the intermediate states X_1, \dots, X_{74} are not permanently stored.)

The 10,000 simulated values of X_{75} from one execution of the Matlab program are summarized in the accompanying table, along with the steady-state probabilities for this chain determined in Sect. 6.4.

j	1	2	3	4	5
# of times	2822	1004	1599	3816	759
$\hat{P}(X_{75} = j)$.2822	.1004	.1599	.3816	.0759
π_j	.2840	.0948	.1659	.3791	.0758

The estimated and theoretical steady-state probabilities are quite similar. Remember that these two rows of probabilities should differ slightly for *two* reasons: first, this is only a *simulation* of 10,000 values of the rv X_{75} , so there is natural simulation error; second, the steady-state probabilities indicate the behavior of X_n as $n \rightarrow \infty$, and we don't expect the rv X_{75} to have exactly this distribution

a	b
<pre> P=not shown; X75=zeros(10000,1); for i=1:10000 current=randsample(5,1); for j=1:75 nextstate= randsample(1:5,1,true,P(current,:)); current=nextstate; end X75(i)=current; end </pre>	<pre> P <- not shown X75 <- NULL for (i in 1:10000){ current <- sample(5,1) for (j in 1:75){ nextstate <- sample(1:5,1,TRUE,P[current,]) current <- nextstate } X75[i] <- current } </pre>

Fig. 6.11 Code for Example 6.34: (a) Matlab; (b) R ■

(although it should be close).

Section 6.6 introduced the notions of time to absorption and mean time to absorption for Markov chains with one or more absorbing states. We can also use simulation to explore properties of time-to-absorption variables and first-passage times.

Example 6.35 Consider again the bank loan application process described in Example 6.25, with lone absorbing state 6 (ultimate acceptance or rejection of the application), and the random variable $T_1 =$ time to absorption from state 1 (document initiation). To simulate the distribution of T_1 , one begins the chain in state 1 and continues to simulate its transitions until it arrives in state 6. The simulation program now must keep track of how many transitions occur, rather than just where the chain ends up. Figure 6.12 shows Matlab and R code for this purpose; again, to save space, the code for entering the matrix \mathbf{P} is not included.

a	b
<pre> P=not shown; T1=zeros(10000,1); for i=1:10000 current=1; steps=0; while current~=6 steps=steps+1; nextstate= randsample(1:6,1,true,P(current,:)); current=nextstate; end T1(i)=steps; end </pre>	<pre> P <- not shown T1 <- NULL for (i in 1:10000){ current <- 1 steps <- 0 while (current!=6){ steps <- steps+1 nextstate <- sample(1:6,1,TRUE,P[current,]) current <- nextstate } T1[i] <- steps } </pre>

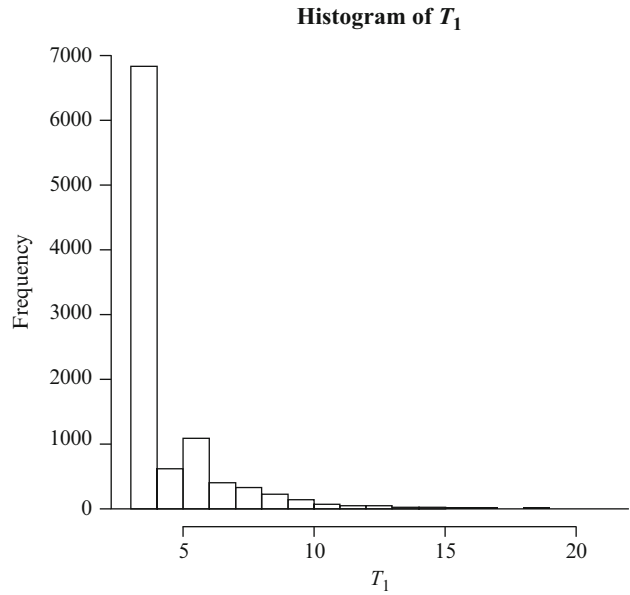
Fig. 6.12 Code for Example 6.35: (a) Matlab; (b) R

The simulated distribution of T_1 from one execution of the R code in Fig. 6.12b appears in Fig. 6.13. These particular 10,000 simulated values had a sample mean of 4.508 and a sample standard deviation of 2.276. Notice the sample mean is quite close to the theoretical expectation, $E(T_1) = 4.5$, determined in Sect. 6.5.

Clearly, the sample mean of the simulated T_1 values is a better estimate of $E(T_1)$ than the approach utilizing the truncated pmf presented in the previous section. Of course, neither is strictly necessary since the mean of T_1 can be found explicitly using the MTTA Theorem. The new information provided by the simulation is a measure of the *variability* of T_1 : we estimate the standard deviation of T_1 to be 2.276, whereas no simple matrix formula exists for its theoretical standard deviation. ■

The preceding examples employed simulations primarily to confirm theoretical results established in earlier sections. (Or, perhaps better put, our earlier theoretical results validate the simulations!) The final two examples of this section indicate situations where we must rely on simulation methods to approximate values of desired quantities.

Fig. 6.13 Simulation distribution of T_1 in Example 6.35



Example 6.36 Refer back to Example 6.13, which described Chinese cell phone users' transitions between three major carriers. Suppose users may renew or change contracts annually, and that annual plans for the three carriers cost the following (in \$US): 550 for China Telecom, 600 for China Unicom, and 525 for China Mobile. Assume that, because of governmental regulations, these prices will remain the same for the next 10 years. If last year the market shares of the three carriers were .4, .2, and .4, respectively and all contracts are about to come up for renewal, what is the average amount a Chinese cell phone customer will pay over the next decade?

We will employ a Markov chain simulation to model the behavior of customers' carrier choices for 10 consecutive years. Critically, we must keep track of how much money a customer spends each year—that is, our three states now have associated quantitative values. (This is a common instance where simulation proves useful.) Let Y = the total cost of ten 1-year calling plans for a Chinese cell phone customer. Figure 6.14 shows code for simulating Y using the techniques of this section.

An initial state x_0 is first determined using the specified initial probability distribution (here, $v_0 = [.4 .2 .4]$). Then, ten steps of the Markov chain are simulated; each of these states is temporarily held in `nextstate`. The vector `AnnualCost` stores the cost of a 1-year calling plan by calling the appropriate element of the `Prices` vector. Once a 10-year chain has been simulated, those annual costs are summed and stored as a simulated value of Y .

```

a
P=[.84 .06 .1;.08 .82 .1;.1 .04 .86];
Prices=[550 600 525];
Y=zeros(10000,1);
for i=1:10000
    v0=[.4 .2 .4];
    AnnualCost=zeros(10,1);
    x0=randsample(1:3,1,true,v0);
    current=x0;
    for n=1:10
        nextstate=
randsample(1:3,1,true,P(current,:));
        AnnualCost(n)=
Prices(nextstate);
        current=nextstate;
    end
    Y(i)=sum(AnnualCost);
end

b
P <- matrix(c(.84,.06,.1,.08,.82,.1,
.1,.04,.86),nrow=3,ncol=3,byrow=TRUE)
Prices <- c(550,600,525)
Y <- NULL
for (i in 1:10000){
    v0 <- c(.4,.2,.4)
    AnnualCost <- NULL
    x0 <- sample(1:3,1,TRUE,v0)
    current <- x0
    for (n in 1:10){
        nextstate <-
sample(1:3,1,TRUE,P[current,])
        AnnualCost[n] <-
Prices[nextstate];
        current <- nextstate
    }
    Y[i]=sum(AnnualCost)
}

```

Fig. 6.14 Code for Example 6.36: (a) Matlab; (b) R

A histogram of the 10,000 simulated Y values appears in Fig. 6.15. Notice that the distribution of Y has three spikes, at \$5250, \$5500, and \$6000. These correspond to customers who keep the same carrier all 10 years; the large probabilities along the main diagonal of the transition matrix indicate reasonably strong customer loyalty. For this particular run, the simulated values of Y had a sample mean and standard deviation of \$5503.40 and \$199.32, respectively, from which we can be 95% confident, using the methods of Sect. 5.3, that μ_Y is between \$5499.49 and \$5507.31.

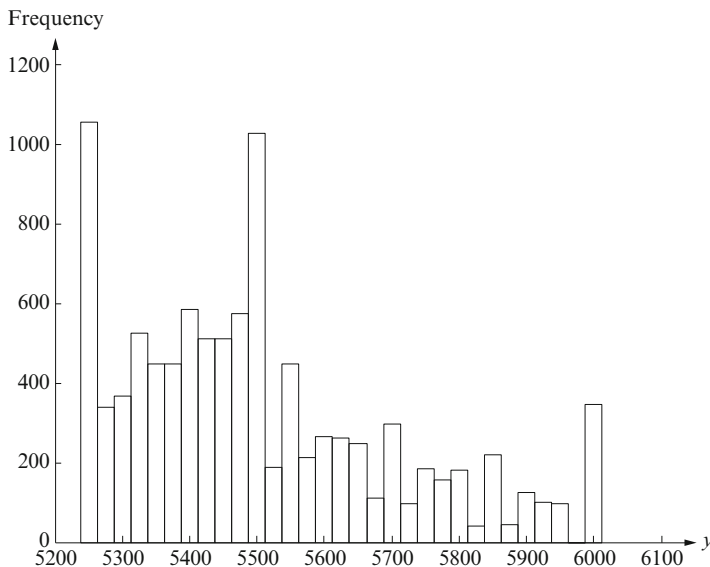


Fig. 6.15 Histogram of values of Y in Example 6.36

Example 6.37 Our taxi driver now makes one last appearance (hopefully to applause). He starts each morning at home in zone 3. Methods from Sects. 6.4 and 6.5 allow us to determine the expected number of fares required for him to return home, or to reach one of the other two zones. But how long does it take him, in the typical day, to visit *all three zones*? Let

T_{all} = number of transitions required to visit every state at least once (not counting the initial state, X_0)

To simulate T_{all} , our program must keep track of which states have been visited thus far. Once all states/zones have been visited, the numerical value of T_{all} for that simulation run can be recorded. Figure 6.16 shows appropriate code.

```

a
P = [.3 .2 .5; .1 .8 .1; .4 .4 .2];

Tall=zeros(10000,1);
for i=1:10000
    current=3;
    visits=[0 0 0];
    Talltemp=0;
    while (sum(visits)<3)
        nextstate=randsample(1:3,1,
            true,P(current,:));
        visits(nextstate)=1;
        current=nextstate;
        Talltemp=Talltemp+1;
    end
    Tall(i)=Talltemp;
end

b
P <- matrix(c(.3,.2,.5,.1,.8,.1,
    .4,.4,.2),nrow=3,ncol=3,byrow=TRUE)
Tall <- NULL
for (i in 1:10000){
    current <- 3
    visits <- c(0,0,0)
    Talltemp <- 0
    while (sum(visits)<3){
        nextstate <-
            sample(1:3,1,TRUE,P[current,])
        visits[nextstate] <- 1
        current <- nextstate
        Talltemp <- Talltemp+1
    }
    Tall[i] <- Talltemp
}

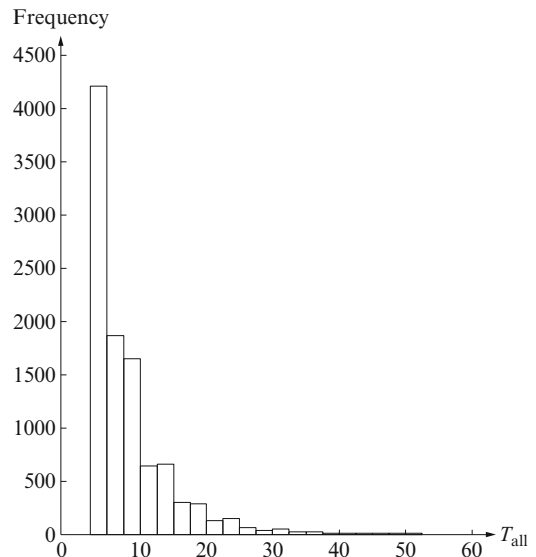
```

Fig. 6.16 Code for Example 6.37: (a) Matlab; (b) R

In both programs, a vector called `visits` keeps a record of which states the chain visits within that particular run. When chain j is visited ($j = 1, 2, 3$), the j th entry of `visits` switches from 0 to 1. Once all three entries of `visits` equal 1, as detected by its sum, the while loop terminates and the temporary count of transitions (`Talltemp`) is stored in `Tall`. The result of the program is 10,000 simulated values of the rv T_{all} , stored in the vector `Tall`.

Figure 6.17 displays a histogram of the 10,000 values resulting from running the Matlab program in Fig. 6.16a. The sample mean and standard deviation of these 10,000 values were $\bar{x} = 8.1674$ and $s = 5.8423$. Hence, we estimate the average number of fares required for the taxi driver to visit all three zones to be 8.1674, with an estimated standard error of $s/\sqrt{n} = 5.8423/\sqrt{10,000} = 0.058423$. Using the techniques of Chap. 5, we can say with 95% confidence that μ_{all} , the true mean of T_{all} , lies in the interval

Fig. 6.17 Simulation distribution of the rv T_{all} in Example 6.37



$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} = 8.1674 \pm 1.96(0.058423) = (8.053, 8.282)$$

Among the 10,000 simulated values of T_{all} , 4204 were at most 5 (so, 3 or 4 or 5). Hence, the estimated probability that the taxi driver visits all three zones within his first five fares is

$$\hat{p} = \hat{P}(T_{\text{all}} \leq 5) = \frac{4204}{10,000} = .4204$$

The estimated standard error of this estimate is given by $\sqrt{\hat{p}(1-\hat{p})/n} = .0049$. Hence we are 95% confident that the true probability $P(T_{\text{all}} \leq 5)$ lies in $.4204 \pm 1.96(.0049) = (.4108, .4300)$. ■

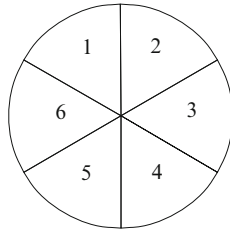
6.6.1 Exercises: Section 6.6 (59–66)

59. Refer back to Exercise 3. Suppose this machine produces 150 units on days when it is fully operational, 75 units per day when partially operational, and 0 units when broken. Consider a month with 20 work days, and assume the machine ended the previous month fully operational.
- Write a simulation of the rv Y = the number of units produced by this machine in 20 work days.
 - Create a histogram of simulated values of Y for at least 10,000 simulation runs.
 - Construct a 95% confidence interval for the mean number of units produced by this machine across 20 work days.
 - Construct a 95% confidence interval for the probability that the machine produces at least 2000 units in such a month.
60. Four friends A, B, C, and D are notorious for sharing rumors amongst themselves. Being very gossipy but not particularly bright, each friend is equally likely to share a rumor with any of the other three friends, even if that friend has already heard it. (For example, if friend B most recently heard the rumor, each of friends A, C, and D is equally likely to hear it next, regardless of how B came to hear the rumor!) Let X_n = the n th person in this foursome to hear a particular rumor.
- Construct the one-step transition matrix for this Markov chain.
 - Friend A has just overheard a particularly nasty rumor about a classmate and is eager to share it with the other three friends. Let T equal the number of times the rumor is repeated within the foursome until all of them have heard the rumor. Write a program to simulate T , and use your program to estimate $E(T)$.
61. A state lottery official has proposed the following system for a new game. In the first week of a new year, a \$10 million prize is available. If nobody gets the winning lottery numbers correct and wins the prize that week, the value doubles to \$20 million for the second week; otherwise, the prize for the second week is also \$10 million. Each week, the prize value doubles if nobody wins it and returns to \$10 million otherwise. Suppose that there is a 40% chance that someone in the state wins the lottery prize each week, irrespective of the current value of the prize. Let X_n = the value of the lottery prize in the n th week of the year.
- Determine the one-step transition probabilities for this chain. [Hint: Given the value of X_n , X_{n+1} can only be one of two possible dollar amounts.]
 - Let M be the maximum value the lottery prize achieves over the course of a 52-week year. Simulate at least 10,000 values of the rv M , and report the sample mean and SD of these simulated values. [Hint: Given the large state space of this Markov chain, don't attempt to construct the transition matrix. Instead, code the probabilities in (a) directly.]

- (c) Let Y be the total amount paid out by the lottery in a 52-week year. Simulate at least 10,000 values of the rv Y , and report a 95% confidence interval for $E(Y)$.
- (d) Repeat (c), but now assume the probability of a winner is .7 each week rather than .4. Should the lottery commission make it easier or harder for someone to win each week? Explain.
62. Write a Markov chain simulation program with the following specifications. The inputs should be the transition matrix \mathbf{P} , an initial state x_0 , and the number of steps n . The output should be a single realization of X_1, X_2, \dots, X_n , as either a row vector or a column vector.
63. Refer back to Exercise 12. Suppose that the typical annual premium for a category 1 (safest) customer is \$500; for category 2, \$600; for category 3, \$1000; and for category 4 (riskiest driver), \$1500.
- (a) Use a Markov chain simulation to estimate the distribution of the rv $Y_1 =$ total premium paid by a customer over 10 years with the insurance company, assuming s/he starts in category 1. Create a histogram of values for Y_1 , and construct a 95% confidence interval for $E(Y_1)$.
- (b) Repeat (a) assuming instead that the customer starts as a category 3 driver.
64. Write a simulation program for Gambler's Ruin. The inputs should be $a =$ Allan's initial stake, $b =$ Beth's initial stake (so $a + b$ is the total stake), $p =$ the probability Allan defeats Beth in any single game, and $N =$ the number of tournaments to be simulated. The program should output two N -by-1 vectors: one recording the number of games played for each of the N runs, and one indicating who won each time. Use your program to determine (a) the average tournament length and (b) the probability Allan eventually wins for the settings $a = b = \$5$ and $p = .4$. Give 95% confidence intervals for both answers.
65. Example 6.3 describes a (one-dimensional) random walk. This is sometimes called a *simple random walk*.
- (a) Write a program to simulate the first 100 steps of a random walk starting at $X_0 = 0$. [Hint: If $X_n = s$, then $X_{n+1} = s \pm 1$ with probability 1/2 each.]
- (b) Run your program in (a) 10,000 times, and use the results to estimate the probability that a random walk returns to its origin at any time within the first 100 steps.
- (c) Let $R_0 =$ the number of returns to the origin in the first 100 steps of the random walk, not counting its initial state. Use your simulation to (1) create a histogram of simulated values of R_0 and (2) construct a 95% confidence interval for $E(R_0)$.
66. A *two-dimensional random walk* is a model for movement along the integer lattice in the xy -plane, i.e., points (x, y) where x and y are both integers. The "walk" begins at $X_0 = (0, 0)$. At each time step, a move is made one unit left or right (probability 1/2 each) and, independently, one unit up or down (also equally likely). If $X_n =$ the (x, y) -coordinates of the chain after n steps, then X_n is a Markov chain.
- (a) Write a program to simulate the first 100 steps of a two-dimensional random walk. [Hint: The x - and y -coordinates of a two-dimensional random walk are each simple random walks. Since they are independent, the x - and y -movements can be simulated separately.]
- (b) Use your program in (a) to estimate the probability that a two-dimensional random walk returns to its origin within the first 100 steps. Use at least 10,000 runs.
- (c) Use your program in (a) to estimate $E(R_0)$, where $R_0 =$ the number of times the walk returns to $(0, 0)$ in the first 100 steps.

6.7 Supplementary Exercises (67–82)

67. A hamster is placed into the six-chambered circular habitat shown in the accompanying figure. Sitting in any chamber, the hamster is equally likely to next visit either of the two adjacent chambers. Let X_n = the n th chamber visited by the hamster.



- Construct the one-step transition matrix for this Markov chain.
 - Is this a regular Markov chain?
 - Intuitively, what should the stationary probabilities of this chain be? Verify these are indeed its stationary probabilities.
 - Given that the hamster is currently in chamber 3, what is the expected number of transitions it will make until it returns to chamber 3?
 - Given that the hamster is currently in chamber 3, what is the expected number of transitions it will make until it arrives in chamber 6?
68. Teenager Mike wants to borrow the car. He can ask either parent for permission to take the car. If he asks his mom, there is a 20% chance she will say “yes,” a 30% chance she will say “no,” and a 50% chance she will say, “ask your father.” Similarly, the chances of hearing “yes”/“no”/“ask your mother” from his dad are .1, .2, and .7, respectively. Imagine Mike’s efforts can be modeled as a Markov chain with states (1) talk to Mom, (2) talk to Dad, (3) get the car (“yes”), (4) strike out (“no”). Assume that once either parent has said “yes” or “no,” Mike’s begging is done.
- Construct the one-step transition matrix for this Markov chain.
 - Identify the absorbing state(s) of the chain.
 - Determine the mean times to absorption.
 - Determine the probability that Mike will eventually get the car if (1) he asks Mom first and (2) he asks Dad first. Whom should he ask first?
69. Refer back to Exercise 14. Suppose Lucas starts in room 1 and proceeds as described in that exercise; however, his mean-spirited uncle has snuck out of the house entirely, leaving Lucas to search interminably. So, in particular, if Lucas enters room 6 of the house, his next visit will necessarily be to room 5. (This really happened one summer!)
- Determine the transition matrix for this chain.
 - Verify that this Markov chain is regular.
 - Determine the steady-state probabilities of this chain.
 - What proportion of time in the long run does Lucas spend in room 2?
 - What is the average number of room transitions between Lucas’ visits to room 1?
70. Refer back to Exercises 20 and 21.
- Suppose all four vans were operational as of Monday morning. What is the expected backlog—that is, the expected number of vans needing repair—as of Friday evening?
 - Suppose instead that two of the four vans were down for repairs Monday morning. Now what is the expected backlog as of Friday evening?

71. Five Mercedes E550 vehicles are shipped to a local dealership. The dealer sells one E550 in any week with probability .3 and otherwise sells none in that week. When all E550s in stock have been sold, the dealer requests a new shipment of five such cars, and it takes 1 week for that delivery to occur. Let X_n = the number of Mercedes E550s at this dealership n weeks after the initial delivery of five cars.
- Construct the transition matrix for this chain. [*Hint*: The states are 0, 1, 2, 3, 4, 5.]
 - Determine the steady-state probabilities for this chain.
 - On the average, how many weeks separate successive orders of five E550s?
72. Refer back to the previous exercise. Let m = the number of Mercedes E550s delivered to the dealership at one time (both initially and subsequently), and let p = the probability an E550 is sold in any particular week ($m = 5$ and $p = .3$ in the previous exercise). Determine the steady-state probabilities for this chain and then the average number of weeks between vehicle orders.
73. Sports teams can have long streaks of winning (or losing) seasons, but occasionally a team's fortunes change quickly. Suppose that each team in the population of all college football teams can be classified as (1) weak, (2) medium, or (3) strong, and that the following one-step transition probabilities apply to the Markov chain X_n = a team's strength n seasons from now:

$$\mathbf{P} = \begin{bmatrix} .8 & .2 & 0 \\ .2 & .6 & .2 \\ .1 & .2 & .7 \end{bmatrix}$$

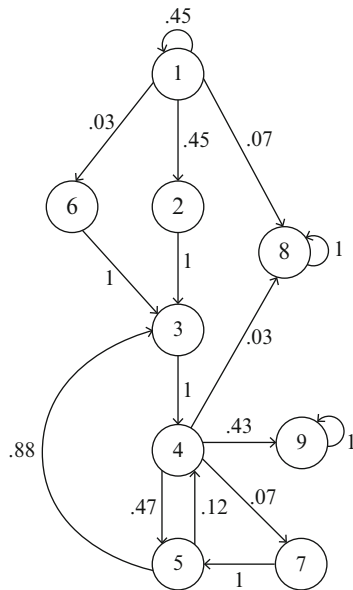
- If a college football team is weak this season, what is the minimum number of seasons required for it to become strong?
 - If a team is strong this season, what is the probability it will also be strong four seasons from now?
 - What is the average number of seasons that must pass for a weak team to become a strong team?
 - What is the average number of seasons that must pass for a strong team to become a weak team?
74. Jay and Carol enjoy playing tennis against each other. Suppose we begin watching them when they are at *deuce*. This means the next person to win a point earns *advantage*. If that same person scores the next point, then s/he wins the game; otherwise, the game returns to *deuce*.
- Construct a transition matrix to describe the status of the game after n points have been scored (starting at *deuce*). [*Hint*: There are five states: (1) Jay wins, (2) advantage Jay, (3) *deuce*, (4) advantage Carol, (5) Carol wins.]
 - Suppose Carol is somewhat better than Jay and has a 60% chance of winning any particular point. Determine (1) the probability Carol eventually wins and (2) the expected number of points to be played, starting at *deuce*. [*Hint*: This should bear surprising similarity to a game played earlier in the chapter by Allan and Beth!]
75. The authors of the article "Pavement Performance Modeling Using Markov Chain" (*Proc. ISEUSAM*, 2012: 619–627) developed a system for classifying pavement segments into five categories: (1) Very good, (2) Good, (3) Fair, (4) Bad, and (5) Very bad. Analysis of pavement samples led to the construction of the following transition matrix for the Markov chain X_n = pavement condition n years from now:

$$\begin{bmatrix} .958 & .042 & 0 & 0 & 0 \\ 0 & .625 & .375 & 0 & 0 \\ 0 & 0 & .797 & .203 & 0 \\ 0 & 0 & 0 & .766 & .234 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notice that a pavement segment either maintains its condition or goes down by one category each year.

- (a) The evaluation of one particular stretch of road led to the following initial probability vector (what the authors call a “condition matrix”): [.3307 .2677 .2205 .1260 .0551]. Use the Markov chain model to determine the condition matrix of this same road section 1, 2, and 3 years from now.
- (b) “Very bad” road segments require repairs before they are again usable; the authors’ model applies to unrepaired road. What is the average time (number of years) that a very good road can be used before it degrades into very bad condition? Make the same determination for good, fair, and bad roads.
- (c) Suppose one road segment is randomly selected from the area to which the condition matrix in (a) applies. What is the expected amount of time until this road segment becomes very bad? [Hint: Use the results of part (b).]

76. A *constructive memory agent* (CMA) is an autonomous software unit that uses its interactions not only to change its data (“memory”) but also its fundamental indexing systems for that data (“structure”). The article “Understanding Behaviors of a Constructive Memory Agent: A Markov Chain Analysis” (*Knowledge-Based Systems*, 2009: 610–621) describes a study of one such CMA as it moved between nine different stages of learning. (Stage 1 is sensation and perception; later stages add on other behaviors such as hypothesizing, neural network activation, and validation. Consult the article for details.) The accompanying state diagram mirrors the one given in the article for the authors’ first experiment.



- (a) Construct the transition matrix for this chain.
- (b) What are the absorbing states of the chain?
- (c) All CMA processes begin in stage 1. What is the mean time to absorption for such a process? Here, “time” refers to the number of transitions from one learning stage to another. [Note: In this particular experiment, absorbing states correspond to any instance of so-called “inductive” learning.]
- (d) Starting in stage 1, what is the probability a CMA will end the experiment in state 8 (constructive learning plus inductive learning)?
77. The authors of the article “Stationarity of the Transition Probabilities in the Markov Chain Formulation of Owner Payments on Projects” (*ANZIAM J.*, v. 53, 2012: C69-C89) studied payment delays in road construction in Australia. States for any payment were defined as follows: k weeks late for $k = 0, 1, 2, 3$; paid (pd), an absorbing state; and “to be resolved” (tbr), meaning the payment was at least 1 month late, which the authors treated as another absorbing state. For one particular project, the following \mathbf{Q} and \mathbf{R} matrices were given for the canonical form of the one-step transition matrix:

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & .959 & 0 \\ 0 & 0 & 0 & .897 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad \mathbf{R} = \begin{matrix} & \begin{matrix} \text{pd} & \text{tbr} \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ .041 & 0 \\ .013 & 0 \\ .804 & .196 \end{bmatrix} \end{matrix}$$

- (a) Construct the complete 6×6 transition matrix \mathbf{P} for this Markov chain.
- (b) Draw the state diagram of this Markov chain.
- (c) Determine the mean time to absorption for payment that is about to come due (i.e., one that is presently 0 weeks late).
- (d) What is the probability a payment is eventually made, as opposed to being classified as “to be resolved”?
- (e) Consider the two probabilities $P(0 \rightarrow 1)$ and $P(3 \rightarrow \text{pd})$. What is odd about each of these values? (The authors of the article offer no explanation for the irregularity of these two particular probabilities.)
78. In a *nonhomogeneous* Markov chain, the conditional distribution of X_{n+1} depends on both the previous state X_n and the current time index n . As an example, consider the following method for randomly assigning subjects one at a time to either of two treatment groups, A or B. If n patients have been assigned a group so far, and a of them have been assigned to treatment A, the probability the next patient is assigned to treatment group A is

$$P((n+1)\text{st patient assigned to A} \mid a \text{ out of first } n \text{ in A}) = \frac{n-a+1}{n+2}$$

Hence, the first patient is assigned to A with probability $(0 - 0 + 1)/(0 + 2) = 1/2$; if the first patient was assigned to A, then the second patient is also assigned to A with probability $(1 - 1 + 1)/(1 + 2) = 1/3$. This assignment protocol ensures that each next patient is more likely to be assigned to the smaller group. Let X_n = the number of patients in treatment group A after n total patients have been assigned ($X_0 = 0$). To simplify matters, assume there are only 4 patients in total to be randomly assigned.

- (a) Let \mathbf{P}_1 denote the transition matrix from $n = 0$ to $n = 1$. Assume the state space of the chain is $\{0, 1, 2, 3, 4\}$. Construct \mathbf{P}_1 . [Hint: Since $X_0 = 0$, only the first row of \mathbf{P}_1 is really relevant.]

To make this a valid transition matrix, treat the “impossible” states 1, 2, 3, and 4 as absorbing states.]

- (b) Construct \mathbf{P}_2 , the transition matrix from $n = 1$ to $n = 2$. Use the same hint as above for states 2, 3, and 4, which are impossible at time $n = 1$.
- (c) Following the pattern of (a) and (b), construct the matrices \mathbf{P}_3 and \mathbf{P}_4 .
- (d) For a nonhomogeneous chain, the multistep transition probabilities can be calculated by multiplying the aforementioned matrices from left to right, e.g., the 4-step transition matrix for this chain is $\mathbf{P}_1\mathbf{P}_2\mathbf{P}_3\mathbf{P}_4$. Calculate this matrix, and then use its first row to determine the likelihoods of 0, 1, 2, 3, and 4 patients being randomly assigned to treatment group A using this method.

[Note: Random assignment strategies of this type were originally investigated in the article “Forcing a Sequential Experiment to be Balanced,” *Biometrika* (1971): 403-417.]

79. A communication channel consists of five relays through which all messages must pass. Suppose that bit switching errors of either kind (0 to 1, or 1 to 0) occur with probability .02 at the first relay. The corresponding probabilities for the other four relays are .03, .02, .01, and .01, respectively. If we define X_n = the parity of a bit after traversing the n th relay, then X_0, X_1, \dots, X_5 forms a nonhomogeneous Markov chain.
- (a) Determine the one-step transition matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$, and \mathbf{P}_5 .
 - (b) What is the probability that a 0 bit entering the communication relay system also exits as a 0 bit? [Hint: Refer back to the previous exercise for information on nonhomogeneous Markov chains.]
80. Consider the two-state Markov chain described in Exercise 39, whose one-step transition matrix is given by

$$\begin{array}{c} 0 \\ 1 \end{array} \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

for some $0 < \alpha, \beta < 1$. Use mathematical induction to show that the k -step transition probabilities are given by

$$\begin{aligned} P^{(k)}(0 \rightarrow 0) &= \delta^k + (1 - \pi)(1 - \delta^k) & P^{(k)}(0 \rightarrow 1) &= \pi(1 - \delta^k) \\ P^{(k)}(1 \rightarrow 0) &= (1 - \pi)(1 - \delta^k) & P^{(k)}(1 \rightarrow 1) &= \delta^k + \pi(1 - \delta^k) \end{aligned}$$

where $\pi = \alpha/(\alpha + \beta)$ and $\delta = 1 - \alpha - \beta$. [Note: Applications of these multistep probabilities are discussed in “Epigenetic Inheritance and the Missing Heritability Problem,” *Genetics*, July 2009: 845-850.]

81. A 2012 report (“A Markov Chain Model of Land Use Change in the Twin Cities, 1958-2005,” available online) provided a detailed analysis from maps of Minneapolis-St. Paul, MN over the past half-century. The Twin Cities area was divided into 610,988 “cells,” and each cell was classified into one of ten categories: (1) airports, (2) commercial, (3) highway, (4) industrial, (5) parks, (6) public, (7) railroads, (8) residential, (9) vacant, (10) water. The report’s authors found that X_n = classification of a randomly selected cell was well modeled by a time-homogeneous Markov chain when a time increment of about 8 years is employed. The accompanying matrix shows the one-step transition probabilities from 1997 (n) to 2005 ($n + 1$); rows and columns are in the same order as the sequence of states described above.

.7388	.0010	.0068	.0010	.0325	.0131	.0000	.0055	.1984	.0029
.0001	.8186	.0201	.0560	.0045	.0227	.0002	.0413	.0350	.0015
.0004	.0107	.9544	.0054	.0058	.0031	.0002	.0094	.0105	.0001
.0004	.0710	.0099	.8371	.0082	.0086	.0011	.0106	.0517	.0014
.0022	.0036	.0031	.0025	.9128	.0062	.0002	.0116	.0364	.0214
.0001	.0193	.0100	.0384	.0569	.7364	.0004	.0223	.1091	.0071
.0000	.0065	.0142	.0201	.0110	.0032	.9139	.0168	.0130	.0013
.0000	.0024	.0024	.0009	.0041	.0023	.0002	.9634	.0230	.0013
.0004	.0141	.0099	.0156	.0513	.0057	.0002	.0988	.7920	.0120
.0001	.0010	.0003	.0014	.0136	.0001	.0000	.0055	.0096	.9684

- (a) In 2005, the distribution of cell categories (out of the 610,988 total cells) was as follows: [4047 20,296 16,635 24,503 74,251 18,820 1505 195,934 200,837 54,160]

The order of the counts matches the category order above, e.g., 4047 cells were part of airports and 54,160 cells were located on water. Use the transition probabilities to predict the land use distribution of the Twin Cities region in 2013.

- (b) Determine the predicted land use distribution for the years 2021 and 2029 (remember, each time step of the Markov chain is 8 years). Then determine the percent change from 1995 to 2029 in each of the ten categories (similar computations were made in the cited report).
- (c) Though it's unlikely that land use evolution will remain the same forever, imagine that the one-step probabilities can be applied in perpetuity. What is the projected long-run land use distribution in Minneapolis-St. Paul?
82. In the article "Reaching a Consensus" (*J. Amer. Stat. Assoc.*, 1974: 118-121), Morris DeGroot considers the following situation: s statisticians must reach an agreement about an unknown population distribution, F . (The same method, he argues, could be applied to opinions about the numerical value of a parameter, as well as many nonstatistical scenarios.) Let F_{10}, \dots, F_{s0} represent their initial opinions. Each statistician then revises his belief about F as follows: the i th individual assigns a "weight" p_{ij} to the opinion of the j th statistician ($j = 1, \dots, s$), where $p_{ij} \geq 0$ and $p_{i1} + \dots + p_{is} = 1$. He then updates his own belief about F to

$$F_{i1} = p_{i1}F_{10} + \dots + p_{is}F_{s0}$$

This updating is performed simultaneously by all s statisticians (so, $i = 1, 2, \dots, s$).

- (a) Let $\mathbf{F}_0 = (F_{10}, \dots, F_{s0})^T$, and let \mathbf{P} be the $s \times s$ matrix with entries p_{ij} . Show that the vector of updated opinions $\mathbf{F}_1 = (F_{11}, \dots, F_{s1})^T$ is given by $\mathbf{F}_1 = \mathbf{P}\mathbf{F}_0$.
- (b) DeGroot assumes that updates to the statisticians' beliefs continue iteratively, but that the weights do not change over time (so, \mathbf{P} remains the same). Let \mathbf{F}_n denote the vector of opinions after n updates. Show that $\mathbf{F}_n = \mathbf{P}^n\mathbf{F}_0$.
- (c) The group is said to *reach a consensus* if the limit of \mathbf{F}_n exists as $n \rightarrow \infty$ and each entry of that limit vector is the same (so all individuals' opinions converge toward the same belief). What would be a sufficient condition on the weights in \mathbf{P} for the group to reach a consensus?
- (d) DeGroot specifically considers four possible weight matrices:

$$\mathbf{P}_A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix} \quad \mathbf{P}_B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad \mathbf{P}_C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{P}_D = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Discuss what each one indicates about the statisticians' views on each other, and determine for which matrices the group ultimately reaches a consensus. If a consensus is reached, write out the consensus "answer" as a linear combination of F_{10}, \dots, F_{s0} .